Grundlagen der Entscheidungstheorie

Wolfgang Spohn

Vorwort

Die Entscheidungstheorie ist ein etwas schwieriges Gebiet, einfach weil sie mit unserer überkommenen Einteilung wissenschaftlicher Disziplinen über Kreuz liegt. Sie fügt sich mathematischem Zugriff, was weidlich ausgenutzt wird; drum sollte man als Entscheidungstheoretiker mathematisch versiert sein. Ein Zweig der Entscheidungstheorie, der statistische, liefert ein wichtiges von mehreren statistischen Verfahren; um seine Qualitäten zu beurteilen, sollte man sich in der Statistik auskennen. Wirtschaftswissenschaftlichem Wirken entsprungen, findet die Entscheidungstheorie auch heute noch am ausgedehntesten dort Anwendung; so sollte man wissen, was es mit diesen Anwendungen auf sich hat. Nun sind die Grundbegriffe der Entscheidungstheorie eindeutig psychologische, was der Psychologie ein Recht auf Vereinnahmung der Entscheidungstheorie einräumt; am besten ist man also noch ausgebildeter Psychologe. Und natürlich sind nicht zuletzt die Philosophen und Wissenschaftstheoretiker auf den Geschmack gekommen; so empfiehlt sich schließlich Vertrautheit mit deren Prüfungen.

Das alles ist nicht gerade zu viel, aber doch ein bißchen viel verlangt; auch von Buchschreibern. Bei einem Buch wie diesem sollte daher eine gewisse Beschränktheit nicht überraschen. Man merkt es ihm überdeutlich an, daß es aus der logisch-philosophischen Ecke stammt. So bitte ich insbesondere bei denjenigen Stellen um Nachsicht, deren amateur-psychologischen Charakter zu kaschieren ich mich nicht ausgesprochen bemüht habe, und bei denjenigen Stellen um Vorsicht, in denen es um formale Präzisierungen geht, von denen ich, wenn es darauf ankommt, nicht lassen kann.

Dabei ist dieses Buch nicht das, was man mathematisch anspruchsvoll nennt. Im Gegenteil; da ich mich jeglichen Vordringens in irgendwelche Unendlichkeiten enthalte, entfällt die hartgesottene Mathematik und beschränkt sich das Rechnerische im wesentlichen auf Addition und Multiplikation. Auch sonst zieht dieses Buch nur Elementares heran, aber eben Elementares aus Gegenständen, die nicht wissenschaftliches Gemeingut sind, wie Mengenalgebra (fast dauernd), Wahrscheinlichkeitstheorie (fast dauernd), Logik (immer mal wieder), Relationsbegriffen (ab und zu) und rekursiven Definitionen (in den Abschnitten 4.3 und 4.4). Insofern ist dem Leser doch dringend Gewandtheit im Umgang mit Formalem anzuempfehlen, jene schwer definierbare Eigenschaft, die einem am ehesten immer noch durch eine nicht nur oberflächliche Auseinandersetzung mit der Mathematik zuwächst. Ohne diese formale Gewandtheit muß der Leser zwar nicht unbedingt scheitern, aber sein Anfangsschwung muß schon groß sein, um ihn durch das ganze Buch zu tragen.

Auch braucht der Leser nicht unbedingt bereits mit der Entscheidungstheorie vertraut zu sein; entscheidungstheoretische Vorbildung wird nicht vorausgesetzt. Natürlich taugt dieses Buch nicht als Lehrbuch der Entscheidungstheorie, das gewiß nicht. Aber die Grundzüge der Standard-Entscheidungstheorie – und nur um die geht es hier – sollten sich ihm aufgrund der Lektüre insbesondere des zweiten Kapitels schon erschließen. Freilich, ohne eine Kenntnis der Entscheidungstheorie, wie sie sich etwa durch Raiffa (1973), Bühlmann, Loeffel und Nievergelt (1975), Krelle (1968) oder Stegmüller (1973a) gewinnen läßt, wird dem Leser der Bezug so mancher Bemerkung und der Hintergrund so manchen Argumentes fehlen. Daher die triviale Empfehlung: Je mehr Vorkenntnisse, desto besser. Dem liegt die Hoffnung zugrunde, daß dieses Buch auch und gerade dem entscheidungstheoretisch Beschlagenen einige Anregungen zu liefern vermag.

Was den Inhalt des Buches und seine Anordnung anlangt – so sei dazu auf das Inhaltsverzeichnis und hinsichtlich der Kapitel 2-5 auf den Abschnitt 2.1 verwiesen. Erst in diesen Kapiteln widme ich mich unmittelbar der Entscheidungstheorie. Im ersten Kapitel wird ein vielerorts nahegelegter, breiterer Rahmen abgesteckt, in dem, wie sich herausstellt, eine ganze Menge nicht nur philosophischer Aktivitäten Platz findet und die Entscheidungstheorie eine zentrale Stellung einnimmt. Ich habe es seinerzeit hauptsächlich entworfen, um in meine eigenen Gedanken und Kenntnisse etwas System zu bringen, und nun den anderen Kapiteln in der Hoffnung vorangestellt, daß dem Leser dieser Rahmen, sofern ihm nicht ohnehin geläufig, von ähnlichem Nutzen ist.

Die einzelnen Themen und Gedankengänge der Reihe nach abzuhandeln, gelang mir nicht; ab einem gewissen, nicht sehr hohen Grad der Verzweigtheit ist das eben nicht mehr möglich. Dies betrifft insbesondere das im Abschnitt 2.5 begonnene Thema, inwieweit es sinnvoll ist, in ein Entscheidungsmodell subjektive Wahrscheinlichkeiten des Handelnden für seine eigenen Handlungen aufzunehmen. Seine Bedeutung schien mir ursprünglich auf die Argumentation in den Abschnitten 2.6 und 2.7 eingegrenzt, aber bei längerem Nachdenken darüber eröffneten sich immer tiefsinnigere Perspektiven, deren Andeutungen sich auf die Abschnitte 3.3, 5.1 und 5.2 verteilen. Ähnlich verzettelt (nämlich vor allem auf die Abschnitte 2.1, 2.8 und 5.3) sind meine Bemerkungen zur Metrisierungsthematik, die ich nicht systematisch angehen wollte und dann doch nicht gänzlich umgehen konnte. Schließlich stoße ich immer wieder auf die Frage, ob es sich empfiehlt, subjunktive Konditionale zur Beschreibung von Entscheidungssituationen heranzuziehen. Wie ich den Text fertiggestellt hatte, merkte ich, daß ich so häufig auf diese Frage gestoßen war, daß sie eigentlich eine systematische Behandlung verdient hätte. Nun – ich beließ es bei den verstreuten Bemerkungen, die zusammen aber doch, wie ich glaube, eine Antwort weisen.

Ansonsten hoffe ich, den Schwierigkeiten, die sich aus der Verzweigtheit des Stoffes ergeben, durch Querverweise und durch das Sachverzeichnis einigermaßen Herr geworden zu sein. Und wo wir gerade bei Verzeichnissen sind, möchte ich den Leser darauf hinweisen, daß er, sollte ihm die Bedeutung dieses oder jenes Symbols entfallen sein, mit Hilfe des Symbolverzeichnisses sein Erinnerung wieder auffrischen kann.

Dieses Buch ist eine Überarbeitung meiner Dissertation, die ich im Jahre 1976 der ehemaligen Philosophischen Fakultät I der Universität München unter dem Titel „Eine Untersuchung der begrifflichen Grundlagen der Entscheidungstheorie“ vorgelegt habe. Gegenüber jener Fassung habe ich zahlreiche mehr oder weniger kleine Änderungen und Umstellungen vorgenommen und das Kapitel 3.3 völlig überarbeitet. Die Abschnitte 4.3 und 4.4 und das Kapitel 5 sind neu hinzugekommen.

Meinem Doktorvater Herrn Prof. Stegmüller danke ich ganz herzlich für die teilnahmsvolle Betreuung meiner Arbeit. Herrn Prof. Sneed und Herrn Prof. Jeffrey danke ich für die Anregung durch einschlägige Diskussionen, ebenso Herrn Prof. Yaari für einen Hinweis, ohne den mir die wirtschaftswissenschaftliche Debatte um Nutzenänderungen vermutlich heute noch unbekannt wäre. Und wie das in der Regel so ist, was alles einen in letztlich entscheidendem Maße formt, ist zu indirekter und diffuser Natur, als daß ich es hier aufzählen könnte. Es geht von vielerlei Begegnungen und Verbindungen aus. – Und von beständiger Zuneigung.

Die ich meine, wußten und wissen darum – und sie spürten und spüren hoffentlich, daß ich das weiß.

München, März 1978 *Wolfgang Spohn*

Inhaltsverzeichnis

KAPITEL 1

Das Umfeld der Entscheidungstheorie

„Entscheidungstheorie" ist eine recht irreführende Bezeichnung für die Entscheidungstheorie; auch das in diesem Zusammenhang oft verwendete Wort „choice" ist dazu geeignet, falsche Assoziationen zu wecken. Es ist jedenfalls nicht von vornherein auszuschließen, daß auch Handlungen in die Domäne der Entscheidungstheorie fallen, von denen man normalerweise nicht sagen würde, der Handelnde habe sich für sie entschieden: Handlungen etwa, denen keine oder keine nennenswerten Überlegungen vorausgegangen sind, wie z.B. Handlungen aus Gewohnheit, oder Handlungen, die unter massivem Druck zustande kommen. Und versteht man den Begriff der Handlung intentional, so ist es nicht einmal korrekt zu sagen, Handlungen bildeten den Gegenstand der Entscheidungstheorie.

Sagen wir lieber: Die Entscheidungstheorie befaßt sich ganz allgemein mit menschlichem Verhalten. Freilich haben es auch viele andere Disziplinen damit zu tun. Von diesen unterscheidet sich die Entscheidungstheorie aber vor allem durch die Begriffe, mit denen sie menschliches Verhalten zu erfassen sucht: Präferenzrelationen und subjektive Wahrscheinlichkeits- und Nutzenfunktionen. Diese Besonderheit wollen wir nun etwas deutlicher herausschälen.

1.1 Glaubens- und Wünschensdispositionen

Menschliches Verhalten fällt unter viele Begriffe – das ohnehin –, aber auch unter viele Begriffstypen. Eine reichhaltige Klassifikation dieser Begriffstypen findet sich z.B. bei Ryle (1949). Eine wichtige Unterscheidung ist dabei die zwischen *Ereignis- oder Vorgangswörtern* und *Dispositionswörtern*. Sagen wir „Marianne fliegt nach London" oder „Konrad setzt Andreas schachmatt" oder „Konrad sagt: 'Du bist schachmatt!'" etc., so machen wir von Ereigniswörtern Gebrauch. Hingegen verwenden wir Dispositionswörter, wenn wir sagen „Uli hat die Nase voll" oder „Marianne kann nicht Auto fahren" oder „Herbert ist ein kühler Rechner" etc.

Natürlich ist das weder eine erschöpfende noch eine sonderlich klare Unterscheidung.[1] Außerdem ist es eine Unterscheidung, die eher auf Verwendungen von Wörtern als auf Wörter selbst zutrifft. Dies wird an dem altbekannten Beispiel klar, daß „Stephan trinkt" sowohl „Stephan ist ein Trinker", als auch „Stephan trinkt gerade" heißen kann. Ferner gibt es Zwitter, die sogenannten Semidispositionswörter, die in einem Zuge auf ein Ereignis und auf eine Disposition verweisen.[2]

Auch unter den Dispositionen allein lassen sich verschiedene Typen unterscheiden. Da gibt es einspurige Dispositionen wie die, Raucher zu sein, die sich nur im Rauchen selbst manifestieren kann, und mehrspurige Dispositionen wie die, Tennisexperte zu sein, die sich auf viele verschiedene Weise äußern kann. Es gibt auch Dispositionen höherer Ordnung, in deren Manifestationsgesetzen wiederum Dispositionswörter vorkommen. Will man inhaltlichere Klassifizierungsmerkmale heranziehen, so kann man kurzfristige Dispositionen wie die, Hunger zu haben, und langfristige Dispositionen wie die, einen Intelligenzquotienten zwischen 115 und 125 zu haben, unterscheiden oder auch Dispositionen nach Fähigkeiten, Gewohnheiten, Charakterzügen etc. gruppieren.

Auf zwei Sorten von Dispositionen, die unser Verhalten entscheidend bestimmen, kommt es mir nun besonders an: auf die Glaubens- und Wünschensdisposi-

---

[1] Auch sollten die paar angegebenen Beispiele nicht dazu dienen, sie klar zu machen. Für detailliertere Ausführungen s. v. Savigny (1974), S. 103-107 und S. 264-267.

[2] Vgl. v. Savigny (1974), S. 119.

tionen, wie ich sie hier – sprachlich nicht gerade schön – nennen will, um zwei möglichst neutrale Bezeichnungen zur Verfügung zu haben und mit der Begriffsbildung in der Psychologie nicht über Kreuz zu kommen.

Die *Glaubensdispositionen bestehen*, grob gesagt, darin, gewisse Überzeugungen zu haben. Besser als durch allgemeine Definitionsbemühungen läßt sich klar machen, was Glaubensdispositionen sein sollen, wenn wir uns umgangssprachliche Ausdrücke ansehen, die sie bezeichnen: „glauben“, „erwarten“, „annehmen“, „überzeugt sein“, „sicher sein“, „für möglich halten“, „für wahrscheinlich halten“, „als unmöglich ansehen“, „als gleichwahrscheinlich erachten“, „eher glauben als“, „mehr zu der Annahme neigen, daß ..., als zu der Annahme , daß ...“, „ahnen“, „voraussehen“, „vollkommen ahnungslos sein“, „auf etwas gefaßt sein“, „sich etwas einbilden“, „mit etwas rechnen“, „vermuten“, „mutmaßen“, und vieles mehr. Darüber hinaus gibt es viele umgangssprachliche Wendungen, die neben dem Vorliegen einer Glaubensdisposition noch mehr ausdrücken: Ein wichtiges Beispiel dafür ist „wissen“, das über „glauben“ zumindest darin hinausgeht, daß das Geglaubte auch wahr sein muß und daß der Wissende irgendwie in einer besonders geeigneten Lage dazu war, zu seinem Wissen zu gelangen; aber auch „sich mit etwas im Irrtum befinden“, „sich verrechnet haben“, „auf etwas bauen“, „überrascht sein“, „verblüfft sein“, „hoffen“, „befürchten“, „vertrauen“, „mißtrauen“, „bezüglich einer Sache pessimistisch eingestellt sein“, „die Hoffnung verlieren“ etc. gehören dazu.

Glaubensdispositionen sind also kognitive Dispositionen, erschöpfen sie aber nicht. So gibt es viele Fähigkeiten, die man als kognitive Fähigkeiten einstufen würde, die aber nicht Glaubensdispositionen im hier verwandten Sinne sind: z.B. Intelligenz, logisches Denkvermögen, Kreativität, Lernfähigkeit, Kombinationsgabe, Orientierungsvermögen oder Gedächtnisleistung, Eidetiker-Sein, Sprachverständnis etc.

Glaubensdispositionen bilden einen zentralen Gegenstand der Psychologie. So beschäftigen sich – jeweils unter anderen Aspekten – Wahrnehmungspsychologie, Lerntheorie, Denkpsychologie und ähnliches intensiv damit, wie sich Überzeugungen *bilden* und *verändern*; und für die *Bewahrung* von Überzeugungen interessieren sich z.B. wiederum die Lerntheorie und die Gedächtnisforschung.

Wünschensdispositionen dagegen sollen all das umfassen, was in der Psychologie gemeinhin unter den Begriff der Motivation fällt. Dieser „stets so umfassend wie möglich gemeinte“[3] Begriff hat in der Psychologie in letzter Zeit zunehmend an Bedeutung, aber nicht unbedingt an Klarheit gewonnen. Nach Thomae (1965), S. 19, lassen sich „wenigstens sechs Gruppen von materialen Bestimmungen des ... Motivationsbegriffs unterscheiden“, hinter denen jeweils verschiedene Theorien stehen. Seine allgemeine Definition, mit der er allen gerecht zu werden versucht, hilft da auch nicht viel weiter: „Motive bzw. Motivationsprozesse sind Abstraktionen aus dem Sinnzusammenhang der menschlichen Aktivität, die in ihrem (sinnvollen) Zusammenhang mit Veränderungen jener Aktivität in bezug auf Intensität, Richtung und Form gesehen werden.“ (Thomae (1965), S. 43.)

Sehen wir lieber von Definitionen und Theorien ab und versuchen wir stattdessen wieder, durch Betrachtung umgangssprachlicher Ausdrücke das fragliche Gebiet deutlicher zu machen. Zur Bezeichnung von Wünschensdispositionen dienen z.B.: „wollen“, „wünschen“, „mögen“, „sehr mögen“, „wenig mögen“, „begehren“, „beabsichtigen“, „intendieren“, „anstreben“, „planen“, „sich etwas vornehmen“, „auf etwas abzielen“, „ein Ziel verfolgen“, „eine Präferenz haben“, „etwas lieber mögen als etwas anderes“, „etwas eher wünschen als etwas anderes“, „vorziehen“, „unentschieden sein“, „herbeisehnen“, „schmachten“, „an etwas interessiert sein“, „nicht leiden können“, „verabscheuen“, „verurteilen“, „mißbilligen“, „gutheißen“, „eine Verpflichtung verspüren“, „sich zu etwas aufgerufen fühlen“, „für etwas eintreten“, „von Gott zu etwas berufen fühlen“, „vom Gewissen gesagt bekommen“, „ein Verlangen haben“, „getrieben werden“, „ein unwiderstehliches Bedürfnis verspüren“, „hoffen“, „fürchten“ und noch viele andere Wendungen . Natürlich ist es hier ähnlich wie bei den Wörtern für Glaubensdispositionen; manche Ausdrücke besagen mehr als bloß das Vorliegen einer Wünschensdisposition. Einige Ausdrücke deuten z.B. die Herkunft der Disposition an, in anderen, „hoffen“ z.B., kommt noch zum Ausdruck, daß der Wünschende glaubt, auf das Erwünschte nicht einwirken zu können. Es bestehen auch grammatische Unterschiede. Allen Wendungen ist aber gemeinsam, daß sie Wünschensdispositionen ausdrücken.

Unter Wünschensdispositionen fallen also sehr disparate Dinge: alle Sorten von Trieben, die die Psychologen unterscheiden mögen – Rohracher (1971), S. 409, unterscheidet z.B. Erhaltungstriebe wie den Geschlechts-, Nahrungs- und Schlaftrieb, Gesellschaftstriebe wie den Vergeltungstrieb, Machtgier oder die

---

[3] Thomae (1965), S. 19.

Triebgrundlagen des Mitleids, Genußtriebe wie das Rauchen und Schlemmen und Kulturtriebe wie den Wissensdrang oder den Gerechtigkeitstrieb –, aber auch politische und gesellschaftliche Zielvorstellungen, Werthaltungen und -vorstellungen, das Gewissen, die moralische Gesinnung, das Pflicht- und das Verantwortungsgefühl, persönliche oder private Ziele und Motive, kurzfristige Begierden und Gelüste, Vorlieben, Marotten, liebgewonnene Gewohnheiten, große und kleine Wünsche, Hoffnungen und Pläne etc. Mit diesen langen Aufzählungen dürfte, wie ich hoffe, einigermaßen klar geworden sein, welch weites Feld unter den dürren Ausdruck „Wünschensdisposition" fallen soll.

Daran, daß all die genannten Dinge bzw. Wörter Dispositionen bzw. Dispositionsbegriffe sein, kann kein Zweifel bestehen. Wenn jemand glaubt, von München nach Stuttgart seien es 500 km, so wird er, wenn er schnell von München nach Stuttgart gelangen will, dazu kein Taxi benutzen, wenn er erfährt, daß es nur 240 km sind, überrascht sein, wenn er danach gefragt wird, sagen, es seien 500 km etc. Umgekehrt wird er, wenn er die Entfernung im Atlas mißt und dabei auf 500 km kommt oder wenn er annimmt, der Eisenbahnkilometer koste 5 Pf, und DM 25,- für eine Eisenbahnfahrt von München nach Stuttgart bezahlen muß, glauben, von München nach Stuttgart seien es 500 km. Wer eine Prüfung bestehen will, wird fleißig lernen, wird Nachhilfe nehmen, wenn ihm seine bisherigen Lernerfolge nicht ausreichend erscheinen, versucht, sich mit seinem Prüfer gut zu stellen, wird vielleicht zu betrügerischen Methoden greifen, wenn er sonst keine Hoffnung mehr hat, etc.

Der Versuch, die möglichen Manifestationen solcher Dispositionen auch nur einigermaßen vollständig aufzulisten, muß freilich kläglich scheitern. Jede Glaubens- und Wünschensdisposition hat unzählige Manifestationen, wobei sich die Angelegenheit dadurch noch ungeheuer kompliziert, daß in den Manifestationsgesetzen solcher Dispositionen vielfach wieder andere Glaubens- und Wünschensdispositionen auftreten. Wir haben es also immer mit einem dicht gewebten Netz aus vielen Glaubens- und Wünschensdispositionen zu tun.

Es ist nun eine wichtige Sache, dieses Netz theoretisch irgendwie in den Griff zu bekommen. Denn nicht nur ist es selbstverständlich, daß unsere Glaubens- und Wünschensdispositionen dann von entscheidender Bedeutung für unser Verhalten sind, wenn wir ausgesprochen überlegt vorgehen. Ebenso unzweifelhaft ist auch, daß sie den meisten unserer gewohnheitsmäßigen und ohne Überlegungen zustande gekommenen Handlungen und Verhaltensweisen unterliegen. Wieso wende ich mich nach meiner Haustür nach links und gehe mit der Einkaufstasche in der Hand die Straße entlang, wie man eben normalerweise geht? Ich möchte mir etwas zu essen verschaffen. Dazu begebe ich mich am besten zum Kaufmann. Andere Methoden haben unangenehme Nebenwirkungen – z.B. die, beim Nachbarn einzudringen und seinen Kühlschrank zu leeren – oder geringe Durchschlagskraft – wie etwa die Methode zu beten, bis Manna vom Himmel fällt – oder beides – wie die Methode vom Fenster aus Tauben zu schießen. Daß ich zu Fuß bin, liegt daran, daß es umständlicher und zeitaufwendiger wäre, das Auto zu benutzen, daß ein Taxi nicht lohnt, daß ich ein Fahrrad nicht besitze, etc. Nach links wende ich mich, weil ich weiß, daß der nächste Weg zum Kaufmann diese Richtung nimmt und weil ich keine Lust habe, länger als nötig draußen herumzumarschieren. Daß ich so gehe, wie man eben geht, liegt z.B. daran, daß ich erwarte, daß ein Straßenpflaster nicht mit derselben Wucht zurückschlägt, wie ich es trete; daß ich vermute, daß ich bei hüpfender Bewegungsweise, die ohnehin anstrengender wäre, alle Leute seltsam anschauen würden, was mir unangenehm wäre; daß ich glaube, daß Rückwärts-Gehen ziemlich unfallträchtig wäre, etc. Daß ich eine Einkaufstasche bei mir habe, liegt daran, daß ich weiß, daß Plastiktüten neuerdings Geld kosten, und dieses Geld sparen möchte; daß der Transport der Waren, die ich zu kaufen wünsche, ohne Behältnis äußerst umständlich wäre, etc. Dies alles überlege ich mir natürlich nicht, bevor oder während ich zur Haustür hinaustrete, und manches habe ich mir nie überlegt. Nichtsdestoweniger habe ich diesen ganzen Wust an Glaubens- und Wünschensdispositionen, was sich unter anderem darin zeigt, daß ich mich nach meiner Haustür nach links wende und mit der Einkaufstasche in der Hand die Straße entlang gehe, wie man nun eben normalerweise geht. Glaubte ich etwa daß die Pflastersteine mit derselben Macht zurückschlagen, wie ich sie trete, so würde ich vermutlich vorsichtig die Straße entlang schleichen oder auch den Rückstoß der Steine zu einer federnd-springenden Fortbewegungsweise ausnutzen. Und wäre ich mir über das Verhalten der Steine bloß unsicher, so würde ich vielleicht erst einige tastende Testschritte machen. (In beiden Fällen würde ich natürlich ziemlich schnell zu der Überzeugung gelangen, daß die Pflastersteine nicht zurückschlagen.)

Glaubens- und Wünschensdisposition liegen also ungeheuer vielen unserer Handlungen zugrunde. Da fällt es vielmehr schwer, Beispiele von Verhaltensweisen zu finden, an denen sie nicht beteiligt sind. Dazu zählen alle Verhaltensweisen, die zwar kontrollierbar, aber normalerweise zu unbedeutend sind, als daß wir sie tatsächlich kontrollieren würden: z.B. die bekannten Kritzeleien während des Telefonierens, nervöse Bewegungen wie Kratzen, Wippen und Wackeln etc. Dann gehören natürlich alle nicht ohne weiteres kontrollierbaren Verhaltensweisen wie Schreckreaktionen, Schnarchen, Niesen etc. dazu. Aber danach wird es schon schwierig. Man ist vielleicht geneigt, auch stark gefühlsbetonte Handlungen dazu zu rechnen – sofern sie nicht gespielt sind und Gefühle nur vortäuschen sollen: Freudensprünge, Wut- und Haßausbrüche, depressive Anwandlungen, Stoßseufzer und dergleichen. Doch hängen Gefühle ganz eng mit den Glaubens- und vor allem mit den Wünschensdispositionen zusammen. Freuen werde ich mich z.B. dann, wenn etwas eintritt, was ich mir sehr gewünscht und kaum mehr erwartet habe. Enttäuscht werde ich sein, wenn es mir mißlungen ist, ein wichtiges Ziel zu erreichen. Aufgeregt bin ich vor und bei Handlungen, mit denen etwas sehr Wichtiges entschieden wird. Und so weiter. Auch den seltsamen Verhaltensweisen psychisch Kranker liegen vielfach Glaubens- und Wünschensdispositionen zugrunde, die nur oft so sehr vom Erwarteten abweichen, daß der Laie sie nicht mehr erkennen vermag..[4]

Glaubens- und Wünschensdispositionen bestimmen also unser Verhalten in einem Maße wie kaum eine andere, ja eigentlich wie keine andere Klasse von Dispositionen. Wenden wir uns ihnen nun eingehender zu. Welche verschiedenen Sorten von Glaubens- und Wünschensdispositionen lassen sich unterscheiden? Wenn ich diese Frage stelle, so will ich damit nicht fragen, wie man die eben erst hergestellte Allgemeinheit wieder inhaltlich etwa nach Herkunft und Gegenstand der Disposition untergliedern kann. Darauf könnte man reagieren, indem man z.B. Glaubensquellen wie direkte Beobachtung, Hörensagen, Erleuchtung und, was es da alles geben mag, unterscheidet, die Gegenstände des Glaubens nach Geographischem, Technischem, Politischem etc. aufgliedert, weitgehend angeborene Motivationen wie die meisten Triebe und weitgehend umweltbestimmte Motivationen wie Konsumwünsche oder auch gewisse Moralvorstellungen voneinander trennt und die verschiedenen Motivationen ihrem Gegenstand gemäß in politische, berufliche und persönliche Ziele etc. unterteilt. Doch geht es mir hier ja um eine logische Analyse, und für die bringen solche Unterscheidungen nicht viel.

Wenn ich also die obige Frage stelle, will ich rein begriffliche Sorten von Glaubens- und Wünschensdispositionen unterscheiden. So weit ich es sehe, kann man

[4] Vgl. Miller, Galanter, Pribram (1973), S. 110-114.

alle diese Dispositionen in jeweils vier Sorten aufteilen: in die – wie ich es nennen will – *qualitativen*, *klassifikatorischen*, *komparativen* und *quantitativen* Dispositionen.[5] Auf der qualitativen Stufe haben wir die Dispositionsausdrücke „*X* glaubt *A*“ und „*X* wünscht *A*“.[6] Um zu erläutern, was ich mit der klassifikatorischen Sorte meine, ist es zunächst nützlich festzuhalten, daß jeder Glaubensgegenstand *A* auf der qualitativen Ebene in drei verschiedene Kategorien eingeteilt werden kann: es gilt entweder „*X* glaubt *A*“ oder „*X* glaubt weder *A* noch nicht-*A*“ oder „*X* glaubt nicht-*A*“. Nun kennen wir in der Umgangssprache aber noch reichere Klassifizierungen, z.B. „für sicher, wahrscheinlich, möglich, unwahrscheinlich oder unmöglich halten“. Das gleiche gilt für den Motivationsbereich. Zur Erfassung solcher Klassifizierungen seien nun die klassifikatorischen Dispositionsausdrücke angesetzt, die man folgendermaßen allgemein formulieren könnte: „*A* gehört für *X* zur *m*-ten von *n* Glaubenskategorien“ und „*A* gehört für *X* zur *m*-ten von *n* Wünschenskategorien“. Was es mit der komparativen und der quantitativen Stufe auf sich hat, dürfte dagegen wieder klar sein: Da haben wir erstens „*X* glaubt eher *A* als *B*“ und „*X* wünscht eher *A* als *B*“ und zweitens „*X* glaubt *A* im Grade *r*“ und *X* wünscht *A* im Grade *r*“.

Nun zu behaupten, jede Glaubens- und Wünschensdisposition lasse sich auf eine der genannten zwei mal vier Formen bringen, ist in einem Sinne klar falsch, in einem anderen Sinne aber trotzdem aufrechtzuerhalten. Klar falsch ist dies Behauptung, wenn man die umgangssprachlichen Wendungen für Glaubens- und Wünschensdispositionen in der üblichen Weise versteht. „*X* fühlt sich zu *A* verpflichtet“ bzw. „*X* weiß *A*“ heißt eben etwas anderes als „*X* wünscht *A*“ bzw. *X* glaubt *A*“; „erwarten“ bzw. „hoffen“ heißt eben nicht bloß „glauben“ bzw. „wünschen“.

---

[5] In einer wichtigen Hinsicht ist die Aufteilung freilich unvollständig; es gibt noch die Sorte der – wie man es nennen könnte – Unabhängigkeits-Dispositionen, d.h. die Disposition, ein Ding hinsichtlich der eigenen Überzeugungen bzw. hinsichtlich der eigenen Wünsche als von einem anderen Ding unabhängig zu behandeln. Diese Dispositionen werden immer dann wirksam, wenn wir etwas außer acht lassen, als irrelevant ansehen, als nicht zur Sache gehörig betrachten, aus unseren Überlegungen ausschalten, von etwas abstrahieren, oder wie immer man das ausdrücken mag, und das tun wir ja dauernd. Dennoch will ich in diesem Kapitel nicht auf sie eingehen, da sie sich zum einen von den eben erwähnten vier Sorten erheblich unterscheiden, z.B. schon in ihren Gegenständen, und zum anderen in existierende Theorien über Glaubens- und Wünschensdispositionen der hier betrachteten Sorte (s. dazu Abschnitt 1.2) nur bei manchen Metrisierungen komparativer Dispositionen explizit auftauchen. Im dritten Kapitel werde ich ausführlich auf sie zu sprechen kommen.

[6] Mit Absicht habe ich die Formulierungen „*X* glaubt, daß *p*“ und „*X* wünscht, daß *p*“ vermieden, um nicht die auf den kommenden Seiten zu erörternde Frage nach der Natur der Glaubens- und Wünschensgegenstände zu präjudizieren.

Richtig ist aber die obige Behauptung, wenn wir eine geeignete Abstraktion vollziehen; wenn wir bei den umgangssprachlichen Wendungen für Glaubens- und Wünschensdispositionen von den Bedeutungskomponenten absehen, die dafür verantwortlich sind, daß diese Wendungen mehr besagen, als nur daß gewisse mehr oder weniger sichere empirische Überzeugungen oder gewisse mehr oder weniger starke Motivationen vorliegen; wenn wir also z.B. bei „*X* weiß *A*" davon absehen, daß *A* wahr ist, daß *X* eine irgendwie privilegierte Position hatte, um zu der Überzeugung, daß *A*, zu gelangen – und was da sonst in „*X* weiß *A*" über „*X* hält *A* für sicher" hinausgehen mag; oder wenn wir die Tatsache, daß *X* sich zu *A* verpflichtet fühlt, darauf reduzieren, daß *X* zu *A* motiviert ist, und uns nicht dafür interessieren, daß der Wunsch nach *A* eigentlich ein von außen an *X* herangetragener ist, dem sich *X* kaum entziehen zu können glaubt; etc.

Wir dürfen dann auch die acht vorgeschlagenen Wendungen nicht ganz im umgangssprachlichen Sinn verstehen; vielmehr müssen wir sie als die Destillate aus diesem Abstraktionsprozeß auffassen, die all den zuvor so unvollständig aufgelisteten umgangssprachlichen Wendungen gemeinsam sind. Dies ist insbesondere für unsere vier normierten Wendungen für den Motivationsbereich zu beachten: Im Gegensatz zum umgangsspachlichen „Wünschen", das zumindest bei Moralvorstellungen, Trieben und Verpflichtungen ein wenig angemessenes Wort ist, soll „Wünschen" im hier verstandenen Sinne den gesamten in den vorstehenden Erläuterungen umrissenen Motivationsbereich abdecken.

Im übrigen ist darauf hinzuweisen, daß der durch diese Abstraktion bedingte Verlust an Ausdrucksmöglichkeiten nicht so schwerwiegend ist, wie es zunächst scheinen mag. Viele Wendungen lassen sich nämlich innerhalb der eingeschränkten, vereinheitlichten Terminologie rekonstruieren. So bedeutet z.B. „*X* erwartet *A*" neben „*X* glaubt *A*" wohl nur noch, daß *A* bezüglich des Zeitpunktes, zu dem *X A* erwartet, in der Zukunft liegt, und „*X* hofft auf *A*" geht über „*X* wünscht *A*" in unserem normierten Sinne nur dadurch hinaus, daß es außerdem noch heißt, daß *X* glaubt, auf *A* nicht einwirken zu können.

Ich hoffe, damit die hier vollzogene Abstraktion klar und so die Behauptung, jede Glaubens- und Wünschensdisposition lasse sich auf eine der genannten acht Formen bringen, einsichtig gemacht zu haben. In allem Folgenden will ich mich nur mehr mit diesen acht Grundformen beschäftigen. Insbesondere sind künftig, wenn von Glaubens- und Wünschensdispositionen die Rede ist, immer diese acht

Grundformen gemeint, und „glauben" und „wünschen" sind immer als Produkt unseres Abstraktionsprozesses zu verstehen.

Wo wir schon die Dispositionswörter, die uns interessieren, vereinheitlicht haben, lohnt es sich auch, für unsere normierten Wendungen einige kurze Symbole einzuführen. Dazu ist zunächst darauf hinzuweisen, daß jede Glaubens- und Wünschensdisposition die Disposition eines Menschen ist[7] und immer zu einem gewissen Zeitpunkt besteht. Unsere normierten Wendungen müssen jeweils einen Namen für eine Person und – was bisher weniger deutlich zum Ausdruck kam – einen Namen für einen Zeitpunkt als Argumente annehmen. Es empfiehlt sich aber, diese Argumente als Indizes zu symbolisieren, die wir dann häufig wieder vernachlässigen können – dann nämlich, wenn nur von einer Person oder nur von einem Zeitpunkt die Rede ist. Damit haben wir auf der qualitativen Stufe:

„$G_{X,t}(A)$" für „X glaubt A zum Zeitpunkt t" und
„$W_{X,t}(A)$" für „X wünscht A zum Zeitpunkt t".

Die klassifikatorische Ebene sei folgendermaßen symbolisiert:

„$G^n_{X,t}(m,A)$" für „A gehört für X zum Zeitpunkt t zur m-ten von n Glaubenskategorien" und
„$W^n_{X,t}(m,A)$" für „A gehört für X zum Zeitpunkt t zur m-ten von n Wünschenskategorien".

Auf der komparativen Stufe sei eine kleine Abweichung von den früher vorgeschlagenen Wendungen vorgenommen:

„$A \preceq^g_{X,t} B$" für „X glaubt zum Zeitpunkt t A höchstens so sehr wie B" und
„$A \preceq^w_{X,t} B$" für „X wünscht zum Zeitpunkt t A höchstens so sehr wie B".

Der Übergang von der irreflexiven zur reflexiven Relation bringt eben einfach gewisse technische Vorteile mit sich. Schließlich haben wir auf der quantitativen Ebene zwei Funktionen

---

[7] Die nötige Qualifikation erfährt diese Behauptung auf S. 11f.

$P_{X,t}$ für den Glaubensgrad von X zum Zeitpunkt t und
$V_{X,t}$ für den Wünschensgrad von X zum Zeitpunkt t.

Diese beiden Funktionen nehmen reelle Zahlen als Werte an.[8]

Anzumerken ist hier zunächst, daß wir uns mit unseren acht Grundbegriffen natürlich in bekannten Regionen aufhalten. Der Leser wird dies längst bemerkt haben. So hat sich die epistemische Logik vorgenommen, die Logik des Prädikates „$G_{X,t}$“ für konstantes X und t zu untersuchen. Die deontische Logik beschäftigt sich mit der Logik von „$W_{X,t}$“ für konstantes X und t, sofern man ihr eine deskriptive Interpretation unterlegt.[9] $P_{X,t}$ wird in der Wahrscheinlichkeitstheorie formal untersucht; dabei ist natürlich von einer personalistischen Wahrscheinlichkeitsinterpretation auszugehen. In $\preceq^{g}_{X,t}$ erkennen wir wieder, was üblicherweise als qualitative Wahrscheinlichkeit bezeichnet wird („qualitativ „ wird dabei anders verwandt als es wir hier tun). In $\preceq^{w}_{X,t}$ sehen wir die Präferenzrelation, von der die Entscheidungstheorie redet, und $V_{X,t}$ ist nichts anderes als die Nutzenfunktion aus der Entscheidungstheorie. Nur „ $G^{n}_{X,t}$“ und „ $W^{n}_{X,t}$“ sind nicht Gegenstand einer prominenten Theorie, wenngleich auch sie da und dort schon Beachtung gefunden haben.[10]

Ferner sind einige Dinge zu den Personen- und Zeitindizes zu bemerken: Bezüglich der zugrunde liegenden Zeitstruktur können wir uns ganz offenhalten; die Variable „t“ kann für Zeitpunkte aus dem Zeitkontinuum stehen oder aber auch für Zeitintervalle aus einer diskreten Zeitstruktur. Das einzige, was wir benötigen werden, ist eine Ordnungsrelation auf dem Bereich, über den „t“ läuft.[11] Was die Variable „X“ anbelangt, so haben wir bisher so getan, als solle sie nur für Menschen stehen. Wir können aber den Bereich, über den „X“ läuft, durchaus weiter fassen. Wir können erstens auch Tiere dazu rechnen, soweit man ihnen Glaubens- und Wünschensdispositionen zuschreiben kann. Für die klassifikatorischen Dispositionen mag dies schwerfallen und für die quantitativen Dispositionen gar unmöglich erscheinen. Aber für die qualitativen und komparativen Dispositionen ist

---

[8] Der Vollständigkeit halber ist hier zu erwähnen, daß im Zusammenhang mit sogenannten lexikographischen Präferenzordnungen auch Wünschensgrade betrachtet wurden, die *Folgen* reeller Zahlen als Werte annehmen. Vgl. dazu etwa Chipman (1960). Ich will jedoch in dieser Arbeit auf diesen Sonderfall nicht weiter zu sprechen kommen.

[9] Wie es z.B. in Hansson (1971), S. 123, oder in Spohn (1973), S. 56f., geschieht.

[10] S. dazu S. 28.

[11] Im Abschnitt 3.1 kommen wir darauf noch etwas ausführlicher zu sprechen.

dies gewiß oft möglich. Und zweitens können wir – und das ist vielleicht wichtiger – auch Gruppen von Menschen (und eventuell auch Tieren) und alle möglichen Arten von Organisationen und Institutionen wie Unternehmen, Verbände, Parteien, Ämter, Staaten etc. dazu rechnen. Der Normalfall wird aber sein, daß X für einen einzelnen Menschen steht, und auf spezielle Probleme, die sich aus der Anwendung unserer acht Grundformen auf andere Gegenstände als Menschen ergeben, werde ich nicht eingehen.

In Vervollständigung der Klärung unserer acht Grundbegriffe fehlt uns nun noch ein entscheidender Punkt: Es gilt noch zu klären, worauf sich die Glaubens- und Wünschensdispositionen beziehen, für welche Entitäten das „A“ in „X glaubt A“ , in „X wünscht A“ etc. steht. In Ermangelung eines eingebürgerten Ausdrucks will ich die Menge von Entitäten, auf die sich einer unserer acht Grundbegriffe sinnvoll anwenden läßt, als seinen Definitionsbereich bezeichnen. Z.B. ist also der Definitionsbereich von $G^n_{X,t}$ die Menge aller A, für die $G^n_{X,t}(m,A)$ für ein $m \leq n$ ein sinnvoller, aber nicht unbedingt wahrer Ausdruck ist, bzw. der von $\preceq^w_{X,t}$ die Menge aller A und B, für die $A \preceq^w_{X,t} B$ ein sinnvoller Ausdruck ist; für Funktionen stimmt mein Sprachgebrauch ohnehin mit dem üblichen überein. Unsere Frage lautet dann: Wie sind die Definitionsbereiche unserer acht Grundbegriffe beschaffen?

Vielleicht gelangen wir zu einem Hinweis, wenn wir uns die genannten Theorien anschauen, die sich mit den verschiedenen Begriffen für Glaubens- und Wünschensdispositionen beschäftigen. Auf den ersten Blick ergibt sich da freilich ein recht uneinheitliches Bild: in der epistemischen Logik wird das von ihr untersuchte Prädikat „$G_{X,t}$“ immer auf Sätze angewandt. Dasselbe gilt auch weitgehend für die deontische Logik. Hier ist allerdings zu erwähnen, daß von Wright in seiner grundlegenden Arbeit (1951) – und mit ihm andere – den Geboten-Operator, der unserem „$W_{X,t}$“ entspricht, auf Namen von Handlungen anwandte – eine Konzeption, von der man wohl sagen kann, sie habe sich nicht durchgesetzt. In der Wahrscheinlichkeitstheorie sind die der Funktion $P_{X,t}$ entsprechenden Wahrscheinlichkeitsmaße immer für irgendwelche anonymen Mengen definiert, die dann Ereignisse getauft werden. Dieses Vorgehen hat auch auf die Behandlung der sogenannten qualitativen Wahrscheinlichkeit $\preceq^g_{X,t}$ abgefärbt. Doch auch hier gibt es eine Abweichung: Die Bestätigungsfunktionen aus Carnaps erster Version der induktiven Logik (1950) und (1952) sind für die Sätze einer freilich sehr einfachen, formalen Sprache definiert. Bei den entscheidungstheoretischen Begriffen wird das Bild noch bunter: In den meisten Versionen der Präferenzlogik, z.B. in von Wright

(1963) und in Rescher (1966), ist die Präferenzrelation, unser $\preceq^{w}_{X,t}$, eine Relation zwischen Sätzen. In der Entscheidungstheorie dagegen wird die Präferenzrelation üblicherweise als eine Relation zwischen Handlungen oder verallgemeinerten Handlungen wie Glücksspielen (s. etwa von Neumann, Morgenstern (1944), Abschnitte 3.5-3.7) oder sogenannten bedingten Entscheidungen (Krantz et al. (1971), Kap. 8) behandelt. Die Nutzenfunktion, unser $V_{X,t}$, ist normalerweise für Dinge definiert, die Konsequenzen genannt werden und die in formalen Darstellungen zu abstrakten mengentheoretischen Gegenständen degenerieren, in ökonomischen Anwendungen (z.B. bei Krelle (1968)) hingegen sich zu Güterbündeln oder Kompliziertem konkretisieren. Jeffrey (1965) schließlich sah in Propositionen die geeigneten Gegenstände für Präferenzrelationen und Nutzenfunktionen.

Doch die Vielfalt ist nicht so groß, wie es zunächst scheinen mag. Denn es lassen sich im wesentlichen zwei Typen von Theorien unterscheiden, die eine unterschiedliche Behandlung des jeweiligen Definitionsbereiches mit sich bringen. Den einen Typ bilden die zumeist von Philosophen und Logikern aufgestellten Theorien, in denen versucht wird, um den zu untersuchenden Begriff herum eine formale Sprache mit einer klaren Syntax, mit Axiomen und Ableitungsregeln und womöglich mit einer Semantik aufzubauen. Dazu gehören z.B. die deontische Logik, die epistemische Logik, die Präferenzlogik etc. Bei einem solchen Vorgehen muß der Definitionsbereich des Begriffs, für den die jeweilige Logik konstruiert wird, natürlich aus sprachlichen Entitäten bestehen, und wie wir gesehen haben, werden fast immer Sätze als die geeigneten sprachlichen Entitäten angesehen. Um Theorien des anderen Typs bemühen sich vorwiegend Mathematiker oder an einer sauberen mathematischen interessierte empirische Wissenschaftler. Diese Theorien sind dadurch gekennzeichnet, daß in ihnen ohne philosophische Skrupel vom mengentheoretischen Formalismus Gebrauch gemacht wird. Paradebeispiel dafür ist die Wahrscheinlichkeitstheorie, aber auch die Entscheidungstheorie wird man dazu rechnen können. Es ist klar, daß bei diesem Vorgehen der Definitionsbereich des jeweils untersuchten Begriffs nur aus mengentheoretischen Entitäten bestehen kann. Diese beiden Typen stehen nun nicht zusammenhanglos nebeneinander. Im Gegenteil, sie besitzen einen großen gemeinsamen Nenner: *Propositionen.*

Die Verbindung der sprachorientierten Theorien oder Logiken zu Propositionen ist leicht hergestellt. Diese Logiken sind nämlich zumeist intensional im strengen Sinne von Carnap (1947), S. 48; d.h. wenn ein Satz aus einem zweiten dadurch entsteht, daß man einen Teilsatz des zweiten Satzes durch einen mit diesem Teilsatz logisch äquivalenten Teilsatz ersetzt, so sind diese beiden Sätze ebenfalls logisch äquivalent. Dies gilt insbesondere auch für Ersetzungen im Bereich des jeweiligen Operators, sei es ein epistemischer, ein deontischer etc. An der Bedeutung dieser Operatoren änderte sich also nichts, wenn man sie statt auf Sätze auf Mengen logisch äquivalenter Sätze – welche man mit Fug und Recht Propositionen nennen könnte – anwendete.[12]

Und wie man die frei schwebenden Mengen(algebren), die etwa als Definitionsbereiche von Wahrscheinlichkeitsmaßen auftreten, an Sprachliches ankoppeln und als Propositionen interpretieren kann, hat uns Carnap (1971) gezeigt. Wie das genau funktioniert, will ich jetzt nicht schildern; im Abschnitt 3.1 kopiere ich Carnaps Verfahren für entscheidungstheoretische Zwecke.

Nun steht eigentlich nur noch die begriffliche Vielfalt in der Entscheidungstheorie dem bereits zu erahnenden Ergebnis unserer Diskussion entgegen. Doch hat Jeffrey (1965), insbesondere in den Abschnitten 4.1, 4.7 und 5.8 gezeigt, daß sich die Gegenstände der entscheidungstheoretischen Präferenzrelation und Nutzenfunktion immer als Propositionen auffassen lassen. So kann man etwa den Wunsch nach gewissen Gütern als den Wunsch ansehen, diese Güter zu besitzen. Und statt Handlungen kann man als Gegenstände der Präferenzrelation Propositionen nehmen, die durch Handlungsbeschreibungen ausgedrückt werden. Daß Propositionen geeignete Gegenstände für die entscheidungstheoretischen Grundbegriffe sind, zeigt sich auch im zweiten und dritten Kapitel, wenn wir auf die Entscheidungstheorie en detail eingehen.

Wir können also davon ausgehen, daß die Definitionsbereiche unserer acht Grundbegriffe aus Propositionen bestehen. Das heißt aber nicht, daß jeder Grundbegriff denselben Definitionsbereich haben müsse. So sollte z.B. der Definitionsbereich von $P_{X,t}$ zumindest eine Algebra von Propositionen sein. Dies scheint auch für die anderen Grundbegriffe für Glaubensdispositionen angemessen zu sein. Dagegen bestehen gute Gründe dafür, den Definitionsbereich der Nutzenfunktion $V_{X,t}$ nicht als Algebra zu konzipieren.[13] Doch will ich mich nun nicht in diese Feinheiten vertiefen. Mir kam es hier zunächst darauf an, den prinzipiellen Charakter der Glaubens- und Wünschensgegenstände herauszuarbeiten.

---

[12] Freilich läßt sich dies in einer formalen Sprache solange nicht realisieren, wie diese formale Sprache keine Ausdrücke für Mengen von Sätzen hat. Diesem Mangel hilft Montague (1970) ab; in der intensionalen Logik die er dort konstruiert, lassen sich sogar ohne Rückgriff auf Mengen von Sätzen Bezeichnungen für Propositionen definieren.

[13] S. dazu Abschnitt 3.4.

Freilich habe wir unser Ergebnis auf recht einseitige Weise erzielt. Wir haben bloß betrachtet, wie unsere Grundbegriffe in den verschiedenen um sie herum aufgebauten Theorien behandelt werden. Insofern dürften wir eigentlich nur sagen, daß wir mit diesen Theorien im Einklang stehen, wenn wir Propositionen als Glaubens- und Wünschensgegenstände nehmen. Aber auch dies ist schon ein erfreuliches Ergebnis. Außerdem verträgt es sich ganz gut mit den grammatischen Eigenschaften umgangssprachlicher Wendungen für Glaubens- und Wünschensdispositionen. Diese lassen sich nämlich immer so formulieren, daß sie sich auf einen Daß-Satz beziehen; und das auf Sätze anzuwendende Wort „daß“ erzeugt ja gerade so etwas wie Propositionen. Dieser Umstand ist natürlich der tiefere Grund dafür, daß sich bei den verschiedenen Theorien in dieser Hinsicht Übereinstimmung beobachten läßt.

Die Angelegenheit kann damit aber nicht als erledigt gelten. Wenn man unser Ergebnis einer genaueren philosophischen und linguistischen Analyse unterzieht – und genau das tut die mittlerweile recht umfangreiche philosophische Debatte um Propositionen und propositionale Einstellungen – so kommen doch beträchtliche Schwierigkeiten zum Vorschein.

Ein Problem hat sich mit besonderer Schärfe bei der epistemischen Logik gestellt, obwohl es zumindest für die anderen mit Glaubensdispositionen befaßten Theorien ebenso besteht: Wenn jemand glaubt, ein Satz sei wahr, so glaubt er auch von jedem mit diesem Satz logisch äquivalenten Satz, er sei wahr – in dieser Annahme der Intensionalität des epistemischen Operators lag, wie wir gesehen haben, die Berechtigung dafür, ihn als auf Propositionen anwendbar zu betrachten. Doch ist diese Annahme fragwürdig. Sie scheint vorauszusetzen, daß jedermann so etwas wie ein perfekter Logiker ist. Dies legt nahe, die Bedingungen für die gegenseitige Substituierbarkeit zweier Sätze im Bereich des epistemischen Operators noch zu verschärfen. Darüber, wie diese Verschärfung aussehen sollte, wurde bisher freilich keine Einigkeit erzielt.[14] Noch schwerere Geschütze hat Quine bei seinem Generalangriff auf den Begriff der Proposition aufgefahren.[15] Sollte man, wie Quine meint, den Begriff der Proposition tatsächlich ganz aufgeben müssen, so kämen all die genannten Theorien natürlich in schwere Bedrängnis.

Ich will mich nun dieser Probleme nicht annehmen; sie erforderten eine eigene sprachphilosophische Arbeit. Außerdem glaube ich, daß diese Probleme, so gravie-

---

[14] Vgl. dazu Blau (1969), S. 57, S. 62ff.

[15] Beginnen mit Quine (1951).

rend sie auch sein mögen. die propositional aufgezogenen Theorien über Glaubens- und Wünschensdispositionen nicht wirklich untergraben. Die Situation scheint mir nämlich ein wenig mit der im Grundlagenstreit der Mathematik bestehenden Situation vergleichbar zu sein: Es ist kein Geheimnis, daß es verschiedene schwere Bedenken gegen die Mengenlehre gibt. Der praktisch arbeitende Mathematiker macht aber, auch wenn er diese Bedenken kennt und ernst nimmt, normalerweise ohne Skrupel von der Mengenlehre, wenngleich nicht unbedingt in ihrer axiomatisierten Form, Gebrauch – steht er doch – mit gutem Recht, wie ich meine – auf dem Standpunkt, daß jeder Vorschlag zur Behebung dieser Probleme mit der Mengenlehre seine Ausdrucksmittel nicht wesentlich beschneiden dürfe, daß also jeder solche Vorschlag erst dann wirklich akzeptabel ist, wenn er gleichzeitig die Mathematik in ihrem Bestand nicht wesentlich angreift. Den gleichen Standpunkt kann man meines Erachtens auch hier einnehmen: Jeder wirklich ernst zu nehmende Vorschlag zur Lösung der Probleme mit dem Begriff der Proposition und mit den propositionalen Einstellungen muß hinreichenden Ersatz an Ausdrucksmöglichkeiten verschaffen; und unsere propositional konzipierten Theorien über Glaubens- und Wünschensdispositionen müssen sich ohne große Abstriche in die vorgeschlagene Ersatzterminologie übersetzen lassen. Die Berechtigung für diesen Standpunkt hängt natürlich wie beim mathematischen Parallelfall davon ab, wie fest die fraglichen Theorien in unserem wissenschaftlichen Gebäude verankert sind; und vielleicht überschätze ich, indem ich mich auf diesen Standpunkt stelle, ihren Grad der Verankerung. Jedenfalls hoffe ich, daß die folgenden Ausführungen, die von einem aus Propositionen bestehenden Definitionsbereich unserer Grundbegriffe ausgeht, von den Problemen mit dem Begriff der Proposition nicht in Mitleidenschaft gezogen werden.

1.2 Theorien über Glaubens- und Wünschensdispositionen: Ein Überblick

Nachdem wir unseren Untersuchungsgegenstand, die für das menschliche Verhalten so wichtigen Glaubens- und Wünschensdispositionen eingeführt und einigermaßen hinreichend geklärt haben, ist es nützlich, einen ganz groben Überblick über dieses Forschungsgebiet zu gewinnen, über seine Probleme und Teilgebiete, über existierende Theorien und Lücken. Natürlich werden wir auch die Entscheidungstheorie in diesem Gebiet lokalisieren.

Obwohl man annehmen sollte, daß dieses Forschungsgebiet zur Psychologie gehört, stammen die Beiträge dazu vor allem von Philosophen, Mathematikern und Ökonomen. Das heißt beileibe nicht, daß sich die Psychologen nicht mit Glaubens- und Wünschensdispositionen beschäftigt hätten; es ist nur so, daß sich die Psychologen mit ihnen kaum auf einer so allgemeinen und abstrakt-formalen Ebene auseinandersetzten wie die Theorien, um die es mir hier geht. Und in der Tat kann man über die Erfolgsaussichten so abstrakter Theorien in einem so vielschichtigen und vielgestaltigen Feld wie dem der Psychologie durchaus geteilter Meinung sein. Daß sie in der Psychologie aber zunehmend höher eingeschätzt werden, zeigt sich an dem Erstarken der noch ziemlich jungen mathematischen Psychologie, die sich als einziger Zweig der Psychologie um vergleichbaren Formulierungsstandards genügende Theorien bemüht.

Die Kehrseite der Medaille ist, daß diejenigen, die so abstrakte Theorien über Glaubens- und Wünschensdispositionen formulieren, wie wir sie hier betrachten wollen, selten auf empirische Theorien aus sind. Am ehesten tun dies noch die Ökonomen, die sich aber – zu Recht – schnell auf die Position zurückziehen, sie machten idealisierte Annahmen und formulierten idealisierte Modelle. Ein anderer beliebter Standpunkt – insbesondere von Entscheidungstheoretikern und Vertretern der personalistischen Wahrscheinlichkeitsinterpretation – besagt, daß in diesen Theorien Rationalitätspostulate aufgestellt würden, daß sie Theorien über rationales Verhalten seien. Der Impetus dieses Standpunktes ist nicht ganz klar; es kann bei ihm einerseits darum gehen, nur auf eine Idealisierung in einer empirisch gedachten Theorie hinzuweisen, andererseits darum, den Rationalitätsbegriff teilweise zu explizieren, und drittens um ein normatives Unterfangen, darum zu sagen, wie man sich am besten verhalten sollte. Die Philosophen und Logiker schließlich spüren der Logik der acht Grundbegriffe für Glaubens- und Wünschensdispositionen nach. Ihre Grundlage dafür ist ziemlich unklar. Häufig berufen sie sich auf ihre Intuition, die sich aber als wenig verläßlich erwiesen hat. Mehr Vertrauen genießt daher zur Zeit die sogenannte semantische Methode, die sich primär um eine Semantik für die jeweilige logische Sprache und erst in zweiter Linie um Axiome und um einen logischen Kalkül bemüht. Die Frage nach der Grundlage wird damit freilich nur um einen Schritt verschoben.[16] Die Philosophen haben in ihrem Streben nach Logik darüber hinaus die Tendenz, dem jeweils untersuchten Grundbegriff einen objektiven oder wenigstens intersubjektiven Anstrich zu geben. Da geht es z.B. nicht darum, für wie wahrscheinlich jemand eine gewisse Annahme aufgrund eines gewissen Erfahrungsdatums hält, sondern darum, wie stark dieses Erfahrungsdatum diese Annahme bestätigt.[17] Nicht was Herr Meier wünscht oder will, ist Gegenstand des Interesses, sondern das, was geboten ist.[18] So objektivierend formuliert, hat es die jeweilige Logik oder Theorie genaugenommen nicht mit Glaubens- und Wünschensdispositionen zu tun. Aber es fällt nicht schwer, sich diesen objektiven Zug wieder hinwegzudenken und so die jeweilige Theorie in eine über Glaubens- und Wünschensdispositionen zurückzuverwandeln. So wollen wir es auch in der kommenden Schilderung halten.

Von all diesen uneinheitlichen Tendenzen will ich nun absehen. Sie zu klären und einander anzupassen, ist natürlich von äußerster Wichtigkeit. Doch geht es mir in diesem Abschnitt nur darum, eine Landkarte von dem großen Feld der Theorien über Glaubens- und Wünschensdispositionen zu zeichnen und zu schauen, welche Flecken auf dieser Landkarte noch weiß sind und welche nicht. Mit welcher Absicht dieser oder jener Fleck ausgefüllt wurde, welcher Status der jeweiligen Theorie zugedacht war, spielt dabei zunächst keine Rolle.

Drei Unterscheidungen legen ein ziemlich feines Raster über dieses Feld. Die erste und wichtigste Unterscheidung haben wir schon im vorigen Abschnitt her-

---

[16] S. dazu Spohn (1973), S. 59f., S. 67ff.

[17] S. Carnap (1950), S. 19ff.

[18] So konzentriert sich z.B. v. Kutschera (1975) auf die normativen und nicht auf die subjektiven Begriffe. Aber er meint auch: „The concept of subjective preference, that a person (or a group of persons) prefers *A* to *B*, for instance, has the same logical structure as the normative notion, that (in a normative system) *A* is to be preferred to *B*. So the sematics and logic of the normative concept may ... also be employed for their subjective counterparts.“ (S. 196)

ausgearbeitet: die Aufteilung der Glaubens- und Wünschensdispositionen auf unsere zwei mal vier Grundtypen. Im folgenden werde ich eine Theorie, die mit den Grundbegriffen *A*, *B* und *C* arbeitet, als *ABC*-Theorie bezeichnen. Die epistemische Logik ist also eine *G*-Theorie, die Entscheidungstheorie eine $\preceq^{w}PV$-Theorie (den Personen- und den Zeitindex können wir uns dabei schenken). Kommt es nur auf die Zahl der verwandten Grundbegriffe an, so will ich von *eingleisigen*, *zweigleisigen* etc. Theorien reden.

Eine zweite wichtige Unterscheidung teilt die Theorien in *personale* und *interpersonale* Theorien ein. In den personalen Theorien dreht es sich immer um die Glaubens- und Wünschensdispositionen ein und desselben Menschen oder Tieres, ein und derselben Menschengruppe oder Organisation etc., wohingegen in interpersonalen Theorien die Zusammenhänge zwischen den Glaubens- und Wünschensdispositionen verschiedener Menschen etc. untersucht werden. Anders ausgedrückt: In den Sätzen einer personalen Theorie taucht immer ein und dasselbe *X*, für wen es auch stehen mag, als Personenindex ihrer Grundbegriffe auf, bei interpersonalen Theorien ist dies nicht der Fall.

Die dritte Unterscheidung schließlich ist die zwischen *statischen* und *dynamischen* Theorien. Statische Theorien untersuchen, wie Glaubens- und Wünschensdispositionen zu *einem* Zeitpunkt aussehen, während sich dynamische Theorien mit den Zusammenhängen zwischen den zu verschiedenen Zeitpunkten bestehenden Glaubens- und Wünschensdispositionen beschäftigen. In den Sätzen einer statischen Theorie ist somit der Zeitindex ihrer Grundbegriffe immer derselbe, dynamische Theorien machen sich von dieser Beschränkung frei.

Später werden wir diese Klassifizierung noch bereichern, doch läßt sich auch schon mit ihr eine recht detaillierte Landkarte zeichnen: Die umfassendste Theorie wäre natürlich eine achtgleisige, interpersonale, dynamische Theorie. Bedauerlicherweise schwebt sie derzeit in völlig unerreichbaren, utopischen Höhen. Immerhin existieren aber zu einer solchen Theorie schon viele Mosaiksteinchen, und die wollen wir nun in unsere Landkarte eintragen.

Fangen wir mit den am besten ausgebauten, weil relativ einfachsten Theorien an: den eingleisigen, personalen, statischen Theorien. Wir erwähnten sie schon im vorigen Abschnitt. Da haben wir als *G*-Theorie die epistemische Logik , als *W*-Theorie die deontische Logik, als $\preceq^{w}$-Theorie die Präferenzlogik und als *P*-Theorie die Wahrscheinlichkeitstheorie. Die Wahrscheinlichkeitstheorie dürfte davon und von allen noch zu erwähnenden die (als Theorie, nicht in ihrer Interpretation) unstrittigste und am weitesten ausgebaute sein. Aber auch in der deontischen Logik gibt es immerhin so etwas wie ein Standardsystem[19], und in der epistemischen Logik ist die Lage ähnlich, sofern man von den dort freilich besonders stark empfundenen Schwierigkeiten mit dem Begriff der Proposition absieht. In der Präferenzlogik läßt sich hingegen im Augenblick nur Nichtübereinstimmung konstatieren.[20] Ansonsten sind mir keine eingleisigen Theorien bekannt. $\preceq^g$ taucht eigentlich nur im Zusammenhang mit der Metrisierungsproblematik auf, und dort liegt auch das Hauptvorkommen von $\preceq^w$ (wir werden darauf gleich noch zu sprechen kommen). Eine eingleisige *V*-Theorie gibt es ebenfalls nicht. Selbst in der reinen Nutzentheorie, in der Wahrscheinlichkeiten noch keine Rolle spielen, wird die Nutzenfunktion V immer im Zusammenhang mit der Präferenzrelation $\preceq^w$ diskutiert. Schließlich sind auch für die überhaupt seltener zu findenden klassifikatorischen Begriffe eingleisige Theorien unbekannt. Das soll nicht bedauernd klingen. Man braucht nicht für jeden Grundbegriff eine eingleisige Theorie aufzumachen. Man kann einen Grundbegriff ebensogut für theoretisch entbehrlich halten. Begründen läßt sich eine solche Meinung aber nur, wenn man diesen Begriff auf einen anderen (oder auch auf mehrere andere) reduziert. Und dies geht nur in einer mehrgleisigen Theorie, die sowohl den reduzierten Begriff als auch die Begriffe, auf die er reduziert wird, enthält.

Damit sind wir schon beim nächsten Thema, den *mehrgleisigen*, *personalen*, *statischen* Theorien. Es empfiehlt sich hier, das den vielfältigeren Möglichkeiten entsprechende, größere Angebot etwas aufzugliedern: und zwar zu einen in *reduktionistische Theorien* – von ihnen war gerade die Rede –, deren Ziel es ist, einen oder mehrere Begriffe auf ihre anderen Grundbegriffe zurückzuführen, und *kooperative* Theorien, wie man die nicht reduktionistischen Theorien nennen könnte; und zum anderen in *unilaterale* Theorien, die nur Begriffe für Glaubensdispositionen oder nur Begriffe für Wünschensdispositionen enthalten, und *bilaterale* Theorien, die sich sowohl mit Glaubens- wie mit Wünschensdispositionen beschäftigen. Man sollte denken, daß reduktionistische Theorien immer unilateral sind, denn wie sollte man Begriffe für Glaubensdispositionen auf Begriffe für Wünschensdispositionen reduzieren oder die Umkehrung davon bewerkstelligen können? Dennoch gibt es überraschenderweise eine Ausnahme; sie wird auf S. 24 zur Sprache kommen.

---

[19] Diese Bezeichnung führte Hansson (1971) ein.

[20] Vgl. etwa die Tabelle in Rescher (1966), S. 53.

Reduktionen können auf verschiedene Weisen durchgeführt werden. Da gibt es einmal die *Reduktion durch Definition*. So liegt es z.B. sehr nahe, die klassifikatorischen Begriffe als vergröberte Versionen der quantitativen Begriffe aufzufassen; man könnte etwa „$G^5(1,A)$" als „$0 \leq P(A) \leq 0{,}2$" und ... und „$G^5(5,A)$" als „$0{,}8 < P(A) \leq 1$" definieren. Ebenso naheliegend ist die Gleichsetzung eines klassifikatorischen Begriffs mit dem entsprechenden, nur endlich viele Grade unterscheidenden komparativen Begriff. Genauer: Zum einen könnte man etwa „$A \preceq^w B$" als „es gibt ein i und ein j mit $W^n(i,A)$, $W^n(j,B)$ und $i \leq j$" definieren. Das umgekehrte Vorgehen sieht so aus: Definiert man „$A \approx^w B$" als „$A \preceq^w B$ und $B \preceq^w A$", so ist $\approx^w$ eine Äquivalenzrelation. Zerlegt nun $\approx^w$ den Definitionsbereich von $\preceq^w$ in endlich viele Äquivalenzklassen $\mathcal{A}_1,...,\mathcal{A}_n$ — die Indizierung sei dabei so gewählt, daß für $A \in \mathcal{A}_i$ und $B \in \mathcal{A}_j$ genau dann $A \preceq^w B$, wenn $i \leq j$ –, so läßt sich „$W^n(j,A)$" als „$A \in \mathcal{A}_j$" definieren. In diesen ziemlich trivialen Reduktionsmöglichkeiten dürfte der Grund dafür liegen, daß die klassifikatorischen Begriffe so wenig Aufmerksamkeit erhalten haben. Eine gründlichere Untersuchung des Sachverhalts wäre aber vielleicht doch wünschenswert. Die trivialste definitorische Reduktion ist freilich die Definition der komparativen Begriffe durch die entsprechenden quantitativen Begriffe, z.B. von „$A \preceq^g B$" als „$P(A) \leq P(B)$". Ein interessanteres Beispiel liefert schließlich noch das Bemühen um die Formulierung induktiver Akzeptationsregeln, das sich als ein Versuch zur definitorischen Reduktion des qualitativen Glaubens auf den quantitativen Glauben deuten läßt.[21]

Als zweites gibt es die *Reduktion durch Konstruktion*, wie ich es einmal nennen möchte. Zu dieser Reduktionssorte gehören alle Metrisierungsresultate, deren Konzept immer folgendermaßen aussieht. Ausgegangen wird von einer gewissen Axiomen genügenden schwachen Ordnung[22] – z.B. $\preceq^g$, (und eventuell von weiteren qualitativen Begriffen wie z.B. einer wahrscheinlichkeitstheoretischen Unabhängigkeitsrelation) – und dann wird bewiesen, daß es eine bis auf gewisse Transformationen eindeutig bestimmte Funktion gibt, die mit der schwachen Ordnung übereinstimmt – in unserem Beispiel ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P, für das $P(A) \leq P(B)$ genau dann gilt, wenn $A \preceq^g B$. Diese Metrisierungen liefern also immer eine Reduktion der quantitativen Ebene auf die komparative Ebene. Beispiele dafür sind die Konstruktion eines Wahrscheinlichkeitsmaßes P

---

[21] S. dazu etwa Hintikka, Hilpinen (1966).

[22] Eine schwache Ordnung ist eine transitive und konnexe Relation.

aus der „qualitativen“ Wahrscheinlichkeit $\preceq^g$ [23] und die vielen Metrisierungsresultate der reinen Nutzentheorie, in denen eine Nutzenfunktion V aus einer Präferenzrelation $\preceq^w$ konstruiert wird.[24] Auf die entscheidungstheoretischen Metrisierungsresultate gehen wir gleich noch ein.

Die letzte Art von Reduktion ist die durch *semantische Festlegung*. Sie ergibt sich, neuerdings immer häufiger, bei dem für Philosophen und Logiker charakteristischen sprachlogischen Vorgehen. Ein typisches Beispiel dafür ist Hanssons deontische Logik (1971), bei der er den deontischen Operator der von ihm konstruierten logischen Sprache mittels einer Präferenzrelation semantisch interpretiert. Er liefert damit – in unserer Terminologie – eine Reduktion von W auf $\preceq^w$. Vorläufig nur erwähnen will ich, daß sich Rescher (1964) meines Erachtens als eine Reduktion von G auf G^n interpretieren läßt, was insofern interessant ist, als hier ein klassifikatorischer Begriff einmal eine grundlegende Rolle spielt. Eine Begründung für diese Behauptung will ich nachliefern, wenn wir die dynamischen Theorien besprechen. Die Beispiele für Reduktionen durch semantische Festlegung ließen sich noch weiter fortsetzen[25], doch glaube ich, bereits vollständig genannt zu haben, welcher Begriff auf welche zu reduzieren versucht wurde, wenngleich ich natürlich nicht darstellen habe – und auch nicht darstellen wollte –, wie all diese Reduktionsversuche genau aussehen.

Bemerkenswert ist hier vielleicht, daß sich bei diesen Reduktionen die komparativen Begriffe jeweils als die fundamentalsten zu erweisen scheinen. Man könnte freilich in Anbetracht der trivialen Reduktion der komparativen auf die quantitativen Begriffe auch die letzteren als die grundlegendsten empfinden. Diese – sicherlich interessanten – Prioritätsfragen will ich aber nicht diskutieren. Stattdessen will ich noch kurz darauf hinweisen, daß die drei genannten Reduktionsarten nicht so verschieden sind, wie es zunächst scheinen mag. Welche gewählt wird, hängt im wesentlichen nur vom sprachlichen Rahmen ab, in dem man sich bewegt. So ermächtigen die Metrisierungsresultate aufgrund ihrer Eindeutigkeitsbehauptung dazu, die durch Metrisierung erhaltene Funktion als *die* Funktion zu definieren, die mit der metrisierten schwachen Ordnung übereinstimmt.[26] Und die Logiker müssen sich

---

[23] S. etwa Krantz et. al. (1971), Kap. 5.

[24] Einen sorgfältigen Überblick darüber gibt Fishburn (1970).

[25] Z.B. mit Gärdenfors (1975), der eine Logik und mit Hilfe von Wahrscheinlichkeitsmaßen eine Semantik für die qualitative Wahrscheinlichkeit formuliert.

[26] Ist die Funktion etwa nur bis auf positiv lineare Transformationen eindeutig bestimmt, so muß man natürlich noch zwei verschiedene Werte der Funktion vorgeben.

nur deswegen auf eine Reduktion durch semantische Festlegung verlegen, weil in der jeweiligen logischen Sprache, die sie untersuchen, nur der reduzierte Begriff, nicht aber der Begriff, auf den reduziert wird, enthalten ist. Wären sie nicht auf eine – natürlich nützliche – Logik für den reduzierten Begriff aus und würden sie stattdessen die Theorie für die sie interessierenden Begriffe mengentheoretisch, analog etwa zur Wahrscheinlichkeitstheorie, aufziehen, so würde die Reduktion durch semantische Festlegung ebenfalls zur definitorischen Reduktion.

Wenden wir uns nun den *kooperativen Theorien* zu. Hier verblüfft zunächst der Umstand, daß es anscheinend keine unilateralen kooperativen Theorien gibt. Doch ist a priori nicht einzusehen, wieso nicht auch eine Theorie sinnvoll sein sollte, die ohne Reduktionsabsicht mit mehreren Begriffen für Glaubens- und Wünschensdispositionen arbeitet. Das einzig stichhaltige Argument dagegen bestünde darin, daß jeweils einer der Begriffe für Glaubens- und Wünschensdispositionen fundamental ist, die anderen also auf ihn reduzierbar sind. Zur konkreten Ausgestaltung dieses Arguments gibt es aber, wie wir gerade gesehen haben, bisher nur Ansätze.

Nach den kooperativen Theorien müssen wir uns also schon unter den bilateralen Theorien umsehen. Aber auch da ist das Angebot mager. In der Tat ist die Entscheidungstheorie die einzig existierende bilaterale und personale Theorie. Dies ist umso bedauerlicher, als doch gerade das Zusammenspiel von Glaubens- und Wünschensdispositionen ein besonders aufregender Untersuchungsgegenstand ist. Insbesondere ist es enttäuschend, daß keine namhaften Bemühungen um eine gemeinsame Theorie des qualitativen Glaubens und Wünschens zu beobachten sind. Es gibt zwar kombinierte Modal- und (nicht deskriptive) deontische Logiken – zur Untersuchung des „ought-can"-Zusammenhangs[27] –, und vielleicht könnte man diese in ihr quasi menschlicheres Gegenstück, in eine personenbezogene, kombinierte epistemische und deskriptiv-deontische Logik umdeuten. Ob bei dieser Umdeutung etwas Vernünftiges herauskommt, ist freilich fraglich.

Es ist nun nicht ganz korrekt, die Entscheidungstheorie ausschließlich als kooperativ zu charakterisieren. Soweit sie eine nur auf der quantitativen Ebene angesiedelte *PV*-Theorie ist, ist sie sicherlich kooperativ. Doch nimmt die Metrisierungsproblematik in der entscheidungstheoretischen Literatur einen breiten Raum ein, und in diesem Kontext stellt sich die Entscheidungstheorie als eine reduktionistische $\preceq^w$ *PV*-Theorie dar, in der sowohl P als auch V auf $\preceq^w$ (durch Konstrukti-

[27] Z.B. Fisher (1962).

on) reduziert werden. (Eine grobe Formulierung solcher Metrisierungstheoreme gebe ich in Abschnitt 2.8.)

Innerhalb der Entscheidungstheorie sind solche Metrisierungen eine natürliche Angelegenheit, zu unserer Aufteilung von Glaubens- und Wünschensdispositionen scheinen sie aber denkbar schlecht zu passen. Zwar habe ich nirgends explizit gesagt, daß Glaubens- und Wünschensdispositionen voneinander unabhängig seien und erst in ihren gemeinsamen Manifestationen zusammenliefen; aber schien nicht gerade darin die raison d'àtre unserer Klassifizierung zu liegen? Ist also die entscheidungstheoretische Präferenzrelation ein aus unserem Rahmen herausfallendes Zwitter? Ich glaube nicht. Daß Glaubensdispositionen schon in die Wünschensdispositionen selbst und nicht erst in ihre Manifestationen eingehen, gilt nicht nur für komparative, sondern für alle Wünschensdispositionen; daß es für die Nutzenfunktion gilt, wird in den Abschnitten 2.3 und 3.6 ganz deutlich, und daß es dann auch für die qualitativen und klassifikatorischen Wünschensdispositionen zutrifft, ist aufgrund der erwähnten Reduktionsmöglichkeiten zumindest plausibel. Insofern haben wir es hier also mit einem ganz allgemeinen Phänomen zu tun, das gerade darum unsere Klassifikation der Dispositionen nicht zu erschüttern braucht, das uns aber dazu zwingt, die Vorstellung, Wünschensdispositionen seien von Glaubensdispositionen unabhängig, zu revidieren.

Wie es zu diesem Phänomen kommt und wie man sich das Verhältnis von Wünschens- und Glaubensdispositionen genau vorstellen soll, will ich nun nicht klären. In Abschnitt 3.6, S. 150-152, werde ich mich diesen Fragen etwas gründlicher zuwenden. Vorläufig sei nur soviel angedeutet: Die Begriffe für Wünschensdispositionen scheinen mir einer systematischen Vieldeutigkeit zu unterliegen – einer Vieldeutigkeit, die sich in der reichlich grobschlächtigen philosophischen Unterscheidung zwischen „gut an sich" und „gut als Mittel" und auch in solchen Äußerungen wie „eigentlich ist mir ja das da lieber, aber insgesamt betrachtet sollte ich vielleicht doch dies hier nehmen" widerspiegelt und die, grob gesagt, daher rührt, daß sich die Glaubensdispositionen in unterschiedlichem Maße in den Wünschensdispositionen berücksichtigen lassen.

Damit sei der Überblick über die personalen, statischen Theorien abgeschlossen. Ein noch wesentlich breiteres Spektrum an Möglichkeiten eröffnet sich in den übrigen Bereichen. Dieses Spektrum irgendwelchen ausgefeilten Klassifikationen zu unterwerfen, lohnt sich aber nicht so recht; dazu ist es an zu wenigen Stellen von konkreten Theorien besetzt. Deshalb will ich mich nun im wesentlichen darauf beschränken, die Theorien, die es da gibt, kurz anzusprechen.

Betrachten wir zunächst die *interpersonalen, statischen Theorien*. Zuallererst ist hier natürlich die Spieltheorie zu erwähnen. Welche unserer acht Grundbegriffe in die Spieltheorie eingehen, ist freilich nicht ganz klar: Auf alle Fälle benötigen wir die Nutzenfunktionen der Spieler. In den früheren, von einer objektivistischen Wahrscheinlichkeitsinterpretation geprägten Darstellungen der Spieltheorie (z.B. in von Neumann und Morgenstern (1944)) wurde ferner das Wissen der Spieler um die objektiven, statistischen Wahrscheinlichkeiten der Zufallszüge vorausgesetzt, d.h. es ist hier vom qualitativen Glauben der Spieler die Rede. Bei einer für die Spieltheorie adäquateren personalistischen Wahrscheinlichkeitsinterpretation ist dagegen der quantitative Glauben der einzelnen Spieler relevant. Es gehen aber noch etliche weitere Voraussetzungen in die Spieltheorie ein, wie bei Luce und Raiffa (1957), Kap. 3, besonders deutlich wird. So wird z.B. das Wissen der Spieler um die Spielregeln und um die Nutzenfunktion der anderen Spieler vorausgesetzt. Formuliert man dies mit Hilfe unserer Grundbegriffe, so erweist sich die Spieltheorie also zumindest als eine *GPV*-Theorie.

Ein anderes prominentes Problem, das eine interpersonale Theorie erfordert, aber nicht zur Spieltheorie gehört, ist das Problem der „social choice“ oder der „group preferences“. Hier geht es darum, die Präferenzrelationen oder die Nutzenfunktionen der Personen $X_1,...,X_m$ zu einer Präferenzrelation der Gruppe $(X_1,..., X_m)$ zu verschmelzen, was bisher auf große Schwierigkeiten stieß und nur in Spezialfällen einer Lösung zugeführt wurde.[28] Zu erwähnen ist vielleicht auch, daß Savage in der Statistik eine Wissenschaft sah, der interpersonale Probleme ganz entscheidend zugrunde liegen.[29]

Damit bin ich auch schon am Ende der Aufzählung angelangt. Andere einschlägige interpersonale Theorien gibt es meines Wissens nicht. Wiederum will ich diesen Zustand nicht beklagen. Bei vielen Möglichkeiten, die wir hier am grünen Tisch an Hand unserer Unterscheidungen und Aufteilungen zusammenstellen können, ist denkbar unsicher, ob sie sich überhaupt durch irgendwelche sinnvolle Theorien auffüllen lassen. Und solange dies nicht nachgeprüft ist, gibt es auch nichts

---

[28] Vgl. dazu etwa Luce, Raiffa (1957), Kap. 14.

[29] So sagt Savage (1954), S. 154: „From the personalistic point of view, *statistics proper* can perhaps be defined as the art of dealing with vagueness and with interpersonal difference in decision situations.“

zu bedauern. Im übrigen ist unsere Bilanz an interpersonalen Theorien nicht so schlecht, wie sie aussieht. Denn es ist nicht zu vergessen, daß die Spieltheorie seit ihrem Bestehen zu einem äußerst umfangreichen und immens verzweigten Moloch geworden ist, der sich der vielfältigsten interpersonalen Problemstellungen annimmt, auf die ich hier weder eingehen kann noch will.

So fehlen nur noch die *dynamischen* Theorien: Da anzunehmen ist, daß Glaubens- und Wünschensdispositionen einer zeitlichen Änderung unterliegen, liegt zunächst die Frage nahe, woher solche Änderungen herrühren könnten. Darauf gibt es die verschiedensten Antworten. Viele Änderungen liegen schon im angeborenen Reifungs- und Alterungsprozeß begründet. So lassen z.B. die motorischen und sexuellen Bedürfnisse im Alter in der Regel nach. Gedächtnisschwund ist ein anderer Faktor mit oft unangenehmen Auswirkungen auf die Glaubensdispositionen. Und wenn man von solchen biologischen und physiologischen Ursachen absieht, so liegt der allgemeinste Grund für Änderungen der Glaubens- und Wünschensdispositionen darin, daß man etwas lernt. In der Tat wird ja in der Lerntheorie Lernen praktisch als Verhaltensänderung definiert,[30] und „Verhaltensänderung“ darf man hier ruhig als „Veränderung von Verhaltensdispositionen“ lesen. Insofern sind Lerntheorien also dynamische Theorien. Freilich passen sie, da meist in Stimulus-Response-Terminologie abgefaßt, nicht so ohne weiteres in unser Begriffsschema.

Eine verwandte Quelle für Änderungen von Glaubens- und Wünschensdispositionen sind Beobachtungen oder Erfahrungen, und nur mit dieser Quelle werden wir es im folgenden zu tun haben. Denn durchweg alle existierenden dynamischen Versionen der bisher angesprochenen statischen Theorien beschäftigen sich mit der aus Beobachtungen oder Erfahrungen resultierenden Dynamik, die freilich eine speziellere Form der Dynamik ist als die, die von der Lerntheorie untersucht wird. Die wichtigste Spezialisierung liegt in der Vernachlässigung des Gedächtnisses. Während die Lerntheorie, auf empirische Adäquatheit bedacht, sich intensiv mit der Rolle des Gedächtnisses bei Lernprozessen auseinandersetzt, gehen die nun zu schildernden Theorien implizit oder explizit von einem perfekten Gedächtnis aus, also davon, daß die gemachten Beobachtungen sich unauslöschlich einprägen und

---

[30] Hilgard, Bower (1975) geben die folgende provisorische Definition: „Learning refers to the change in a subject’s behavior to a given situation brought about by his repeated experience in that situation, provided that the behavior change cannot be explained on the basis of native response tendencies, maturation, or temporary states of the subject (e.g. fatigue, drugs, etc.).“ (S. 17)

nicht wieder in Vergessenheit geraten können. Aber auch in anderen Hinsichten ist diese Form der Dynamik beschränkt; so berücksichtigt sie z.B. nicht die Rolle der Begriffsbildung.[31] Trotzdem sind natürlich alle Untersuchungen zu dieser eingeschränkten Dynamik wertvoll; es ist ja durchaus gute Forschungsstrategie, erst einmal diese einfache, wenn auch immer noch genügend komplizierte Form der Dynamik zu klären, bevor man sich an realistischere und damit komplexere Formen heranwagt.

Das einfachste Beispiel für eine dynamische Theorie dürfte das wohlbekannte Prinzip der Konditionalisierung in der Wahrscheinlichkeitstheorie sein, das folgendes besagt: Wenn P_t das subjektive Wahrscheinlichkeitsmaß von X zum Zeitpunkt t ist und wenn X zwischen t und t' die Erfahrung A sammelt, so gibt das durch A bedingte Maß $P_t(.|A)$ die subjektiven Wahrscheinlichkeiten von X zum Zeitpunkt t' an. Auf dieses Prinzip und seine Verallgemeinerungen werden wir im Abschnitt 4.2 noch ausführlicher zu sprechen kommen. Zur dynamischen P-Theorie gehören auch die von den Vertretern einer personalistischen Interpretation zur Stützung ihrer Interpretation bewiesenen Resultate, daß sich die anfänglich möglicherweise recht unterschiedlichen subjektiven Glaubensgrade verschiedener Leute mit wachsender Erfahrung den beobachteten relativen Häufigkeiten und somit auch einander immer mehr angleichen.[32] Überhaupt lassen sich große Teile der Theorie der stochastischen Prozesse als dynamische P-Theorie auffassen, soweit man die dort auftretenden Wahrscheinlichkeitsmaße sinnvoll als subjektive Glaubensgrade interpretieren kann.

Ferner sind hier gewisse G- und W-Theorien zu erwähnen, die zwar wohl nicht direkt als dynamische Theorien gedacht waren – so ist in ihnen nirgends von Zeitpunkten die Rede –, die sich aber doch so interpretieren lassen. Die W-Theorie, an die ich dabei denke, ist die dyadische deontische Logik. Sie beschäftigt sich mit dem Problem der „contrary-to-duty imperatives", d.h. mit der Frage, welche Gebote bzw. Wünsche bestehen, nachdem die ursprünglichen Gebote bzw. Wünsche nicht erfüllt worden sind. Man erkannte, daß sich diese Frage mit einem einstelligen deontischen Operator nicht adäquat behandeln läßt, und führte daher einen zweistelligen deontischen Operator ein, zu lesen als „unter den Umständen A ist B geboten" bzw. in unserer personenbezogenen Terminologie als „unter den Um-

---

[31] Ein Punkt, der Suppes (1966) besonders am Herzen liegt.

[32] S. etwa de Finetti (1964), Kap. V.

ständen A wünscht X B“.[33] Für diesen bedingten deontischen Operator kann man ebenfalls ein Prinzip der Konditionalisierung formulieren: Ist W_t bzw. $W_{t'}$ das zum Zeitpunkt t bzw. t' auf X zutreffende qualitative Wünschensprädikat und sammelt X zwischen t und t' die Erfahrung A, so gilt $W_{t'}(B)$ genau dann, wenn $W_t(B \mid A)$. Darin liegt die gewünschte dynamische Interpretation der dyadischen deontischen Logik. Sie läßt sich aber in der Logik selbst nicht ohne weiteres explizit machen. Dazu müßte man in der fraglichen logischen Sprache zumindest auch über Zeitpunkte reden können.[34]

Ähnlich wie das Problem der „contrary-to-duty imperatives“ Anlaß für eine dynamisch interpretierbare W-Theorie gab, forderte das Problem der kontrafaktischen Konditionale eine dynamische G-Theorie heraus; zumindest scheint man mir die auf S. 22 ohne Begründung als reduktionistisch ausgegebene Theorie von Rescher (1964) so interpretieren zu können. Rescher führt darin das Problem der kontrafaktischen Konditionale zunächst auf das Problem der „belief-contravening hypothesis“ zurück: auf das Problem, wie ein Glaubenskorpus – eine Menge von geglaubten Aussagen –, der durch Hinzunahme einer neuen Annahme inkonsistent geworden ist, zu revidieren sei derart, daß der revidierte Korpus die neue Annahme enthält und wieder konsistent ist. Rescher schlägt dann ein solches Revisionsverfahren vor, das darauf basiert, daß die Menge aller Aussagen auf eine endliche Anzahl sogenannter Modalkategorien verteilt wird. Wie es genau funktioniert und wie sinnvoll es ist, braucht uns jetzt nicht zu interessieren. Wichtig ist nur, daß sich diese Modalkategorien, wie Rescher selbst deutlich macht ((1964), S. 38), epistemisch, d.h. gerade als unser klassifikatorisches Prädikat G^n interpretieren lassen – daher meine Behauptung, Rescher reduziere G auf G^n – und daß sich die ganze Geschichte auf folgende Weise dynamisch interpretieren läßt: Sei G_t bzw. $G_{t'}$ das zum Zeitpunkt t bzw. t' auf X zutreffende Glaubensprädikat, d.h. sei die Extension von G_t bzw. $G_{t'}$ der Glaubenskorpus von X zu t bzw. t', und mache X zwischen t und t' die Erfahrung A. Ist A mit der Extension von G_t konsistent – der triviale Fall –, so gilt neben $G_{t'}(B)$ für alle B mit $G_t(B)$ auch noch $G_{t'}(A)$ mit allen Folgerungen, die sich daraus ergeben mögen. Ist A dagegen mit der Extension von G_t inkonsistent – der nicht triviale Fall –, so liefert uns gerade Reschers Revisionsverfahren die

[33] Vgl. Dazu Hansson (1971).

[34] Blau (1974) ist sogar der Ansicht, daß sich ohne Einbeziehung des Zeitpunktes, zu dem ein Gebot gilt, ohnehin keine angemessene deontische Logik formulieren läßt.

Extension von $G_{t'}$. Dies ist freilich wiederum eine von außen herangetragene Interpretation, die sich nicht so einfach in die Logik selbst einarbeiten läßt.[35]

Diesen drei Beispielen für dynamische Theorien scheint das gleiche, auch für andere dynamische Theorien nutzbare Schema zugrunde zu liegen: ein Schema, das allerdings – wie der Leser schon bemerkt haben wird – von einer gewissen Erweiterung unseres Bestandes an Grundbegriffen ausgeht. Nicht, daß es wesentlich neue Begriffe verwendete. Doch benötigt es von unseren acht Grundbegriffen noch jeweils *bedingte Versionen*: „unter Umständen *A* wünscht *X B*", „unter Umständen *A* glaubt *X* eher *B* als *C*", „unter Umständen *A* wünscht *X B* im Grade *r*", etc. Dabei können wir wiederum ohne Beschränkung der Allgemeinheit annehmen, daß auch *A* für Propositionen steht – diese Diskussion brauchen wir hier nicht zu wiederholen. In der Wahrscheinlichkeitstheorie ist diese Erweiterung trivial; bedingte Wahrscheinlichkeiten lassen sich mit Hilfe nicht bedingter definieren. Hingegen liegt in der deontischen Logik eine wesentliche Erweiterung vor; wie wir sahen, reichte der einstellige deontische Operator nicht zur Festlegung des bedingten dyadischen deontischen Operators aus. Und auch Reschers oben erwähntes Revisionsverfahren läßt sich als Definition eines bedingten Glaubensprädikats auffassen. Aber auch anderswo treten bedingte Begriffe auf. So wurden häufig bedingte komparative Glaubensrelationen untersucht[36], freilich weniger im Hinblick auf eine dynamische Interpretation als vielmehr im Zusammenhang der Metrisierungsproblematik. Savage (1954) stützt sich bei seiner simultanen Konstruktion eines Wahrscheinlichkeitsmaßes und einer Nutzenfunktion aus einer nicht bedingten Präferenzrelation ganz wesentlich auf die Definition einer bedingten Präferenzrelation aus dieser nicht bedingten Präferenzrelation; auf S. 23 streift er außerdem eine dynamische Interpretation für bedingte Präferenzrelationen. Bedingte Nutzenfunktionen schließlich werden auch gelegentlich erörtert, z.B. von Jeffrey (1965), S. 80, der dort auch eine dynamische Interpretation anklingen läßt; sie spielen insgesamt aber nur eine untergeordnete Rolle.

In der Tat scheint man ohne die bedingten Begriffe die von Beobachtungen oder Erfahrungen herrührende Dynamik von Glaubens- und Wünschensdispositio-

---

[35] In diesem Zusammenhang ist darauf hinzuweisen, daß Lewis (1973) in zumindest formal wesentlich ansprechenderer Weise eine Logik und unter Rückgriff auf den Begriff der komparativen Ähnlichkeit eine Semantik kontrafaktischer Konditionale vorgelegt hat. Es stellt sich somit die interessante Frage, ob und, wenn ja, wie sich dieser Begriff der komparativen Ähnlichkeit mit unseren Grundbegriffen für Glaubensdispositionen in Verbindung bringen läßt.

[36] Z.B. in Krantz et. al. (1971), Abschnitt 5.6.

nen nicht in den Griff zu bekommen. Denn wie sollte man die Auswirkungen der Erfahrungen des Subjektes X zwischen den Zeitpunkten t und t' auf seine Glaubens- und Wünschensdispositionen theoretisch erfassen können, wenn nicht bereits zum früheren Zeitpunkt t festliegt, wie diese Dispositionen von X zum späteren Zeitpunkt t' aussähen, wenn X die und die Erfahrungen zwischen t und t' machte? Und genau das wird ja durch die bedingten Begriffe für Glaubens- und Wünschensdispositionen ausgedrückt.

Zuletzt ist noch darauf hinzuweisen, daß sich auch die Entscheidungs- und Spieltheorie selbst bei augenscheinlich statischer Formulierung dynamisch verstehen lassen. Dies liegt daran, daß in ihnen wesentlich von Strategien die Rede ist. Denn nehmen wir an, der Entscheidende X habe sich in einer bestimmten Situation zur Strategie S durchgerungen. Macht nun X die Erfahrung, daß A, so steht er genaugenommen vor einem neuen Entscheidungsproblem, das als Teilproblem des ursprünglichen Problems freilich eine einfache Lösung hat: X wird sich in der neuen Situation gerade dafür entscheiden, was die Strategie S ihm für den Fall, daß A, vorschreibt. So trivial diese Überlegung ist, so unbestreitbar ist ihr dynamischer Charakter. Im vierten Kapitel kommen wir noch ausführlicher darauf zu sprechen. Im übrigen ist zu erwähnen, daß der dynamische Charakter der Spiel- und Entscheidungstheorie klar zutage tritt, wenn man sie nicht unter Rückgriff auf Strategien in Normalform, sondern in extensiver Form formuliert.[37]

Damit ist unser Streifzug durch die Landschaft der Theorien über Glaubens- und Wünschensdispositionen zu einem vorläufigen Ende gekommen . Nun haben wir bei diesem Streifzug freilich nur auf drei Merkmale geachtet, wir haben nur danach geschaut, ob diese Theorien eingleisig oder mehrgleisig, ob sie personal oder interpersonal und ob sie statisch oder dynamisch sind. Dies ermöglichte bereits eine recht reichhaltige Klassifizierung; doch haben wir bisher– ich deutete es schon zu Beginn dieses Abschnitts an – zumindest ein wesentliches Unterscheidungsmerkmal außer acht gelassen, auf das ich zum Abschluß noch kurz eingehen will. Ich denke dabei an die Beschaffenheit der zugrundeliegenden Propositionsmengen, der Definitionsbereiche unserer acht Grundbegriffe. Diese Definitionsbereiche hatten wir bisher nur mit einer logischen Struktur versehen, indem wir für Propositionen die aussagenlogischen Operationen wie Konjunktion etc. erklärten. Man könnte nun noch weiter nach der genauen Beschaffenheit dieser logischen

[37] Vgl. Dazu etwa Luce, Raiffa (1957), Kap. 3.

Struktur differenzieren: also z.B. P- bzw. $\preceq^g$-Theorien danach unterscheiden, ob $P_{X,t}$ bzw. $\preceq^g_{X,t}$ in ihnen für eine σ-Algebra oder nur für eine Algebra von Propositionen definiert ist; bei W- bzw. G-Theorien darauf achten, ob im Bereich des deontischen Operators $W_{X,t}$ bzw. des epistemischen Operators $G_{X,t}$ nur aussagenlogisch zusammengesetzte Sätze auftauchen oder ob auch Quantoren einbezogen werden; Versionen der Entscheidungstheorie mit einer nur für disjunkte Propositionen definierten Nutzenfunktion und andere Versionen mit einer auf einer Propositionenalgebra definierten Nutzenfunktion betrachten; etc. Mir kommt es jetzt aber weniger auf die technischen Einzelheiten der logischen Struktur, als vielmehr auf unterschiedliche, ins Inhaltliche gehende Strukturierungen der fraglichen Definitionsbereiche an. Und darüber haben wir bisher noch überhaupt nichts gesagt.

Freilich geht es mir bei dieser inhaltlichen Differenzierung nicht um eine Unterscheidung danach, ob in den Propositionen dieser Definitionsbereiche von Einkommensverhältnissen, von der Politik, vom Wetter oder von anderen Dingen die Rede ist. Das brächte auf der allgemeinen Ebene, auf der wir uns hier bewegen, nichts Neues. Dennoch gibt es wenigstens drei Substantiierungen der Definitionsbereiche unserer Grundbegriffe, die auch für die philosophische Grundlagenarbeit von Interesse sind und die völlig neue theoretische Schwierigkeiten mit sich bringen.

Die erste solcher Substantiierung beruht auf dem Umstand, daß sich aus der Anwendung unserer acht Grundbegriffe auf Propositionen Sätze ergeben, aus denen man in der üblichen Weise Propositionen abstrahieren kann: Propositionen, auf die man wieder unsere acht Grundbegriffe anwenden kann. Wir ziehen hier also in Betracht, daß sich in den Definitionsbereichen unserer acht Grundbegriffe Propositionen befinden, die selbst wiederum aus Glaubens- und Wünschensdispositionen zuschreibenden Sätzen abstrahiert sind. Mithin dreht es sich, kurz gesagt, um *Iterationen* unserer Grundbegriffe, also um die theoretische Erfassung solcher Konstruktionen wie

$$„G_{X,t}(W_{X,t}(A))“,\ „G_{Y,t}(A) \preceq^w_{X,t} G_{Y,t}(B)“,\ „P_{X,t}(W_{X,t'}(A)) = r“,$$

etc. Bedingte Begriffe und höhergradige Iterationen haben wir dabei in diese Beispiele noch gar nicht einbezogen. Diese Iterationen komplizieren die Materie also ganz ungeheuerlich.

Die verschiedenen Iterationen lassen sich immerhin analog zu unseren bisherigen Unterscheidungen klassifizieren. Da gibt es eingleisige Iterationen, in denen ein und derselbe Grundbegriff iteriert wird, und mehrgleisige Iterationen, die mehrere Grundbegriffe enthalten, personale Iterationen, die sich auf ein und dieselbe Person beziehen, und interpersonale Iterationen, bei denen zumindest zwei verschiedene Personen durch die Personenindizes der iterierten Begriffe bezeichnet werden, und entsprechend statische und dynamische Iterationen. Diese Iterationen sind im Prinzip alle sinnvoll; bei mehr als vier- oder fünffach geschachtelten Iterationen fällt es freilich schwer, ihren Sinn noch nachzuvollziehen. Dies dürfte damit zusammenhängen, daß mit zunehmendem Iterationsgrad immer unklarer wird, worin sich solche iterierten Dispositionen denn letztlich manifestieren. Bei iterierten Sätzen ist also das Problem, wie sie denn zu verstehen sind, entschieden drängender als bei nicht iterierten Sätzen; es sollte daher vorrangig angegangen werden. Als besonders akut empfinde ich dieses Interpretationsproblem bei personalen, statischen Iterationen. Hier erscheinen schon die einfachsten Iterationen problematisch: z.B. „*X* wünscht zum Zeitpunkt *t*, daß *X* zu *t* wünscht, daß *A*“ oder „*X* glaubt zum Zeitpunkt *t*, daß *X* zu *t* glaubt, daß *A*“.

Insofern ist es bedauerlich, daß es in praktisch allen existierenden Theorien, die sich mit Iterationen befassen, um personale, statische und zudem eingleisige Iterationen geht. So gibt es gut ausgebaute iterierte epistemische Logiken, z.B. Hintikka (1962), und iterierte deontische Logiken, z.B. Hintikka (1971), und neuerdings eine iterierte Präferenzlogik, Jeffrey (1978), und sogar eine iterierte Logik für die qualitative Wahrscheinlichkeit, d.h. für $\preceq^{g}$ (Gärdenfors (1975)). Diese sind aber, wie gesagt, allesamt personal und statisch.

Ansonsten ist mir nur noch eine Theorie bekannt, in der Iterationen vorkommen; es ist – wir erwähnten sie schon – die Spieltheorie: insofern nämlich, als sie annimmt, daß jedem Spieler die Nutzenfunktionen der anderen Spieler bekannt sind, was sich ja nur mit einer Iteration von *G* und *V* ausdrücken läßt. Dennoch kann man nicht sagen, daß sich die Spieltheorie explizit mit diesen Iterationen beschäftigt; vielleicht sollte sie das tun. Bemerkenswert ist immerhin, daß sich die namhafteste interpersonale Theorie wesentlich auf interpersonale Iterationen stützt. Dies ist vielleicht als Hinweis darauf zu werten, daß die Betrachtung interpersonaler Iterationen allgemein für interpersonale Fragestellungen fruchtbar ist. Analoges gilt für dynamische Probleme. Jedenfalls werden wir im vierten Kapitel feststellen, daß der Witz strategischen Denkens gerade in (personalen) dynamischen Iterationen liegt.

Mit einer anderen inhaltlichen Spezifizierung der Propositionen haben wir es zu tun, wenn wir in die Definitionsbereiche unserer acht Grundbegriffe *statistische Propositionen* aufnehmen, also Abstraktionen aus solchen Formulierungen wie „die statistische Wahrscheinlichkeit von A ist r“ oder eventuell komplizierteren Formulierungen, in denen noch von Versuchen und Versuchsanordnungen die Rede ist. Diese Bereicherung dürfte vor allem für die Definitionsbereiche der Grundbegriffe für Glaubensdispositionen bedeutsam sein. Ein fast selbstverständliches Gesetz, das in diesen Bereich fällt ist: Wenn X glaubt, daß die statistische Wahrscheinlichkeit von A r ist, dann glaubt X im Grade r, daß A.

Zu den Theorien über Glaubens- und Wünschensdispositionen, die auch statistische Propositionen berücksichtigen, gehören zum einen alle Versionen der Entscheidungs- oder der Spieltheorie, die eine objektivistische oder statistische Wahrscheinlichkeitsinterpretation zugrunde legen. Mit dem eben formulierten Gesetz gehen sie aber in die entsprechenden Versionen mit personalistischer Wahrscheinlichkeitsinterpretation über. Wichtiger ist freilich, daß zum anderen große Teile der statistischen Grundlagenforschung als zu diesem Bereich gehörend aufgefaßt werden können. Um ein Beispiel dafür zu geben: Hacking (1965) entwickelt sehr sorgfältig eine Theorie der Stützung für statistische Hypothesen: und zwar eine Theorie der komparativen Stützung (Kap. III und V) wie der quantitativen Stützung (Kap. IX). Er setzt sich also mit den folgenden Grundbegriffen auseinander: „e stützt h besser als h'“ und „e stützt h im Grade r“ wobei „e“, „h“ und „h'“ für statistische Propositionen Hackingscher Provenienz stehen. Liest man nun diese unpersönlichen, objektivierenden Formulierungen – analog zu früheren Fällen – personenbezogen als „X glaubt der Bedingung e eher, daß h, als daß h'“ bzw. als „X glaubt unter der Bedingung e im Grade r, daß h“, so entpuppt sich Hackings Theorie gerade als bedingte $\preceq^{g}$ P-Theorie für statistische Propositionen.[38] Das oben genannte Gesetz ist dann gerade Hackings „frequency principle“ (S. 135).

Eine letzte inhaltliche Spezifikation von philosophischem Interesse – bestimmt gibt es noch mehr – erscheint mir noch erwähnenswert: Sie besteht darin, den Definitionsbereichen unserer Grundbegriffe Propositionen hinzuzufügen, die *Sprach-*

---

[38] P-Theorien für statistische Propositionen dürfen natürlich nicht mit iterierten P-Theorien verwechselt werden; statistische Propositionen beschreiben ja keine Glaubensdispositionen – jedenfalls solange man statistische Wahrscheinlichkeiten als etwas eigenständig Sinnvolles anerkennt und nicht durch personelle Wahrscheinlichkeiten wegerklären will.

verhalten beschreiben. Natürlich gibt es auch darauf zugeschnittene Theorien: z.B. Lerntheorien, die sich mit dem Sprachbereich beschäftigen. Insbesondere möchte ich hier aber neben Grice (1957) Lewis (1969) anführen, der sich zunächst unter Rückgriff auf spieltheoretische Konzeptionen (Kap. I) und dann durch ausgiebige Verwendung interpersonaler Iterationen (Kap. II) um eine Klärung von Konventionen allgemein und speziell von Sprachkonventionen (Kap. IV und V) bemüht. Jedenfalls ist anzunehmen, daß Arbeiten in dieser Richtung Linguistik und Sprachphilosophie ganz entscheidend bereichern können.

Damit ist unser kartographisches Unternehmen abgeschlossen. Der Maßstab unserer Landkarte ist zwar noch klein, aber nicht so klein, daß sich nicht bereits die wichtigsten Formationen klar erkennen ließen. In mancherlei Hinsicht ist unsere Landkarte freilich noch oberflächlich oder einseitig. So haben wir, wie zu Anfang erläutert, von den unterschiedlichen Zielrichtungen der verschiedenen Theorien einfach abgesehen. Ferner haben wir überhaupt nicht angesprochen, wie adäquat die einzelnen Theorien sind; da wird es ja erst interessant. Und schließlich haben wir uns, philosophisch orientiert, auf abstrakt-allgemeine und formal einigermaßen präzisierte Theorien konzentriert und der umfangreichen psychologischen Detailarbeit keinen Platz eingeräumt. Dies alles hätte auch den Rahmen gesprengt; es ging hier ja zunächst nur darum, das begriffliche und theoretische Umfeld der Entscheidungstheorie zu sondieren, um so eine genauere Lokalisierung des Standorts der Entscheidungstheorie zu ermöglichen. Ein Resultat unserer Aufzählung und Klassifizierung verdient freilich, noch einmal ausdrücklich festgehalten zu werden: Was sich im ersten Ausschnitt dieses Kapitels mit den Bemerkungen über die entscheidende Bedeutung der Glaubens- und Wünschensdispositionen für unser Verhalten andeutete, setzte sich nun im zweiten Abschnitt auf theoretischer Ebene fort; mit Glaubens- und Wünschensdispositionen beschäftigen sich nicht nur viele einzelwissenschaftliche Theorien, sondern in einem sehr erheblichen Maße auch die analytische Philosophie. Keine Frage, daß diese Bilanz noch eindrücklicher ausfiele, wenn man auch die weniger präzisierten philosophischen Beiträge einbezöge.

KAPITEL 2

Einige Konzeptionen der Entscheidungstheorie

2.1 Einleitung

Nachdem wir im ersten Kapitel ausgehend von der weitgehenden Steuerung menschlichen Verhaltens durch Glaubens- und Wünschensdispositionen einen Rahmen umrissen haben, in den sich viele wissenschaftliche und insbesondere auch philosophische Theorien einordnen lassen, und so einerseits den Standort der Entscheidungstheorie in diesem Rahmen bestimmen konnten und andererseits auf einige bei einer eingehenderen Beschäftigung mit der Entscheidungstheorie zu beachtende Punkte stießen, wollen wir uns nun einer detaillierteren Analyse der Entscheidungstheorie zuwenden.

Dazu sei zunächst das, was wir da unter die Lupe nehmen wollen, etwas genauer abgegrenzt, denn schließlich ist die Entscheidungstheorie alt genug, um schon einen beträchtlichen Grad der Wucherung erreicht zu haben. Kurz gesagt, wollen wir uns mit einigen allgemeinen Formulierungen der Entscheidungstheorie im Hinblick auf ihre logisch-begriffliche Beschaffenheit auseinandersetzen, und zwar mit solchen Formulierungen, die es als sinnvoll erachten, den Gegenständen der Theorie – seien es nun Menschen, Organisationen oder Tiere – sowohl subjektive Wahrscheinlichkeiten als auch subjektive Werte[1] zuzuschreiben, und die dann das Bay-

[1] Nachdem ich mich im ersten Kapitel noch an den am ehesten eingebürgerten Begriff „Nutzen" hielt, will ich in den systematischen Kapiteln stattdessen den mir passender erscheinenden Begriff „subjektiver Wert" verwenden. „Nutzen" hat doch einen etwas irreführenden Beigeschmack von Egoismus und Materialismus (im nicht philosophischem Sinn). Die Mehrdeutigkeit des Wortes „Wert" kann dabei wegen der permanenten Hinzufügung des Adjektivs „subjektiv" kein Unheil anrichten.

essche Prinzip der Maximierung des erwarteten subjektiven Wertes – je nach Zuschnitt der Formulierung – zur Beschreibung dieser Gegenstände, als Empfehlung an sie oder auch als Explikation des Rationalitätsbegriffs verwenden. Konzeptionen dieser Art sind nun einmal – ob zu Recht oder nicht – die bei weitem am allgemeinsten formulierten, am detailliertesten erforschten und am umfassendsten angewandten. Repräsentativ für solche Konzeptionen dürften die folgenden, in den kommenden Abschnitten dieses Kapitels in ihrer chronologischen Reihenfolge behandelten Theorien sein: erstens die auch als statistisch bezeichnete Theorie von Savage (1954), die wohl mit Wald (1950) die erste und nach wie vor die wichtigste allgemeine Formulierung der Entscheidungstheorie darstellt und an die sich die meisten Detailuntersuchungen anschließen; zweitens die Konzeption von Fishburn (1964), die sich von der Savages schon im Ansatz unterscheidet; drittens die wieder ganz anders geartete, eher von philosophischer Seite kommende Konzeption von Jeffrey; und viertens schließlich die Theorie von Luce und Krantz (1971), die sich zwar auf Fishburn beruft,[2] die merkwürdigerweise aber doch eher als Synopsis von Savages und Jeffreys Theorie denn als Weiterentwicklung von Fishburns Konzeption zu bewerten sein wird.

Worum es hier gehen soll, wird allerdings (zumindest für den Informierten) noch klarer, wenn wir die fünf Punkte nennen, um die es hier nicht gehen soll. So wollen wir erstens die sogenannten probabilistischen Modelle in der Entscheidungstheorie völlig außer acht lassen, die insbesondere von psychologischer Seite aus entwickelt worden und deren Charakteristikum es ist, daß sie nicht wie sonst üblich von Präferenzrelationen, sondern von Präferenzwahrscheinlichkeiten reden, also die Wendung „die statistische Wahrscheinlichkeit dafür, daß der Entscheidende X die Alternative f der Alternative g vorzieht, beträgt r“ zu einem ihrer Grundbegriffe machen.[3] Wir halten uns hier also an die üblichen, von Luce, Suppes (1965) auch als algebraisch bezeichneten Konzeptionen der Entscheidungstheorie.

---

[2] S. Luce, Krantz (1971), S. 253.

[3] Probabilistisch heißen diese Modelle also nicht deswegen, weil sie dem Entscheidenden subjektive Wahrscheinlichkeiten unterstellen, sondern weil sie den Entscheidenden lediglich statistisch und nicht deterministisch zu erfassen trachten. Einen guten Überblick über solche probabilistische Modelle geben Luce, Suppes (1965), Abschnitte 5-8. Luce (1959), S. 77-79, zufolge unterscheiden sich diese von den hier zu behandelnden algebraischen Theorien nur dadurch, daß sie noch auf das Phänomen der unvollkommenen Diskriminierung von Alternativen eingehen. Insofern kann eine Analyse algebraischer Theorien auch für sie von gewissem Nutzen sein, auch wenn sie sich im wesentlichen mit den von dem probabilistischen Ansatz geschaffenen, dem algebraischen Ansatz weitgehend fremden Problemen beschäftigen.

Während diese probabilistischen Modelle ausgesprochen im Hinblick auf empiristische Angemessenheit ersonnen wurden, sind die hier betrachteten Konzeptionen diesbezüglich weniger eindeutig. Doch können wir unsere logische Analyse zweitens auch von dem Problem freihalten, wie die Entscheidungstheorie zu interpretieren sei: ob als rein empirische Theorie, als unmittelbar normative Theorie, als Logik der Ausdrücke für Glaubens- und Wünschensdispositionen oder als eine Theorie über rationale Individuen, was wiederum verschieden gedeutet werden kann; nämlich als idealisierte empirische Theorie, als bloße (partielle) Explikation des Rationalitätsbegriffs oder auch normativ – man soll sich ja rational verhalten. Es wäre zwar wichtig, sich darüber mehr Klarheit zu verschaffen; doch sind die Probleme, mit denen sich unsere Analyse beschäftigen wird, von diesen Interpretationsproblemen ganz unabhängig. Anstatt nun hier aber dauernd komplizierte Formulierungen zu wählen, die möglichst allen Interpretationen gerecht werden sollen, werde ich der Einfachheit halber in den nächsten Kapiteln von einer bestimmten Interpretation ausgehen: Ich werde dauernd so tun, als gelte es, ohne Wenn und Aber eine empirische Theorie über menschliches Verhalten zu formulieren. Diese Interpretation wähle ich deshalb, weil sie mir die unkomplizierteste und geradlinigste zu sein scheint; ob sie auch die vernünftigste und haltbarste ist, kann dabei ganz dahingestellt bleiben. Durch dieses Vorgehen wird aber die Allgemeinheit der kommenden Ausführungen in keiner Weise eingeschränkt. Denn jede Äußerung, die diese Interpretation voraussetzt, läßt sich leicht gemäß den anderen Interpretationen umformulieren. Wenn ich z.B., von der empirischen Interpretation ausgehend, behaupte, daß der Handelnde seinen erwarteten subjektiven Wert maximiert, so muß man, will man andere Interpretationen zugrunde legen, stattdessen einfach sagen: „Man *soll* seinen erwarteten subjektiven Wert maximieren!“ oder „Ein rationales Individuum maximiert seinen erwarteten Nutzen“ etc.

Aus unserem Interesse an einer begrifflichen Analyse ergibt sich drittens, daß wir den vielfältigen Verästelungen der Entscheidungstheorie, den zahlreiche Untersuchungen spezieller Entscheidungsprobleme mit zusätzlichen Annahmen über die Entscheidungssituation, über die subjektiven Wahrscheinlichkeiten oder über die subjektiven Werte nicht nachzuspüren brauchen, auch wenn sie es sind, die der Entscheidungstheorie erst Substanz geben. Vielmehr interessiert uns nur die allgemeine Formulierung, der Theoriekern, da seine Spezialisierungen, logisch gesehen, nichts Neues bringen.

Viertens wollen wir uns hier auch nicht auf die verschiedentlich vorgebrachten Zweifel am Prinzip der Maximierung des erwarteten subjektiven Wertes und auf die deswegen vorgeschlagenen komplizierteren Entscheidungsregeln einlassen.[4] Daß wir uns auf die von der Bayesschen Regel ausgehenden Standardkonzeptionen der Entscheidungstheorie beschränken, bedeutet gleichzeitig, daß wir das, was Luce, Raiffa (1957), S. 13, Entscheidung unter Unsicherheit getauft haben, nicht behandeln. Wir lassen also die Bemühungen außer acht, auch dann noch eine angemessene Entscheidungsregel zu formulieren, wenn der Entscheidende über keine noch so spärliche probabilistische Einschätzung seiner Situation verfügt.[5]

Fünftens und letztens berühren uns hier nicht einmal Zweifel an der Berechtigung der Zuschreibung quantitativer Glaubens- und Wünschensgrade. Dieser Punkt ist für unser gesamtes weiteres Vorgehen und insbesondere für unsere Behandlung der eingangs genannten Konzeptionen ganz wesentlich. Diese lassen sich nämlich von diesen Zweifeln sehr beeindrucken und sind daher zu einem großen Teil mit ihrer Beseitigung beschäftigt. Die vorherrschende Methode dazu besteht im Beweis geeigneter Metrisierungsresultate, die, grob gesagt, zeigen, daß sich die problematisierten quantitativen Glaubens- und Wünschensgrade in hinreichend eindeutiger Weise aus dem als weniger problematisch empfundenen komparativen Präferenzrelationen gewinnen lassen.[6] Demgegenüber werde ich in dieser Arbeit das ganze Metrisierungsthema links liegen lassen und mich auf die Formulierung quantitativer Entscheidungsmodelle konzentrieren.

Ist das überhaupt durchführbar? Ich glaube schon. Denn nur wenn man sich vom Zweifel am Sinn quantitativer Glaubens- und Wünschensgrade völlig gefangen nehmen läßt, kann einem entgehen, daß zwei Fragen streng auseinanderzuhalten sind: erstens die Frage, wie eine Theorie quantitativer Glaubens- und Wünschensgrade am besten zu formulieren ist, und zweitens die Frage, auf welche Wei-

---

[4] So waren etwa Allais (1953) und Ellsberg (1961) mit den von ihnen konstruierten Paradoxa darauf aus, die Bayessche Entscheidungsregel als normativ unbefriedigend zu erweisen; vgl. dazu die meines Erachtens überzeugenden Entkräftigungsversuche von Raiffa (1975), Abschnitte 4.9 und 5.4. Und erst recht wurde die empirische Adäquatheit der Bayesschen Regel angezweifelt; vgl. dazu etwa Luce, Suppes (1963) Abschnitt 4.4. oder Tversky (1975).

[5] Diese Bemühungen, ausführlich dargestellt z.B. in Luce, Raiffa (1957), Kap. 13, waren eine Zeit lang hauptsächlich deswegen prominent, weil man weitgehend noch eine objektivistisch-statistische Interpretation des Wahrscheinlichkeitsbegriffs zugrunde legte. Mittlerweile hat sich die Erkenntnis durchgesetzt, daß subjektive Wahrscheinlichkeiten die einzigen der Entscheidungstheorie angemessenen sind; und damit wurde das Problem der Entscheidung unter Unsicherheit weniger drängend.

[6] Die gründlichste Darstellung der Metrisierungstheorie im allgemeinen bzw. im Rahmen der Entscheidungstheorie dürfte Krantz et. al (1971) bzw. Fishburn (1970) sein.

se eine solche quantitative Theorie empirische Signifikanz gewinnt bzw. (wenn man sie als empirische Theorie auffaßt) wie sie sich überprüfen läßt. Natürlich läßt sich im allgemeinen die erste Frage nicht unabhängig von der zweiten behandeln. In unserem speziellen Fall besteht aber doch eine relative Unabhängigkeit. Dies liegt daran, daß die Last der empirischen Signifikanz eines entscheidungstheoretischen Modells vor allem auf der in ihm angenommenen Entscheidungsregel beruht und daß wir hier lauter in dieser Hinsicht gleichartige Modelle diskutieren, nämlich solche, die das hier nicht problematisierte Prinzip von der Maximierung des erwarteten subjektiven Wertes akzeptieren. Für eine Einstufung dieser Modelle gemäß der ersten Frage müssen also andere Kriterien maßgeblich sein, als sie sich aus einer Beantwortung der zweiten Frage ergeben können. Wir können daher die Formulierungsfrage für quantitative Modelle weitgehend losgelöst von ihrer Überprüfungsproblematik untersuchen. Dies darf natürlich nicht darüber hinwegtäuschen, daß für eine voll ausgereifte Entscheidungstheorie, verwende sie nun die Bayessche Entscheidungsregel oder doch eine andere, die Behandlung der Überprüfungsproblematik ganz entscheidend ist. Im Abschnitt 5.3 will ich noch etwas ausführlichere Bemerkungen dazu machen.

Was sind nun aber diese anderen Kriterien, die noch zwischen im Kern ähnlichen Formulierungen quantitativer Entscheidungsmodelle zu unterscheiden vermögen? Da sind einmal solche Dinge wie technische Eleganz und Verallgemeinerungsfähigkeit. Doch ist Eleganz zwar schön, aber nicht unabdingbar, und der technischen Verallgemeinerungsfähigkeit können wir hier nicht nachspüren, da wir unsere Untersuchung, wie im Vorwort schon erwähnt, mathematisch elementar halten wollen. Im übrigen dürften in dieser letzten Hinsicht alle hier diskutierten Konzeptionen gleichgestellt sein. Vergleiche, die auf die logisch-begriffliche Beschaffenheit der verschiedenen Konzeptionen abheben, erscheinen mir daher ergiebiger und maßgeblicher. Konkret kommt es mir dabei vor allem auf die folgenden drei Fragen an:

Die Allgemeinheitsfrage: Wie allgemein oder wie speziell ist die jeweilige Konzeption? D.h., gehen in die jeweilige Konzeption spezielle Voraussetzungen über die Entscheidungssituation oder über die quantitativen Glaubens- und Wünschensdispositionen des Handelnden ein und, wenn ja, welche? Dies ist natürlich die Hauptfrage, und sie wird den meisten Raum in Anspruch nehmen. (Es dürfte klar sein, daß sie von der Frage nach der technischen Verallgemeinerungsfähigkeit verschieden ist.)

Die Explizitheitsfrage: Welche theoretische Klärung erfahren die Grundentitäten der jeweiligen Konzeption, und inwieweit lassen sich diese Grundentitäten, sofern sie noch komplexer Natur sind, in der jeweiligen Konzeption selbst analysieren? Insistieren möchte ich dabei auf der Klärung durch die Theorie selbst. Denn daß ein Autor die Grundentitäten seiner Theorie erfolgreich zu erläutern und in konkreten Anwendungen zu identifizieren vermag, spricht erst einmal für den Autor und nicht unbedingt für seine Theorie.

Die Anwendungsfrage: Läßt sich innerhalb der jeweiligen Konzeption sagen, was als *eine* Entscheidungssituation oder als *ein* Entscheidungsproblem anzusehen ist? Der Hintergedanke bei dieser Frage ist der, daß es ein Qualitätsmerkmal einer jeden Theorie ist, wenn sie selbst ihre Anwendungen theoretisch abzugrenzen vermag. (Z.B. ist es in der Newtonschen Mechanik möglich, diejenigen Dinge, auf die sie sich anwenden läßt, nämlich abgeschlossene System in äußeren Kraftfeldern, zu definieren.)

Unser weiteres Vorgehen ist damit ziemlich klar. In diesem Kapitel wollen wir die anfangs genannten Konzeptionen vor allem in Hinblick auf diese drei Fragen darstellen und untersuchen, und im nächsten Kapitel wollen wir dann unsere Lehren daraus ziehen (und noch ein bißchen mehr tun). Bei alledem ist zu beachten, daß sich dieses und das nächste Kapitel noch ganz auf die Formulierung und Untersuchung statischer Entscheidungsmodelle beschränken. Die notwendige Erweiterung des Blickfeldes auf dynamische Fragestellungen erfolgt erst im vierten Kapitel. Bevor wir uns nun in medias res stürzen, ist noch die folgende Bemerkung zur Schreibweise nötig, die sich daraus ergibt, daß die Entscheidungstheorie eine personale Theorie ist. Deswegen können wir nämlich im Kommenden – wie auch schon im ersten Kapitel – stellvertretend für alle Anwendungsobjekte der Entscheidungstheorie durchweg von einem Handelnden X reden, und auch dies brauchen wir nur in nicht formalen Ausführungen zu tun; denn da in der jeweiligen formalen Theorie immer über dasselbe Individuum gesprochen wird, empfiehlt es sich, die Notation der Theorie von der Last der expliziten Anführung von X zu befreien. Ähnliches gilt für die Zeitpunkte: Sie bedürfen der Kenntlichmachung erst, wenn wir uns mit dynamischen, also mehrere Zeitpunkte involvierenden Überlegungen befassen.

2.2 Das Grundmodell von Savage

Eröffnet sei der Reigen mit dem – wie man es nennen könnte – entscheidungstheoretischen Grundmodell von Savage (1954), das sich durch seinen elementaren Charakter jedem Lehrbuch für Entscheidungstheorie als Einführung und uns als Ausgangspunkt unserer Untersuchung empfiehlt. Gerade die Allgemeinheits- und die Explizitheitsfrage lassen sich an ihm beispielhaft diskutieren, so daß unsere späteren Ausführungen davon nur profitieren können, zumal alle zu erörternden Konzeptionen sich in der einen oder anderen Form an ihm orientieren.

Zunächst sei der Gedankengang, der zur Konstruktion des Grundmodells führt, kurz rekapituliert, freilich ohne es noch ausführlich an Beispielen zu illustrieren.[7]

Das Grundmodell geht davon aus, daß dem Handelnden X in einer bestimmten Situation verschiedene Handlungen $f_1,...,f_m$ offenstehen und daß X nun zu entscheiden hat, welche dieser Handlungen er ausführen soll. Seine Entscheidung wird davon abhängen, zu welchen Konsequenzen die verschiedene Handlungen führen Die Konsequenzen von f_i stehen nun aber im allgemeinen für X nicht von vornherein fest; vielmehr wird X annehmen, daß sie auch durch die genauen Umstände der Situation, in der er sich befindet, bestimmt werden. Das Grundmodell geht nun davon aus, daß X zwar nicht weiß, welcher der möglichen Umstände $w_1,...,w_n$ in seiner Situation vorliegt, daß er aber, wenn er dies wüßte, auch wüßte, zu welchen Konsequenzen die ihm offenstehenden Handlungen führen. Die durch f_i und w_j nach Xs Meinung eindeutig bestimmten Konsequenzen seinen mit c_{ij} bezeichnet.

Wenn man Xs Situation so, wie eben geschildert, sieht, was ist damit gewonnen? Nun, die verschiedenen Konsequenzen c_{ij} sind für ihn von mehr oder weniger großem subjektiven Wert. Die genauen subjektiven Werte seien in der subjektiven Wertfunktion V von X festgehalten, die jedem c_{ij} eine reelle Zahl $V(c_{ij})$ zuordnet. Wenn X nun schon die Konsequenzen jeder Handlung kennte, so würde er sich natürlich zu einer Handlung entscheiden, deren Konsequenzen maximalen subjektiven Wert für ihn haben. Nun hatten wir ja aber angenommen, daß X sich der jeweiligen Konsequenzen nicht sicher ist. Um auch in diesem Fall der Entscheidung von X theoretisch gerecht zu werden, nimmt das Grundmodell neben den subjekti-

---

[7] Für illustrierte Darstellungen s. etwa Jeffrey (1965), Kap. 1, oder Stegmüller (1973a), S. 288-298.

ven Werten von X auch noch subjektive Wahrscheinlichkeiten $P(w_j) \geq 0$ von X für die verschiedenen möglichen Umstände $w_1,...,w_n$ an. Da angenommen wird, daß es für X sicher ist, daß genau einer der möglichen Umstände $w_1,...,w_n$ vorliegt, muß natürlich $\sum_{j=1}^{n} P(w_j) = 1$ gelten. Man kann den Sachverhalt dann auch so ausdrücken, daß f_i für X mit der Wahrscheinlichkeit $P(w_j)$ zur Konsequenz c_{ij} führt. Daß es überhaupt sinnvoll ist, X eine subjektive Wertfunktion V und eine subjektive Wahrscheinlichkeitsfunktion P zuzuschreiben, hatten wir ja zur Voraussetzung dieses ganzen Kapitels gemacht.

Wie wird sich X nun entscheiden? Die Grundannahme aller hier zu diskutierender Theorien besteht darin, daß X – wo er schon den subjektiven Wert selbst nicht maximieren kann (da die Konsequenzen unsicher sind) – dann wenigstens den erwarteten subjektiven Wert maximiert. D.h. X wird sich für eine Handlung f_i entscheiden, deren erwarteter subjektiver Wert $\sum_{j=1}^{n} V(c_{ij}) \cdot P(w_j)$ maximal ist. Dieses Bayessche Prinzip der Maximierung des erwarteten subjektiven Wertes wollen wir hier, wie gesagt, ebenfalls nicht problematisieren, da es sich ja im Augenblick nicht um die Richtigkeit gewisser empirischer Behauptungen oder um die Angemessenheit gewisser Empfehlungen dreht, sondern um eine begriffliche Analyse. Und an diesem Prinzip gibt es außer der in es eingehenden Begriffe nicht zu klären.

Begünstigt durch die oft sehr einfach gehaltenen Illustrationen mag sich leicht der Eindruck einstellen, dieses so charakterisierte Grundmodell sei in der Anwendung sehr beschränkt. Doch täuscht man sich da. Denn erstens sind die in das Grundmodell eingehenden Endlichkeitsannahmen (m Handlungen, n mögliche Umstände und demnach höchstens $m \cdot n$ verschiedene Konsequenzen) nicht sonderlich einschränkend – es gibt ja recht große natürliche Zahlen – und auch nicht wesentlich; sie resultieren ja vor allem aus dem Einführungscharakter des Grundmodells und lassen sich auf technisch etwas höherem Niveau ohne Mühe beseitigen.[8] Zweitens muß man sich vergegenwärtigen, daß das, was da so harmlos Umstände und Konsequenzen genannt wird, in Wirklichkeit höchst komplexe Dinge sein können. So können ja z.B. in die möglichen Umstände der Entscheidungssituation eines disponierenden Bauers von der Wetterentwicklung des gesamten nächsten Jahres bis hin zu der langfristigen Preisentwicklung diverser landwirtschaftlicher Erzeugnisse alles Mögliche eingehen, und auch die möglichen Konse-

[8] Wie es z.B. Savage (1954) tut.

quenzen von den Erträgen seiner Felder bis hin zu den familiären Auswirkungen der anvisierten Arbeitsbelastung die verschiedensten Dinge umfassen. Der dritte Verallgemeinerungspunkt dürfte allerdings der wichtigste sein: Was da als offenstehende Handlung bezeichnet wird, braucht ja nicht eine einzelne Handlung im üblichen Sinne sein. Es können damit auch sich über längere Zeiträume erstreckende Handlungsabläufe und insbesondere auch Strategien gemeint sein, also wohl ausgefeilte Handlungspläne, die auf alle Eventualitäten vorbereitet sind. Man braucht also keine Angst zu haben, daß das Grundmodell zu klein dimensioniert ist. Richtig ist freilich, daß das Grundmodell als solches noch keine sonderlich substantielle Theorie liefert. Dazu gehört eine Ausarbeitung des Modells, die Untersuchung von besonderen Fällen und Problemen etc., wie es in Lehrbüchern der Entscheidungstheorie geschieht.

Fassen wir zunächst den formalen Gehalt des bisher Gesagten zu einer Definition zusammen. Danach können wir uns einer eingehenden Diskussion des Grundmodelle zuwenden:

Definition 1: $\langle W, C, H, P, V, U\rangle$ ist ein *Savage-Modell* genau dann, wenn gilt:

(1) W und C sind nicht leere, endliche Mengen,

(2) H ist eine nicht leere Menge von Funktionen von W und C,

(3) P ist ein Wahrscheinlichkeitsmaß auf Pot(W),[9]

(4) V ist eine Funktion von C in $\mathbf{R}$,[10]

(5) U ist die auf H definierte Funktion, für die alle $f \in H$

$$U(f) = \sum_{w \in W} V(f(w)) \cdot P(W).$$

Dazu sind einige Erläuterungen nötig: Der Handelnde X ist gemäß unserer Vereinbarung am Ende des Abschnitts 2.1 in dieser Definition nicht erwähnt. W repräsentiert die Menge der möglichen Umstände, die wir auch *als Möglichkeitsraum* bezeichnen wollen; C stellt die Menge der möglichen Konsequenzen dar. Weiterhin soll H die Menge der dem Handelnden offenstehenden Handlungen oder auch Strategien repräsentieren. Um dies einzusehen, müssen wir einen gedanklichen Trick Savages nachvollziehen. Dieser besteht darin, daß wir in Anbetracht dessen, daß jede Handlung zusammen mit einem Umstand die dann eintretende Konse-

---

[9] Pot(W) bezeichnet die Potenzmenge von W, d.h. die Menge aller Teilmengen von W.

[10] $\mathbf{R}$ bezeichnet immer die Menge der reellen Zahlen.

quenz eindeutig festlegt, Handlungen als Funktionen auffassen können, die jedem möglichen Umstand aus W eine Konsequenz aus C zuordnen. Und gerade das tun wir im Definition 1,(2). Daß Savage Handlungen auf diese Weise repräsentiert, hat vornehmlich technische Gründe, auf die wir hier nicht einzugehen brauchen.

Die Teilmengen von W bezeichnen wir, wie in der Wahrscheinlichkeitstheorie üblich, als *Ereignisse*. Ein $A \subseteq W$ drückt also das Ereignis aus, daß einer der Umstände aus A tatsächlich vorliegt. Und P gibt dann die subjektiven Wahrscheinlichkeiten des Entscheidenden für die verschiedenen Ereignisse an. Wir gestatten uns dabei, für $w \in W$ statt $P(\{w\})$ einfach $P(w)$ zu schreiben. Natürlich hätte es in dem hier durchweg angenommenen endlichen Fall genügt, lediglich den möglichen Umständen subjektive Wahrscheinlichkeiten zukommen zu lassen, da ja hier für jedes $A \subseteq W$ $P(A) = \sum_{w \in A} P(w)$ gilt. Doch wird sich die hier gewählte Festlegung noch als nützlich erweisen (und im hier nicht diskutierten unendlichen Fall wäre sie ohnehin unentbehrlich).

V drückt die subjektiven Werte des Entscheidenden für die möglichen Konsequenzen aus. Und schließlich ist es vorteilhaft, die erwarteten subjektiven Werte der Handlungen direkt in das jeweilige Entscheidungsmodell aufzunehmen; dies wird durch die Funktion U gewährleistet.

Die mit Savage-Modellen verbundene empirische Behauptung – wir wollen ja, wie gesagt, der Einfachheit halber von einer empirischen Interpretation der Entscheidungstheorie ausgehen – lautet dann einfach: Wird X durch das Savage-Modell $\langle W, C, H, P, V, U \rangle$ charakterisiert, so führt X ein $f \in H$ aus, für das $U(f)$ maximal ist.

Wie steht nun das Grundmodell im Lichte unserer drei Fragen da? Bezüglich der Anwendungsfrage ist ihm bisher nichts zu entnehmen. Doch äußert sich Savage zu ihr ausführlich; wir wollen darauf im nächsten Abschnitt eingehen. Und bezüglich der Allgemeinheits- und der Explizitheitsfrage sind die Antworten im Grunde bekannt. Wir wollen sie hier trotzdem ausführlich darstellen, da sie für ein Verständnis sowohl des Grundmodells wie der späteren Abschnitte ganz wesentlich ist.

Zur Explizitheitsfrage: Zur Anwendung des Grundmodells gilt es natürlich nicht nur, die jeweilige Entscheidungssituation geeignet abzugrenzen. (Darauf zielt ja gerade die Anwendungsfrage ab.) Zuallererst muß man wissen, welche Dinge überhaupt jeweils als die möglichen Handlungen, als die möglichen Umstände und als die möglichen Konsequenzen anzusehen sind. Dies ist nicht überraschend; jede

Theorie setzt zu ihrer Anwendung das Verständnis ihrer Grundbegriffe voraus. Freilich ist es wünschenswert, daß die Theorie einem dabei möglichst weitgehend hilft und daß das Verständnis des von der Theorie selbst ungeklärten Restes möglichst geringe Anforderungen stellt. Wie steht es in dieser Hinsicht mit dem entscheidungstheoretischen Grundmodell?

Bei aller Geringfügigkeit verdient es zunächst festgehalten zu werden, daß man den Elementen der Mengen W, C und H eines Savage-Modells ihren propositionalen Charakter nicht ansieht. (Daß sie ihn haben, stellten wir ja schon im ersten Kapitel fest.) Wichtiger ist freilich, was wir aus dem Grundmodell über die Besonderheiten der Elemente dieser Mengen erfahren.

Was unter einer (möglichen) Handlung zu verstehen ist, sagt das Grundmodell selbst nicht; es verläßt sich da ganz auf das intuitive Verständnis. Das Grundmodell steht in dieser Hinsicht, wie wir noch sehen werden, allerdings nur genauso schlecht da wie alle anderen Konzeptionen der Entscheidungstheorie. Jede Konzeption setzt den Handlungsbegriff schlankweg voraus, und ob sich hier im Kontext der Entscheidungstheorie eine Verbesserung erreichen läßt, ist zumindest fraglich. Seine Klärung ist vielleicht auch eher Aufgabe der Handlungstheorie als die der Entscheidungstheorie.

Über die möglichen Umstände erfahren wir mehr: Das Grundmodell sagt erstens, daß die möglichen Umstände in Xs Sicht einander ausschließen und daß das Vorliegen eines möglichen Umstandes für X sicher ist. Und zweitens geht aus dem Grundmodell hervor, daß die möglichen Umstände nach Xs Meinung handlungsunabhängige Ereignisse sind: die subjektiven Wahrscheinlichkeiten der möglichen Umstände stehen für X fest, egal welche Handlung er ausführt. Und mehr läßt sich über die möglichen Umstände allgemein wohl nicht sagen: Prinzipiell jede Menge handlungsunabhängiger Ereignisse, die einander ausschließen und insgesamt erschöpfend sind, kann als Menge möglicher Umstände eines Entscheidungsproblems dienen. (Dies zeigt freilich die Dringlichkeit der Anwendungsfrage gerade in bezug auf die möglichen Umstände.)

Mit den möglichen Konsequenzen ist es aber wiederum schlecht bestellt. Über sie erfahren wir nur, daß sie durch Handlungen und Umstände eindeutig bestimmt sind. Und das ergibt wenig Hinweise für die konkrete Situation. Auch fühlt man sich hier viel eher als vom Grundmodell im Stich gelassen als bezüglich der Handlungen; denn was Handlungen sind, ist intuitiv weitgehend klar; was dagegen alles zu den Konsequenzen einer Handlung zu rechnen ist, ist weit weniger klar.

Und von einer expliziten Theorie ist zu erhoffen, daß sie in dieser Hinsicht mehr an die Hand gibt.

Zu betonen ist allerdings noch einmal, daß all dies keine Argumente gegen die Anwendbarkeit des Grundmodells sind. Im Gegenteil, es bleibt ganz unbestritten, daß sich das Grundmodell oft ohne Schwierigkeit anwenden läßt. Nur läßt sich dies nicht als Verdienst der Formulierung des Grundmodells ausgeben.

Schließlich ist zur Explizitheitsfrage noch zu bemerken, daß sich die Komplexität, die die Handlungen, Umstände und Konsequenzen annehmen können, in Savage-Modellen, wie wir sie formuliert haben, nicht widerspiegelt. Für die technische Ausarbeitung der Theorie ist dies sogar von Vorteil. Man denke nur an die gewaltige technische Vereinfachung, die man in der Spiel- und auch in der Entscheidungstheorie erzielen kann, indem man von ihrer extensiven Form zu ihrer Normalform übergeht. Allerdings böte es auch keine Schwierigkeit, in Savage-Modellen auszudrücken, wie sich komplexe Umstände, Konsequenzen und Handlungsabläufe oder gar Strategien aus Einzelteilen zusammensetzen.

Zur Allgemeinheitsfrage: Die Annahmen des Grundmodells über die subjektive Wertfunktion V von X sind schwerlich Quelle irgendwelcher spezieller Voraussetzungen. Schon der Definitionsbereich von V, die Menge der möglichen Konsequenzen, erwies sich als reichlich unbestimmt, und sonst erfahren wir von V nur, daß es reelle Zahlen als Werte annimmt.[11]

Ein Punkt ist hier aber doch von Interesse. Es liegt die naive Frage nahe, wieso denn die subjektive Wertfunktion nur für Konsequenzen, nicht aber für Handlungen oder gar für mögliche Umstände definiert ist. Nun stimmt aber schon die Voraussetzung der Frage nicht; die subjektive Wertfunktion kann durchaus auch mögliche Handlungen und Umstände decken, da man ja Handlungen und Umstände sehr wohl als Komponenten der möglichen Konsequenzen ansehen kann. Ein solcher Schachzug wird jedenfalls vom Grundmodell nicht ausgeschlossen, auch wenn er zu Lasten der ohnehin ungenügenden Klarheit des Begriffs der möglichen Konsequenz geht. Handlungen kann man ja vielleicht noch als triviale Konsequenz ihrer selbst ansehen; aber mögliche Umstände, die ja eingestandenermaßen – nach Xs Meinung wenigstens – handlungsunabhängig sind, ebenfalls zu den Bestandteilen möglicher Konsequenzen zu rechnen, stellt zumindest eine unerwartete Dehnung des Begriffs der möglichen Konsequenz dar. Außerdem wären dann mögli-

---

[11] Wenn man die Endlichkeitsannahmen des Grundmodells fallen ließe, bedeutete dies genau genommen eine gewisse Einschränkung; vgl. dazu Anm. 8 zu Kap. 1.

che Umstände in Savage-Modellen – unschönerweise – doppelt repräsentiert; nämlich einmal als Elemente von W und zum anderen als Bestandteile der Elemente von C.

Nun könnte man freilich argumentieren, daß diese Maßnahme insofern nicht nötig sei, als Xs eventuelle subjektive Bewertung der möglichen Umstände für die Lösung seines Entscheidungsproblems irrelevant sei. Denn, so läuft das Argument, da es in Xs Augen nichts gibt, was er zur Realisierung des einen oder anderen möglichen Umstandes beitragen könnte, bleiben subjektive Werte für die Umstände ohne Einfluß darauf, welche Handlungen von X maximalen erwarteten subjektiven Wert haben. Doch hat dieses Argument eine entscheidende Lücke, es setzt stillschweigend voraus, daß die möglichen Umstände von den restlichen Bestandteilen der möglichen Konsequenz *wertmäßig unabhängig* sind. Wenn z.B. für Xs Entscheidung der handlungsunabhängige Zustand des Wetters von Belang ist, so können für sie dennoch Xs subjektive Werte für die verschiedenen Wetterumstände wesentlich sein – dann nämlich, wenn diese subjektiven Werte z.B. von Xs Handlungen abhängen. Und dieser Fall kann durchaus auftreten; wie wichtig einem schönes Wetter ist, hängt oft davon ab, was man gerade eventuell vorhat. Will man also die oben beschriebene, unschöne Maßnahme vermeiden, so muß man mit dem Grundmodell die Annahme machen, daß Umstände und Konsequenzen wertmäßig voneinander unabhängig sind. In den Abschnitten 2.7 und 3.4 werden wir darauf wieder zurückkommen.

Echte Anwendungsbeschränkungen des Grundmodells werden aber erkennbar, wenn wir uns anschauen, wie es Xs quantitative Glaubensdispositionen konzipiert. Augenscheinlich macht das Grundmodell die Annahme, daß sich Xs Glaubensdispositionen gänzlich auf eine Menge handlungsunabhängiger Ereignisse, nämlich auf die Menge der möglichen Umstände beschränken. So ausgedrückt, ist das freilich nicht richtig. In der Repräsentation von Handlungen als Funktionen von W in C sind nämlich noch weitere Glaubensdispositionen von X enthalten, die man sogar mit seinen subjektiven Wahrscheinlichkeiten hätte ausdrücken können. In der Tat hätte man auch durch Handlungen und Umstände bedingte subjektive Wahrscheinlichkeiten für Konsequenzen zulassen und festlegen können, daß es zu jedem $f \in H$ und $w \in W$ genau ein $c \in C$ (nämlich gerade $f(w)$) gibt derart, daß $P(c \mid f, w) = 1$ und $P(c' \mid f, w) = 0$ für alle $c' \in C$ mit $c' \neq c$. Davon ausgehend kann man auch handlungsbedingte Wahrscheinlichkeiten für Konsequenzen berechnen. Es gilt ja:

$$P(c \mid f) \;=\; \sum_{w \in W} P(c \mid f, w) \cdot P(w \mid f) \;=\; \sum_{f(w)=c} P(w)$$

(da natürlich $P(w \mid f) = P(w)$ für alle $f \in H$). Definiert man für jedes $f \in H$ P_f durch $P_f(c) = P(c \mid f)$, so ist also jedes P_f ein Wahrscheinlichkeitsmaß für die Konsequenzen, nämlich gerade das Bildmaß von P bezüglich f.[12] X glaubt dann im Grade $P_f(c)$, daß c eintritt, sofern er f ausführt.

Die entscheidende Voraussetzung des Grundmodells besteht dann – in drei verschiedenen Varianten korrekt ausgedrückt – in der Repräsentierbarkeit der Handlungen als Funktion von W in C bzw. in der Annahme, daß die durch Handlungen und Umstände bedingten Wahrscheinlichkeiten für Konsequenzen nur die trivialen Werte 0 und 1 annehmen, bzw. in der Annahme, daß die handlungsbedingten Wahrscheinlichkeiten für Konsequenzen alle durch Bildmaße eines Wahrscheinlichkeitsmaßes für handlungsunabhängige Ereignisse bezüglich der Handlungen repräsentierenden Funktionen gegeben sind. Für die folgende Diskussion dieser Voraussetzung empfiehlt es sich, sie in zwei Annahmen aufzusplittern: in die *Annahme* 1: „In der Entscheidungssituation von X läßt sich eine Menge disjunkter und erschöpfender und von X probabilistisch beurteilter Ereignisse ausmachen derart, daß sich aus Xs Glaubensgraden bezüglich dieser Ereignisse seine sonstigen Glaubensgrade in der angeführten trivialen Weise ergeben", und in der *Annahme* 2: „Die Ereignisse, von denen in Annahme 1 die Rede ist, sind handlungsunabhängig."

Den Effekt dieser Annahmen kann man auch so beschreiben, daß durch sie die einzig relevanten quantitativen Glaubensdispositionen von X aus dem diffusen Bereich der Konsequenzen herausgehalten und ausschließlich auf eine Menge handlungsunabhängiger Ereignisse konzentriert werden. Leicht vergröbernd, aber doch prägnant läßt sich das Grundmodell also insgesamt dadurch charakterisieren, daß es in Entscheidungssituationen eine säuberliche Trennung zwischen Glaubens- und Wünschensbereich vollzieht – hier Umstände mit Wahrscheinlichkeiten, da Konsequenzen mit subjektiven Werten –, die erst durch die Berechnung der erwarteten subjektiven Werte der möglichen Handlungen wieder miteinander in Berührung kommen.

Doch läßt sich diese Trennung immer durchführen? Oder gibt es Entscheidungssituationen, in denen die Annahmen 1 und 2 nicht zutreffen? Betrachten wir

---

[12] Zum Begriff des Bildmaßes vgl. etwa Bauer (1968), §7.

stellvertretend für viele andere Fälle ein ganz simples Beispiel: *X* sitze in einer Prüfung und habe eins von zwei vorgelegten Themen zu bearbeiten. Es sei dabei angenommen, daß es *X* nur auf das Bestehen der Prüfung ankommt . Ohne tiefergehende Analyse würde man doch sagen, daß *X* dasjenige Thema wählt, bei dem er die Aussichten, durch die Prüfung zu kommen, höher einschätzt. Wie aber läßt sich diese Situation im Grundmodell beschreiben?

Als die *X* offenstehenden Handlungen würde man natürlich seine Bearbeitung des ersten bzw. des zweiten Themas ansehen, und als mögliche Konsequenz bieten sich „Bestehen" und „Durchfallen" an. Worin sollen aber die möglichen Umstände dieser Entscheidungssituation bestehen? Die Möglichkeit, nur einen einzigen Umstand, dessen Wahrscheinlichkeit dann 1 beträgt, herauszunehmen, ist durch die Annahme 1 verbaut: *Xs* handlungsbedingte Wahrscheinlichkeiten für die Konsequenzen lassen sich daraus keinesfalls entnehmen. Und die andere naheliegende Möglichkeit, die möglichen Umstände mit den Konsequenzen gleichzusetzen, scheitert offensichtlich an der Annahme 2.

Was also tun? Man könnte die Annahmen 1 und 2 durch eine genauere Analyse der Situation zu erfüllen suchen. Man könnte die möglichen Handlungen nicht nur nach der Themenwahl, sondern nach dem genauen Wortlaut von *Xs* Arbeit differenzieren. Man könnte in die möglichen Umstände die Anforderungen des Korrektors an eine Prüfungsarbeit, seine Stimmung während der Beurteilung und dergleichen einbeziehen. Auf diese Weise könnte man schließlich zu Handlungen und Umständen gelangen, die zusammen jeweils eindeutig eine der beiden Konsequenzen bestimmen. So verdienstvoll eine solche Analyse wäre, sie bedeutete aber doch das Eingeständnis, daß man die obige, einfache, intuitive Überlegung im Grundmodell nicht direkt nachvollziehen kann. Die Lage scheint hoffnungslos; doch es führen drei Wege aus ihr heraus. Zwei davon führen in ihrer Geradlinigkeit ungeschickterweise geradewegs aus dem Grundmodell heraus; während der dritte, ein ganz raffinierter, noch innerhalb des Grundmodells verbleibt.

Die zwei naheliegenden, aber systemsprengenden Möglichkeiten bestehen darin, eine der beiden Annahmen 1 und 2 fallen zu lassen. Geben wir die Annahme 1 auf, so brauchen *Xs* handlungsbedingte Wahrscheinlichkeiten für Konsequenzen nicht bereits durch seine Wahrscheinlichkeiten für Umstände festgelegt zu sein. Das obige Beispiel läßt sich dann leicht mit den schon genannten Handlungen und Konsequenzen und dem sicheren Ereignis als einzigem Umstand beschreiben. Freilich ist der Verzicht auf die Annahme 1 ein drastischer Schritt: Die für das

Grundmodell charakteristische Trennung von Glaubens- und Wünschensdispositionen muß dann ganz anders aufgezogen werden, wobei die möglichen Umstände die ihnen vom Grundmodell zugedachte Rolle verlieren. In der Tat entfällt mit der Annahme 1 jegliche Motivation für eine Auszeichnung der möglichen Umstände als eines besonderen Bestandteils einer Entscheidungssituation.

Weniger gravierend scheint dagegen die Aufgabe der Annahme 2 zu sein. *X*s subjektive Wahrscheinlichkeiten für die Umstände können dann ebenfalls von seinen Handlungen abhängen, und zu einer adäquaten Beschreibung des obigen Beispiels gelangen wir, indem wir die möglichen Umstände einfach mit den möglichen Konsequenzen (Bestehen oder Durchfallen) identifizieren. In der Tat wurde dieses Vorgehen verschiedentlich als naheliegende Verallgemeinerung des Grundmodells empfunden.[13] In Wirklichkeit jedoch wird das Grundmodell auch dadurch aus seinen Angeln gehoben. Die an sich geringfügige Tatsache, daß die Bezeichnung „Umstand“ nun eigentlich nicht mehr angemessen ist, ist dafür ein Indiz. Sie weist darauf hin, daß sich nach dem Verzicht auf die Annahme 2 keine klare Grenze mehr zwischen möglichen Umständen und Konsequenzen finden läßt. In unserem Beispiel kam es ja sogar zu der Merkwürdigkeit, daß die möglichen Umstände und Konsequenzen miteinander identisch sind. Kraß ausgedrückt, wird dann nur mehr zwischen als mögliche Umstände bezeichneten Konsequenzen beliebiger handlungsbedingter Wahrscheinlichkeiten und Konsequenzen mit durch Handlungen und Konsequenzen erster Sorte bedingten 0-1-Wahrscheinlichkeiten unterschieden, und das wirkt bloß noch fadenscheinig. Außerdem bricht auch mit dem Verzicht auf Annahme 2 die Trennung des Grundmodells von Glaubens- und Wünschensbereich in sich zusammen, da dann das auf S. 47 angeführte Argument für die Irrelevanz subjektiver Werte für die möglichen Umstände jeglicher Grundlage entbehrt.

Beide Vorgehensweisen werden also unserem Prüfungsbeispiel gerecht, erschüttern aber das Grundmodell in seinen Grundfesten. Fishburn (1964) hat als erster die Konsequenzen aus diesen Überlegungen gezogen, auf die möglichen Umstände als gesondertem Bestandteil einer Entscheidungssituation gänzlich verzichtet und beliebige handlungsbedingte Wahrscheinlichkeiten für Konsequenzen zugelassen. Die Konzeption, zu der er auf diese Weise gelangt, ist dabei im Grunde noch einfacher als das Grundmodell. Aber darauf werden wir in Abschnitt 2.4 noch ausführlicher zurückkommen.

---

[13] So etwa von Jeffrey (1965), Kap. 1.

Der dritte angekündigte Weg besteht in einem ganz cleveren Kniff, den sich Fishburn (1964), S. 34, zur Rettung des Grundmodells ausgedacht hat. Mit ihm kann man das obige Beispiel doch noch mit dem Grundmodell beschreiben, ohne einen Haufen neuer Dinge wie die Eigenschaften des Korrektors und dergleichen einzubeziehen. Man nehme als die möglichen Umstände einfach folgende vier Ereignisse:

w_1: Wenn *X* Thema 1 wählte, so bestünde er die Prüfung; wenn er Thema 2 wählte, so bestünde er die Prüfung ebenfalls.
w_2: Wenn *X* Thema 1 wählte, so bestünde er die Prüfung; wenn er dagegen Thema 2 wählte, so fiele er durch.
w_3: Wenn *X* sich zu Thema 1 entschiede, so fiele er durch; bei Thema 2 hingegen käme er durch.
w_4: Welches Thema *X* auch nähme, er fiele durch.

w_1, w_2, w_3, und w_4 sind handlungunabhängig, bestimmen zusammen mit den Handlungen eindeutig jeweils eine Konsequenz und qualifizieren sich so als mögliche Umstände im Sinne des Grundmodells. Dieser Gedankengang läßt sich sogar ausnutzen, im Endlichen die formale Äquivalenz des Grundmodells und der Konzeption von Savage mit der Konzeption von Fishburn und der von Krantz und Luce zu beweisen.[14] So hätten wir es also doch noch geschafft, unser so widerspenstiges Beispiel in das Prokrustesbett des Grundmodells zu zwängen?

Bei näherem Hinsehen erweist sich dieser Weg allerdings als wenig attraktiv. Es ist schon unangenehm, daß man es auf diese Weise mit einer gewaltigen Zahl an möglichen Umständen zu tun bekommt: bei m Handlungen und r Konsequenzen benötigt man offensichtlich r^m mögliche Umstände.[15] Entscheidend scheint mir aber folgende Überlegung zu sein: Keines der Wenn-Sos in $w_1, ..., w_4$ darf man als materiale Implikation verstehen. Denn verstünde man sie als materiale Implikation, so wäre w_1 mit „*X* besteht die Prüfung" und w_2 mit „*X* besteht die Prüfung genau dann, wenn *X* Thema 1 wählt" äquivalent. w_1 und w_2 wären dann weder handlungsunabhängig, noch schlössen sie einander aus. Ähnliches gälte natürlich für die nur stilistisch variierten w_3 und w_4. Man muß also die Wenn-Sos schon als subjunktive

---

[14] S. dazu etwa Krantz et. al. (1971), S. 414ff., und diese Arbeit S. 69.

[15] Dieser Sachverhalt ist besonders Krantz und Luce ein Dorn im Auge; s. Krantz et. al. (1971), S. 416.

Konditionalsätze verstehen, und so waren sie ja auch formuliert. Damit geraten wir aber in ganz unsichere Gewässer. Schon daß wir die Entscheidungssituation mit subjunktiven Konditionalen beschreiben müssen, ist unangenehm genug. Aber wir haben es ja dann auch noch mit subjektiven Wahrscheinlichkeiten für subjunktive Konditionale und Konjunktionen aus ihnen zu tun, und da geraten wir in ganz finstere Regionen.[16] Wir sind also doch wohl zu dem Eingeständnis gezwungen, daß man das Grundmodell auf solche Situationen wie der in unserem Prüfungsbeispiel nicht direkt anwenden kann.

Sowohl die qualitativen Glaubensgrade wie auch„ sofern man den Begriff der möglichen Konsequenz nicht arg weit faßt, die quantitativen Wünschensgrade des Entscheidenden müssen also speziellen Bedingungen genügen, um in einem Savage-Modell repräsentiert werden zu können. Darüber war sich im übrigen Savage völlig im klaren, als er fragte:

> „Is it good, or even possible, to insist, as this preference theory does, on a usage in which acts are without influence on events and events without influence on well-being?“ (Savage (1967), S. 307f.)

Betrachten wir nun, was Savage zur Anwendungsfrage sagt.

---

[16] Es liegt nahe, die Wahrscheinlichkeit eines subjunktiven Konditionals „wenn *A* der Fall wäre, so wäre *B* der Fall“ als bedingte Wahrscheinlichkeit $P(B \mid A)$ aufzufassen. Allgemein geht das jedoch nur unter fatal speziellen Voraussetzungen, wie Lewis (1976), S. 300-302, zeigt. Außerdem liefe das der Intention des dritten Weges ganz zuwider. Was wir in unserem Beispiel dann hätten, wären ja gerade wieder handlungsbedingte Wahrscheinlichkeiten für Konsequenzen und nicht nicht-bedingte Wahrscheinlichkeiten für Umstände. Im übrigen ist es auch bei dieser Interpretation noch unklar, wie man zu Wahrscheinlichkeiten für Konjunktionen aus subjunktiven Konditionalen kommt.

2.3 *Small Worlds*

Savage ist, soweit mir bekannt ist, einer der wenigen, der auf die mit der Anwendungsfrage verbundene theoretische Herausforderung ausführlicher eingegangen ist. Seine Ansicht zu dieser Frage nimmt freilich zunächst eine eigenwillige Wendung:

> „In application of the theory, the question will arise as to which world [Möglichkeitsraum] to use in a given context. Thus, if the person is interested in the only brown egg in a dozen, should that egg or the whole dozen be taken as the world? It will be seen as the theory is developed that in principle no harm is done by taking the larger of the two worlds as a model of the situation. One is therefore tempted to adopt, one and for all, one world sufficiently large, say example 7.“ (S. 9).

Beispiel 7 ist dabei ein Beispiel für einen Gegenstand, über den der Handelnde *X* nicht genau Bescheid weiß, nämlich:

> „The exact and entire past, present and future history of the universe, understood in any sense, however wide.“ (S. 8)

Savage stellt sich also im Prinzip auf den Standpunkt, daß sein Modell bezüglich einer Person *X* zunächst nur eine universale Anwendung hat, in der *X* die Gestaltung seines gesamten zukünftigen Lebens zu entscheiden hat. Natürlich erkennt Savage selbst, daß dies, so extrem formuliert, eine völlig unsinnige Idealisierung und kein Standpunkt ist, mit dem man sich zufrieden geben darf. Die Herausdestillierung von „small worlds“, also von begrenzten Entscheidungssituationen angemessenen Möglichkeitsräumen, aus der einen „grand world“ ist seiner Meinung nach allerdings womöglich

> „a matter of judgment and experience about which it is impossible to enunciate complete and sharply defined general principles“ (S. 16),

aber auch

> „an operation in which we all have much experience, and one in which there is in practice considerable agreement.“ (S. 16f.)

Dennoch mißt Savage dieser Frage größte theoretische Bedeutung zu:

> „Any claim to realism made by this book – or indeed by almost any theory of personal decision of which I know – is predicted on the idea that some of the individual decision situations into which actual people tend to subdivide the single grand decision do recapitulate in microcosm the mechanism of the idealized grand decision.“ (S 83)

Das Problem ist also

> „to say as clearly as possible what constitutes a satisfactory isolated decision situation“ (S. 83),

und sein Lösungsweg besteht darin,

> „to talk in terms of the grand situation – tongue in cheek – and in those terms to analyse and discuss the isolated decision situations.“ (S. 83)

Daß er aber immer noch auf die umfassende Situation rekurrieren muß, scheint ihm freilich großen Kummer zu bereiten – unnötigerweise meines Erachtens. Denn daß eine Entscheidungssituation isolierbar ist, heißt ja, grob gesagt, daß alles außerhalb der Situation Gelegene vom Standpunkt des Entscheidenden aus nichts mit ihr zu tun hat. Wie aber sollte man dies von innen heraus feststellen können? Ein Rückgriff auf die umfassende Situation erscheint da ganz unvermeidlich; er sollte aber so sparsam wie möglich ausfallen, und gerade darum bemüht sich ja Savage.

Um nun die Definition einer „kleinen Welt“, einer begrenzten Entscheidungssituation, bei Savage verständlich zu machen, wollen wir zunächst ein kleines Beispiel betrachten. Vorher möchte ich dem Leser jedoch davor warnen, daß dieser Abschnitt zunehmend mühselig wird, was, wie ich hoffe, weniger an mir als an Savage selbst liegt. Wem dieser Abschnitt zu mühselig wird, kann ihn auch überschlagen; er wird es allenfalls im Abschnitt 3.6 vermissen, ihn nicht gelesen zu haben. Nun aber das Beispiel:

Unser *X* hat eine Geliebte und überlegt sich nun, ob er ihr einen Smaragd- oder einen Saphirring schenken soll (Diamanten sind ihm doch etwas zu teuer). Leider kennt er ihren Schmuck noch nicht. Seine Entscheidungssituation ist also durch die folgenden möglichen Umstände gekennzeichnet: Seine Geliebte besitzt bereits einen Smaragd- und einen Saphirring bzw. einen Smaragd-, aber keinen Saphirring bzw. keinen Smaragd-, aber einen Saphirring bzw. weder einen Smaragd- noch einen Saphirring. Und die möglichen Konsequenzen sind: Seine Geliebte wird nun auch einen Smaragd- bzw. Saphirring bzw. noch einen weiteren Smaragd- bzw. Saphirring besitzen. So weit die arg primitiv konzipierte „Stecknadelkopfwelt" unseres *X*.

Diese ist nun Savage zufolge in die „große Welt" unseres *X* einzubetten. Doch seien wir diskret genug, nun nicht allzu tief in *X*s Liebesleben einzudringen. Enthüllen wir lediglich, daß er zu seinen Leidwesen auch nicht weiß, ob seine Geliebte schon verheiratet ist oder nicht, und erweitern wird so die eben beschriebene Stecknadelkopfwelt zu einer „Fingerhutwelt": Sie enthält bereits acht mögliche Umstände, nämlich die obigen vier einmal mit dem Zusatz, daß seine Geliebte verheiratet ist, und einmal mit dem Zusatz, daß sie es nicht ist. Die *X* offenstehenden Handlungen sind natürlich dieselben wie oben. Die möglichen Konsequenzen sind aber auch komplizierter; in sie gehe jetzt noch ein, ob seine Geliebte sein Geschenk, ihr wievielter Smaragd- bzw. Saphirring es auch sei, nur heimlich tragen kann oder nicht. Doch sollten wir ihn seinen Sorgen überlassen und uns stattdessen dem genauen Verhältnis zwischen der Stecknadelkopfwelt und der Fingerhutwelt als Stellvertreter der „großen Welt" zuwenden:

Das Verhältnis der möglichen Umstände beider Welten ist schnell analysiert. Die Umstände der Stecknadelkopfwelt sind einfach Ereignisse, und zwar disjunkte und in ihrer Gesamtheit erschöpfende Ereignisse, in der Fingerhutwelt. Savage beschreibt daher den Möglichkeitsraum W_0 der Stecknadelkopfwelt als eine Zerlegung des Möglichkeitsraums W_1 der Fingerhutwelt. Für jedes $w \in W_1$ wollen wir dabei dasjenige Zerlegungselement von W_0, das w enthält, mit $\overline{w}$ bezeichnen.

Interessanter ist das Verhältnis der möglichen Konsequenzen beider Welten: Die möglichen Konsequenzen in einer Welt – sei sie die Stecknadelkopf-, die Fingerhut- oder die große Welt – sind von ihr selbst aus betrachtet immer Dinge, die mit keinerlei Unsicherheiten behaftet sind. Vom Standpunkt der umfassenden Entscheidungssituation aus sind aber die möglichen Konsequenzen in der begrenzteren Entscheidungssituation prinzipiell unsichere Dinge: Der subjektive Wert des so-und-so-vielten Smaragd- bzw. Saphirrings hängt eben auch von davon ab, ob er nur heimlich getragen werden kann oder nicht, und dies hängt wiederum davon ab, ob die Geliebte verheiratet ist oder nicht. D.h. vom Standpunkt der Fingerhutwelt aus kann man die Konsequenzen in der Stecknadelkopfwelt als Funktionen auffassen, die für verschiedene Umstände der Fingerhutwelt als Argumente verschiedene

Fingerhutwelt-Konsequenzen als Werte annehmen, und somit formal als Handlungen in der Fingerhutwelt gleichsetzen. Und genau dies tut Savage.

So formal korrekt diese Gleichsetzung ist, so unvorsichtig scheint mir freilich Savages Usus zu sein, „small world consequence" und „grand world act" synonym zu verwenden. Denn dies ist nur bezüglich des formalen Sinnes von „Handlung" zulässig, nach dem *jede* Funktion vom Möglichkeitsraum in die Menge der Konsequenzen eine Handlung ist. Bei der Anwendung von Savages Theorie, für die ja die normale Bedeutung von „Handlung" ausschlaggebend ist, kann dieser Sprachgebrauch aber nur verwirren, da die Konsequenz in der begrenzten Entscheidungssituation den Handlungen (im üblichen Sinne) aus der umfassenderen Situation denkbar unähnlich sein können – auch wenn man unter „Handlung" ganze Handlungsabläufe oder gar Strategien versteht. In der Tat zeigt sich dies schon an unserem Beispiel. Diesem Einwand will ich in der kommenden Definition 2 dadurch Rechnung tragen, daß ich nicht fordere, daß die Menge der Konsequenzen in der Stecknadelkopfwelt eine Teilmenge der Menge der Fingerhutwelt-Handlungen ist.

Die Stecknadelkopfwelt-Handlungen sind dann nach bewährter Manier formal einfach wieder als Funktionen zu repräsentieren, die jedem möglichen Umstand der Stecknadelkopfwelt eine Stecknadelkopfwelt-Konsequenz zuordnen. Wichtig ist nun, daß jeder Handlung in der Stecknadelkopfwelt eine Handlung aus der Fingerhutwelt entspricht: Sei f_0 eine Stecknadelkopfwelt-Handlung und w ein Fingerhutwelt-Umstand. Wenn w vorliegt, so liegt auch der Stecknadelkopfwelt-Umstand $\overline{w}$ vor. f_0 führt dann also zur Stecknadelkopfwelt-Konsequenz f_0 $(\overline{w})$; bezeichnen wir sie mit c_0. c_0, vom Standpunkt der Fingerhutwelt aus betrachtet eine unsichere Konsequenz, führt aber selbst wiederum zur Fingerhutwelt-Konsequenz $c_1 = c_0(w)$. Und so entspricht der Stecknadelkopfwelt-Handlung f_0 eine Fingerhutwelt-Handlung f_0*, die jedem w die Fingerhutwelt-Konsequenz $[f_0(\overline{w})](w)$ zuordnet. Natürlich werden wir verlangen, daß zu jeder Handlung der Stecknadelkopfwelt die entsprechende Handlung in der Fingerhutwelt enthalten ist, da man ja ansonsten kaum sagen könnte, die Stecknadelkopfwelt sei eine Verkleinerung der Fingerhutwelt.

Schließlich ist noch ein Wort zu den subjektiven Wahrscheinlichkeiten und Werten zu sagen. Da die Stecknadelkopfwelt-Ereignisse alle auch Ereignisse der Fingerhutwelt sind, sollen die subjektiven Wahrscheinlichkeiten für sie in beiden Welten natürlich dieselben sein. Bei den subjektiven Werten ist die Sachlage komplizierter. Die Reckndelkopfwelt-Konsequenzen sind ja vom Standpunkt der Fingerhutwelt aus unsicher und haben so keinen subjektiven Wert. Ihr subjektiver Wert ist daher mit ihrem *erwarteten* subjektiven Wert in der Fingerhutwelt zu identifizieren. Der erwartete subjektive Wert der Stecknadelkopfwelt-Handlungen wird dann wie üblich berechnet.

Damit haben wir alle Elemente für die folgende Definition beieinander:

Definition 2: $\langle W_0, C_0, H_0, P_0, V_0, U_0\rangle$ ist ein *potentielles Savage-Submodell von* $\langle W_1, C_1, H_1, P_1, V_1, U_1\rangle$ genau dann, wenn gilt:

(1) $\langle W_1, C_1, H_1, P_1, V_1, U_1\rangle$ ist ein Savage-Modell,

(2) W_0 ist eine Zerlegung von W_1,

(3) C_0 ist eine nicht leere Menge von Funktionen von W_1 in C_1,

(4) H_0 ist eine nicht leere Menge von Funktionen vom W_0 in C_0 derart, daß für jedes $f_0 \in H_0$, die durch $f_0{}^*(w)=[f_0(\overline{w})](w)$ definierte Funktion $f_0{}^*$ von W_1 in C_1 ein Element von H_1 ist,

(5) P_0 ist das Wahrscheinlichkeitsmaß auf Pot(W_0), für das für alle $\overline{w} \in W_0$ $P_0(\{\overline{w}\}) = P_1(\overline{w})$,

(6) V_0 ist die durch $V_0(c_0) = \sum_{w \in W_1} V_1(c_0(w)) \cdot P_1(w)$ für alle $c_0 \in C_0$ definierte Funktion von C_0 in $\mathbf{R}$,

(7) U_0 ist die durch $U_0(f_0) = \sum_{\overline{w} \in W_0} V_0(f_0(\overline{w})) \cdot P_0(\overline{w})$ definierte Funktion von H_0 in $\mathbf{R}$.

Natürlich gilt: Wenn Δ_0 ein potentielles Savage-Submodell von Δ_1 ist, so ist Δ_0 ein Savage-Modell.

Es liegt die Frage nahe, ob die Relation „ist ein potentielles Savage-Submodell von" reflexiv und transitiv ist. Schon aufgrund der Konstruktion der Konsequenzen im potentiellen Submodell muß die Antwort darauf trivialerweise „nein" lauten. Diese Frage soll uns aber gleich noch eingehender beschäftigen.

Vorher soll freilich erläutert werden, was es mit dem Attribut „potentiell" auf sich hat. Dies habe ich in Definition 2 hinzugefügt, weil Definition 2 nur die formale Bedingung dafür enthält, daß man von der Stecknadelkopfwelt sinnvoll als einer Verkleinerung der Fingerhutwelt reden kann. Damit aber die Stecknadelkopfwelt wirklich eine getreue Verkleinerung der Fingerhutwelt ist, muß außerdem noch folgendes gelten: Da die Handlung f_0 aus der Stecknadelkopfwelt und die entsprechende Handlung f_0^* aus der Fingerhutwelt, definiert durch $f_0^*(w) = [f_0(\overline{w})](w)$, dieselbe Handlung repräsentieren, muß auch ihr erwarteter subjektiver Wert in beiden Welten derselbe sein. Andernfalls kann von einer getreuen Verkleinerung nicht die Rede sein. Man könnte meinen, daß diese Bedingung immer erfüllt ist. Dem ist aber, wie wir sehen werden, keineswegs so. Zunächst sei jedoch definiert:

Definition 3: $\Delta_0 = \langle W_0, C_0, H_0, P_0, V_0, U_0\rangle$ ist ein *Savage-Submodell* von $\Delta_1 = \langle W_1, C_1, H_1, P_1, V_1, U_1\rangle$ genau dann, wenn gilt:
(1) Δ_0 ist ein potentielles Savage-Submodell von Δ_1,
(2) für jedes $f_0 \in H_0$ ist $U_0(f_0) = U_1(f_0^*)$.

Natürlich interessiert uns jetzt, wann ein potentielles Savage-Submodell auch ein echtes Savage-Submodell ist. Die Antwort darauf gedenke ich auf einem freilich lohnenden Umweg zu geben. Denn der Umweg führt uns auch zur Beantwortung der Frage, ob denn wenigstens etwas gilt, was der Reflexivität und der Transitivität der (potentiellen) Submodell-Relation nahekommt.[17]

Eine Antwort bezüglich der Reflexivität gibt uns der folgende einfache

Satz 1: Sei $\Delta_1 = \langle W_1, C_1, H_1, P_1, V_1, U_1\rangle$ ein Savage-Modell. Sei $W_0 = \{\{w\} \mid w \in W_1\}$, d.h. also $\overline{w} = \{w\}$ für jedes $w \in W_1$, und sei P_0 gemäß Definition 2, (5), festgelegt. Sei für jedes $c \in C_1$ k_c die durch $k_c(w) = c$ definierte konstante Funktion von W_1 in C_1, sei $C_0 = \{k_c \mid c \in C_1\}$, und sei V_0 die durch $V_0(k_c) = V_1(c)$ auf C_0 definierte Funktion. Sei H_0 die Menge aller Funktionen f_0 von W_0 in C_0, für die es ein $f_1 \in H_1$ gibt, so daß für alle $w \in W_1$ $f_0(\overline{w}) = k_c$ genau dann, wenn $f_1(w) = c$. Sei U_0 schließlich dadurch auf H_0 definiert, daß für je zwei solche $f_0 \in H_0$ und $f_1 \in H_1$ $U_0(f_0) = U_1(f_1)$ gilt. dann ist $\Delta_0 = \langle W_0, C_0, H_0, P_0, V_0, U_0\rangle$ ein Savage-Submodell von Δ_1.

Satz 1 besagt also gerade so etwas wie die Reflexivität der Submodell- und damit auch der potentiellen Submodell-Relation. Denn das in Satz 1 konstruierte Δ_0 ist zu Δ_1 isomorph – genauer: Damit Δ_0 ein Submodell von Δ_1 sein kann, müssen

[17] Die folgenden Ausführungen sind im Grunde simpel, erfordern aber ein in der einen oder anderen Form leider unvermeidlichen Gewurstel im Hinblick auf Notation und Formulierung, für das ich mich im voraus entschuldigen möchte.

die Umstände in Δ_0 eine Zerlegung von W_1 bilden; wir nahmen da die feinste Zerlegung, die es gibt. Und ferner müssen dazu die Konsequenzen in Δ_0 Funktionen von W_1 in C_1 sein; daher haben wir in Δ_0 statt der sicheren Konsequenzen $c \in C_1$ konstante Funktionen k_c als Konsequenzen genommen, die wegen ihrer Konstanz ebenfalls sichere Konsequenzen repräsentieren. H_0, P_0, V_0 und U_0 sind dann einfach die entsprechenden Modifikationen von H_1, P_1, V_1 und U_1. Δ_0 repräsentiert also genau dieselbe Entscheidungssituation wie Δ_1.

Läßt sich nun, wenn Δ_2 ein Submodell von Δ_3 und Δ_1 ein Submodell von Δ_2 ist, auf ebenso natürliche Weise ein zu Δ_1 isomorphes Entscheidungsmodell Δ_0 konstruieren, das Submodell von Δ_3 ist? Die vom folgenden Satz 2 geleistete Antwort fällt nur bedingt positiv aus – bedingt insofern, als eine zusätzliche Voraussetzung erfüllt sein muß, damit das aus Δ_1 konstruierte Modell Δ_0 auch nur ein potentielles Submodell von Δ_3 ist.

Satz 2: Sei $\Delta_2 = \langle W_2, C_2, H_2, P_2, V_2, U_2\rangle$ ein (potentielles) Submodell von $\Delta_3 = \langle W_3, C_3, H_3, P_3, V_3, U_3\rangle$ und $\Delta_1 = \langle W_1, C_1, H_1, P_1, V_1, U_1\rangle$ ein (potentielles) Savage-Submodell von Δ_2. Definieren wir nun $\Delta_0 = \langle W_0, C_0, H_0, P_0, V_0, U_0\rangle$ in drei Schritten:

(1) Zunächst sei $W_0 = W_1$ und $P_0 = P_1$.[18]

(2) Nun kommt die entscheidende Definition von C_0: C_1 enthält lauter Funktionen von W_2 in C_2, wobei C_2 selbst aus Funktionen von W_3 in C_3 besteht. C_0 muß aber auch aus Funktionen von W_3 in C_3 bestehen. legen wir also fest: Für $c \in C_1$ sei k_c diejenige Funktion von W_3 in C_3, für die für alle $w \in W_3$ $k_c(w) = [c(\overline{w})](w)$ ist; es sei dann $C_0 = \{k_c \mid c \in C_1\}$.

(3) Es sind nur noch H_0, V_0 und U_0 entsprechend festzulegen: Sei V_0 die durch $V_0(k_c) = V_1(c)$ definierte Funktion von C_0 in $\mathbf{R}$; sei H_0 die Menge aller Funktionen f_0 von W_0 in C_0, für die es ein $f_1 \in H_1$ gibt, so daß $f_0(\overline{\overline{w}}) = k_c$ genau dann, wenn $f_1(\overline{\overline{w}}) = c$; und sei U_0 dann dadurch festgelegt, daß für je zwei solche $f_0 \in H_0$ und $f_1 \in H_1$ $U_0(f_0) = U_1(f_1)$ gilt.

Sei zuletzt $\mathcal{A}$ die Menge der Ereignisse $A \subseteq W_3$, für die es ein $c_1 \in C_1$ und ein $c_2 \in C_2$ gibt, so daß $c_1(\overline{w}) = c_2$ für genau die w aus A, und sei entsprechend $\mathcal{B}$ die Menge der Ereignisse $B \subseteq W_3$, für die es ein $c_2 \in C_2$ und ein $c_3 \in C_3$ gibt, so daß $c_2(w) = c_3$ für genau die w aus B. Sind dann $\mathcal{A}$ und $\mathcal{B}$ bzgl. P_3 unabhängig, d.h. gilt

[18] Das ist genaugenommen nicht ganz korrekt; aber wem das aufgefallen ist, der wird es auch korrigieren können.

für alle $A \in \mathcal{A}$ und $B \in \mathcal{B}$ $P_3(A \cap B) = P_3(A) \cdot P_3(B)$,[19] so ist Δ_0 ein (potentielles) Savage-Submodell von Δ_3.

Ich will hier Satz 2 weder als Satz über potentielle Submodelle noch als Satz über Submodelle beweisen. Der Beweis ist elementar, erfordert aber einige Rechnerei. In der Zeit, in der Leser ihn hier nachvollzogen hätte, hätte er ihn auch selbst erstellt.

In kurzen Worten besagt Satz 2 also: Wenn Δ_1 ein Submodell von Δ_2 und Δ_2 ein Submodell von Δ_3 ist, so läßt sich jedenfalls dann ein Submodell von Δ_3 angeben, das dieselbe Entscheidungssituation beschreibt wie Δ_1, wenn die von Δ_3 aus betrachtet unsicheren Konsequenzen aus Δ_1 und Δ_2 bzw. die von ihnen erzeugten Ereignisse probabilistisch unabhängig sind.

Er sei zwischendurch bemerkt, daß Savages Konstruktion der Konsequenzen des Submodells, wenngleich in seinem Rahmen durchaus zwingend und nicht anders zu bewerkstelligen, technisch nicht gerade günstig ist. Die Submodell-Relation wird dadurch technisch etwas kompliziert, Reflexivität und auch nur bedingte Transitivität gehen genau genommen verloren, und die Ersatztheoreme 1 und 2 lassen sich nur recht umständlich formulieren. Was davon zu halten ist, daß die Submodell-Relation nur unter zusätzlichen Voraussetzungen (im laxen Sinne) transitiv ist, wollen wir später ansprechen.

Eine Brücke von Satz 2 zur Frage, wann potentielle Submodelle tatsächlich Submodelle sind, schlägt die folgende Beobachtung:

Satz 3: Sei $\Delta_2 = \langle W_2, C_2, H_2, P_2, V_2, U_2 \rangle$ ein Savage-Modell. Sei $W_1 = \{W_2\}$ und P_1 das triviale Wahrscheinlichkeitsmaß auf $\mathrm{Pot}(W_1)$. ($\mathrm{Pot}(W_1)$ enthält ja nur das sichere und das unmögliche Ereignis.) Sei ferner $C_1 = H_2$ und $V_1 = U_2$. Sei schließlich für $c \in C_1$ f_c die Funktion auf W_1, die W_2 c zuordnet, $H_1 = \{f_c \mid c \in C_1\}$ und $U_1(f_c) = V_1(c)$ für alle $c \in C_1$. Dann ist $\Delta_1 = \langle W_1, C_1, H_1, P_1, V_1, U_1 \rangle$ ein Savage-Submodell von Δ_2.

Dieser Sachverhalt ist denkbar trivial, aber auch interessant: Denn Satz 3 besagt, daß man jedes Savage-Modell auf ein Minimalmodell, das nur durch das sichere und das unmögliche Ereignis und somit nur für diese Ereignisse, also praktisch

---

[19] Bündig formuliert besagt diese Voraussetzung also, daß die Funktionen aus C_1 von den Funktionen aus C_2 bzgl. P_3 unabhängig sind.

keine subjektiven Wahrscheinlichkeiten enthält, dessen Konsequenzen die Handlungen des umfassenderen Modells sind, dessen Handlungen gerade wieder diese Handlungen, nur in anderer formaler Gestalt, sind und dessen subjektive Wertfunktion gerade den erwarteten Wert dieser Handlungen angibt. Solche simplen Gebilde sind also auch Savage-Modelle, und jedes Savage-Modell hat ein solches Gebilde als Submodell. Dieser Tatbestand erweist sich für unsere abschließende Beurteilung von Savages Verfahren zur Isolierung von „small worlds“ aus der einen „grand world“ als ganz wesentlich.

Aus Satz 2 und Satz 3 folgt nun der

Satz 4: Sei $\Delta_2 = \langle W_2, C_2, H_2, P_2, V_2, U_2\rangle$ ein potentielles Savage-Submodell von $\Delta_3 = \langle W_3, C_3, H_3, P_3, V_3, U_3\rangle$. Sei $\mathcal{A}$ die Menge der Ereignisse $A \subseteq W_3$, für die es ein $f_2 \in H_2$ und ein $c_2 \in C_2$ gibt, so daß $f_2(\overline{w}) = c_2$ für genau die $w \in A$, und sei entsprechend $\mathcal{B}$ die Menge der Ereignisse $B \subseteq W_3$, für die es ein $c_2 \in C_2$ und ein $c_3 \in C_3$ gibt, so daß $c_2(w) = c_3$ für genau die $w \in B$. Sind dann $\mathcal{A}$ und $\mathcal{B}$ bzgl. P_3 unabhängig,[20] so ist Δ_2 ein Savage-Submodell von Δ_3.

Natürlich kann man Satz 4 direkt und ohne Rückgriff auf die Sätze 2 und 3 (wiederum elementar) beweisen. Zieht man aber diese Sätze heran (indem man Δ_1 aus Δ_2 bzw. Δ_0 aus Δ_3 wie in Satz 3 konstruiert, und dann zeigt , daß Δ_0 gerade das in Satz 2 aus Δ_1 konstruierte Modell ist), so wird verständlich, wieso zur Transitivität der potentiellen Submodell-Relation (im laxen Sinne) und zur Verstärkung potentieller Submodelle zu echten Submodellen ähnliche Unabhängigkeitsannahmen benötigt werden. Der Grund dafür liegt, kurz gesagt, darin: Das potentielle Submodell Δ_2 von Δ_3 ist ja gerade dann ein echtes Submodell von Δ_3, wenn einander entsprechende Handlungen in beiden Modellen denselben erwarteten subjektiven Wert erhalten. Nun hat Δ_2 bzw. Δ_3 nach Satz 3 ein minimales Submodell Δ_1 bzw. Δ_0, in dem im wesentlichen nur noch Handlungen und deren erwartete subjektive Werte auftauchen. Δ_2 ist also genau dann ein Submodell von Δ_3, wenn Δ_1 und Δ_0 wenn schon nicht gleich, so doch isomorph sind; und dies wiederum ist gerade dann der Fall, wenn für Δ_3, Δ_2, und Δ_1 Transitivität im laxen Sinne besteht.

Insgesamt sind wir mit Satz 4 bei einem leicht durchschaubaren Ergebnis gelandet: Hinreichend dafür, daß ein potentielles Savage-Submodell eines Savage-

[20] Bündig formuliert besagt diese Voraussetzung also, daß die Funktionen auf H_2 von den Funktionen aus C_2 bzgl. P_3 unabhängig sind.

Modells ein echtes Savage-Modell ist, ist die Bedingung, daß die Ereignisse im Submodell von den Konsequenzen des Submodells bzw. von den von ihnen erzeugten Ereignissen probabilistisch unabhängig sind.

Wenden wir dieses Ergebnis auf unser Beispiel an: Die Stecknadelkopfwelt unseres *X* ist jedenfalls dann eine zulässige Verkleinerung seiner Fingerhutwelt, wenn er die Ereignisse der Stecknadelkopfwelt – z.B. „die Geliebte besitzt weder einen Smaragd- noch einen Saphirring“ – als unabhängig von den Stecknadelkopf-Konsequenzen bzw. den von ihnen erzeugten Fingerhutwelt-Ereignissen – z.B. „die Geliebte ist unverheiratet“ – ansieht.

Savage selbst ist mit diesem Ergebnis nicht ganz glücklich. Denn um diese Unabhängigkeit feststellen zu können, muß man schon die subjektiven Wahrscheinlichkeiten für die Ereignisse der umfassenden Entscheidungssituation kennen. Und bedenkt man, daß es nach Savage in der umfassenden Entscheidungssituation genaugenommen um die Festlegung des gesamten zukünftigen Lebens geht, so ist dies wahrlich viel verlangt. Hätte man jedoch eine qualitative Charakterisierung dieser subjektiven Unabhängigkeit zwischen Ereignissen, so wäre diese Forderung wesentlich gemildert. Aus diesem Grunde hält Savage (1954, S. 91) eine solche Charakterisierung für wünschenswert.

Spätestens mit einer solchen qualitativen Charakterisierung probabilistischer Unabhängigkeit scheint aber Savages Vorschlag zur Abgrenzung einzelner Entscheidungssituationen durchaus überzeugend zu sein; in der Tat könnte man ja in der in Satz 4 angeführten Unabhängigkeitsbedingung nachgerade eine Explikation meiner zu Anfang dieses Abschnitts gebrauchten laxen Formulierung erblicken, daß alles außerhalb der Situation Liegende – vom Standpunkt des Entscheidenden aus – nichts mit ihr zu tun hat.

Dennoch glaube ich, daß Savages Ausführungen über „small worlds“ ganz anders zu beurteilen sind; trotz seiner anders lautenden Ankündigungen schlägt er nämlich meines Erachtens am Ende keine Lösung des Problems vor, wie sich isolierte Entscheidungssituationen aus der „grand world“ heraustrennen lassen. Vielmehr definiert er mit seinen Submodellen so etwa wie getreue Verkleinerung umfassenderer Modelle – in diese Wendung glitten wie ja auch früher schon –, und Verkleinerbarkeit ist ganz etwas anderes als Isolierbarkeit. Bei letzterem geht es darum, aus einem großen Entscheidungsproblem ein Teilproblem auszusondern, das sich vollkommen losgelöst vom Rest des großen Problems behandeln läßt. Bei ersterem dreht es sich dagegen darum, ein großes Entscheidungsproblem zu verkleinern und damit überschaubarer zu machen, ohne daß damit die Absicht verbunden wäre, Teilprobleme herauszulösen. Natürlich bemüht man sich bei der Lösung großer Entscheidungsprobleme um beides: um die Verkleinerung des Problems wie um die Abtrennung von Teilproblemen. Doch ändert dies nichts daran, daß man sich da um Unterschiedliches bemüht.

Deutlichstes Indiz dafür, daß Savage entgegen seiner Absicht mit der getreuen Verkleinerung von Entscheidungsmodellen befaßt ist, ist der Satz 3. Denn von dem dort angegebenen Submodell – ich nannte es minimal, weil es praktisch nur noch Handlungen und deren erwartete subjektive Werte enthält – läßt sich schwerlich sagen, es repräsentiere eine isolierte Entscheidungssituation, die sich aus der umfassenden Entscheidungssituation ausgliedern ließe. In die umfassendere Entscheidungssituation geht ja gerade all das ein, was für die Beurteilung der in ihr zur Entscheidung anstehenden Handlungen relevant ist. Wie können da diese Handlungen allein eine isolierte Entscheidungssituation bilden? Demgegenüber macht es Sinn, wenn wir das Submodell aus Satz 3 als extreme Verkleinerung und Vereinfachung des umfassenderen Entscheidungsmodell ansehen – wird doch ein Entscheidungsmodell nachgerade trivial, wenn man die erwarteten subjektiven Werte der möglichen Handlungen kennt.

So hat denn auch die in Satz 4 auftauchende Unabhängigkeitsbedingung entgegen äußerem Anschein gar nicht die Aufgabe, Isolierungsmöglichkeiten zu markieren. Ihre Quelle liegt vielmehr in den speziellen, in Abschnitt 2.2 aufgezeigten Annahmen Savages über die quantitativen Glaubensdispositionen des Entscheidenden, die sich in seiner Submodell-Relation nicht nur wiederholen, sondern geradezu multiplizieren: Sei nämlich Δ_0 ein Savage-Submodell von Δ_1, wobei Δ_0 und Δ_1 die gleichen Handlungen, wenn auch in unterschiedlicher formaler Gestalt, enthalten mögen. Die Δ_0-Umstände lassen sich dann als Teilumstände von Δ_1 auffassen, die erst kombiniert mit anderen Teilumständen von Δ_1 – nennen wir sie Restumstände – volle Δ_1-Umstände ergeben. Und ebenso kann man die Δ_0-Konsequenzen als Teilkonsequenzen von Δ_1 verstehen, aus denen man erst dann die vollen Δ_1-Konsequenzen erhält, wenn man sie mit weiteren Teilkonsequenzen von Δ_1 – nennen wie sie Restkonsequenzen – paart. Beschreibt man nun den Entscheidenden sowohl mit Δ_1 wie mit Δ_0, so wird nicht einfach wie bei allen Savage-Modellen angenommen, daß mit je einer Handlung und einem vollen Δ_1-Umstand eine volle Δ_1-Konsequenz eindeutig festgelegt ist; vielmehr wird angenommen, daß nach Ansicht des Entscheidenden zunächst jede Handlung zusammen mit einem *Teil*zustand bereits eine *Teil*konsequenz festgelegt und daß dann diese Teilkonsequenz zusammen mit einem *Rest*umstand auch noch eine *Rest*konsequenz eindeutig bestimmt. Dies liegt zum einen daran, daß auch Δ_0 ein Savage-Modell ist, und zum anderen daran, daß die Teilkonsequenzen, d.h. die Δ_0-Konsequenzen, wie die Handlungen als Funktionen von Δ_1-Umständen in Δ_1-Konsequenzen konzipiert sind. Daß man dem Entscheidenden mit Submodellen gleichzeitig solche speziellen Ansichten über die Welt zuschreiben muß, scheint mir ein innerhalb Savages Rahmen zwar konsequenter, insgesamt aber doch unerwünschter Nebeneffekt seiner Methode zur Verkleinerung von Entscheidungsmodellen zu sein.

Im dritten Kapitel werde ich ausführlich auf die hier angesprochenen Dinge noch einmal zurückkommen, und zwar sowohl auf das Problem, auf das Savage abzielte, als auch auf das Problem, das Savage dann traf. Ersteres soll im Abschnitt 3.5 unter dem Stichwort „Zerlegungen von Entscheidungsmodellen" geschehen, zweiteres im Abschnitt 3.6 unter der Überschrift „Reduktionen von Entscheidungsmodellen", wie ich dann statt „Verkleinerung" sagen werde. Unsere Beurteilung Savages bestätigt sich dort insofern, als man zu Savages Submodellen nicht durch Zerlegung, sondern nur durch Reduktion von Entscheidungsmodellen gelangt.

Es ist noch nachzutragen, daß unsere Darstellung von Savages Original doch in einigen Punkten abgewichen ist. Zum ersten will Savage auch in der „kleinen Welt" eine komplette Metrisierung der subjektiven Wahrscheinlichkeiten und Werte durchführen. Dies habe ich aus den bekannten Gründen hier außer acht gelassen.

Zum zweiten – eine technische Bemerkung – gelangt Savage zu einer etwas schwächeren hinreichenden Bedingung dafür, daß ein potentielles Submodell ein echtes Submodell ist: In unserem Satz 4 hätte auch schon die Bedingung hingereicht, daß die von den Funktionen $V_2 \circ f_2$ ($f_2 \in H_2$) erzeugte Ergebnisalgebra $\mathcal{A}'$ und die von den Funktionen $V_3 \circ c_2$ ($c_2 \in C_2$) erzeugte Ergebnisalgebra $\mathcal{B}'$ voneinander unabhängig sind. Und bei Savage tut es statt $\mathcal{B}'$ sogar die von den Differenzen $V_3 \circ c_2 - V_3 \circ c'_2$ ($c_2 \circ c'_2 \in C_2$) erzeugte Algebra. Dies liegt daran, daß er an echte Submodelle nicht wie wir die Forderung stellt, daß eine Handlung im Submodell denselben erwarteten subjektiven Wert hat wie die entsprechende Handlung aus der „großen Welt", sondern nur daran interessiert ist, daß die Präferenzordnung zwischen den Handlungen des Submodells dieselbe ist wie die zwischen den entsprechenden Handlungen aus der „großen Welt".

Drittens kann Savage seine hinreichende Bedingung außerdem auch als notwendig erweisen. Dies rührt daher, daß er, bedingt durch sein Metrisierungsanliegen, H_2 denkbar umfangreich ansetzt, nämlich als die Menge *aller* Funktionen von W_2 in C_2. (Auch unsere Bedingung in der eben genannten schwächeren Version wird notwendig, wenn H_2 eine gewisse Reichheit aufweist – eine Reichheit freilich, die man bei Anwendungen nicht voraussetzen kann.) Jedoch wird der Kern von Savages Vorschlag zur, wie sich herausstellte, getreuen Verkleinerung umfassenderer Entscheidungssituationen durch diese Abweichungen nicht berührt.

2.4 *Die Konzeption von Fishburn*

Die Konzeption von Fishburn (1964) läßt sich am besten als Reaktion auf die bei der Behandlung der Allgemeinheitsfrage in Abschnitt 2.2 (S. 46-52) aufgedeckte Beschränkung des Grundmodells und der Theorie von Savage verstehen. Die Diskussion des Prüfungsbeispiels aus Abschnitt 2.2 (S. 49ff.) zeigte – um dies kurz zu rekapitulieren –, daß man ihm und ähnlichen Situationen am besten dadurch gerecht wird, daß man die Annahme 1 oder die Annahme 2 des Grundmodells (S. 48) fallen läßt. Beide Verfahren haben aber, wie wir feststellten, einen ähnlichen, für das Grundmodell vernichtenden Effekt: Die für das Grundmodell charakteristische Trennung von Glaubens- und Wünschensbereich bricht zusammen, und die möglichen Umstände verlieren ihre spezifische, für das Grundmodell so wesentliche Funktion. Geben wir die Annahme 1 auf, so ist es nicht mehr möglich, sich auf subjektive Wahrscheinlichkeiten für die möglichen Umstände zu beschränken; geben wie die Annahme 2 auf, so ist überhaupt kein entscheidender Unterschied mehr zwischen Umständen und Konsequenzen zu entdecken.

Fishburn zieht nun, wie schon früher angedeutet, aus dieser Sachlage die naheliegende Konsequenz, indem er in seiner Konzeption die möglichen Umstände als gesonderte Elemente einer Entscheidungssituation völlig verschwinden läßt und ganz allgemein handlungsbedingte subjektive Wahrscheinlichkeiten für Konsequenzen zuläßt. Sehen wir uns sein Vorgehen etwas genauer an:

Fishburn geht zum einen davon aus, daß der Entscheidende verschiedene Strategien in Betracht zieht. Fishburn nimmt in seine Konzeption also von vornherein den Begriff der Strategie auf, den er in der folgenden Weise erläutert:

> „In general, a strategy is a plan or program of action that may be adopted by the decision maker and implemented by him and/or by other persons working with or responsible to the decision maker.“ (S. 21)

> „Grammatically, a strategy is a set of imperative clauses, some of which may be modified by conditional phrases. The imperative clauses are prescriptions of things to be done or actions to carry through ...“(S. 23)

Die spezielleren Handlungsabläufe sind dann einfach Strategien, die nur aus kategorischen, nicht bedingten Imperativen bestehen. Und zum zweiten nimmt Fishburn in sein Entscheidungsmodell eine Menge mit den Strategien verknüpfter Konsequenzen herein. Eine Konsequenz ist dabei

> „ a set of things which may be done by the individual and, perhaps, by others working with or for him, and of things which may happen to the individual, or to other people, or to entities with which he is concerned in one way or another. It is relative to a given decision maker in a given decision situation. The statement of a consequence is, in the ideal, a complete description of the individual's future insofar as it is affected by or concerned with the decision situation at hand." (S. 25)

Da die möglichen Umstände als spezielle Komponente einer Entscheidungsituation eliminiert wurden – wenngleich natürlich handlungsunabhängige Ereignisse nach wie vor eine Rolle spielen können, nämlich als Bestandteile der Konsequenzen –, treten bei Fishburn nur durch Strategien bedingte subjektive Wahrscheinlichkeiten des Entscheidenden für die Konsequenzen auf. Die subjektive Wertfunktion des Handelnden erstreckt sich dagegen wie im Grundmodell über die Menge der Konsequenzen.

Damit dürfte die folgende Definition klar sein:

Definition 4: $\langle C, H, P, V, U\rangle$ ist ein *Fishburn -Modell* genau dann, wenn gilt:

(1) C und H sind nicht leere Mengen,
(2) P ist eine auf H definierte Funktion derart, daß für jedes $f \in H$ P_f ein Wahrscheinlichkeitsmaß auf Pot(C) ist,
(3) V ist eine Funktion von C in $\mathbf{R}$,
(4) U ist die Funktion von H in $\mathbf{R}$, für die alle $f \in H$ $U(f) = \sum_{c \in C} V(c) \cdot P_f(c)$.

Die mit Fishburns Konzeption verbundene empirische Behauptung – Fishburn selbst faßt die Entscheidungstheorie allerdings präskriptiv auf – lautet dann: Wird X durch das Fishburn-Modell $\langle C, H, P, V, U\rangle$ charakterisiert, so führt X eine Handlung $f \in H$ aus, für die $U(f)$ maximal ist.

Prüfen wir nun auch Fishburns Konzeption hinsichtlich unserer drei Fragen: Die Allgemeinheitsfrage, die bei der Diskussion des Grundmodells unseren längsten Kommentar herausforderte, ist hier schnell erledigt. Im Fishburn-Modell sollen ja gerade die dort festgestellten Mängel und Beschränkungen des Grundmodells beseitigt werden; und es bewältigt diese Aufgabe erfolgreich. Ganz deutlich zeigt dies der folgende

Satz 5: Sei $\langle W, C, H, P, V, U\rangle$ ein Savage-Modell. Sei $C' = W \times C$ und $H' = H$. Sei für jedes $f \in H'$ P'_f das Wahrscheinlichkeitsmaß auf Pot(C'), für das für alle $w \in W$ und $c \in C$ $P'_f(\langle w, c\rangle) = \begin{cases} P(w), & \text{falls } f(w) = c \\ 0, & \text{falls } f(w) \neq c \end{cases}$, und sei P' die Familie aller dieser P'_f ($f \in H'$). Sei weiter V' auf C' durch $V'(\langle w, c\rangle) = V(c)$ für alle $w \in W$ und $c \in C$ definiert. Sei zuletzt $U'(f) = U(f)$ für alle $f \in H' = H$. Dann ist $\langle C', H', P', V', U'\rangle$ ein Fishburn-Modell.

Von entscheidender Bedeutung ist dabei, daß dieser Satz nicht bloß ein formaler Taschenspielertrick ist, sondern eine *natürliche* Einbettung von Savage-Modellen in Fishburn-Modellen liefert. Dies sieht man daran, daß sich hinter dem einzigen technischen Kunstgriff von Satz 5, nämlich der Identifikation von Fishburn-Modell-Konsequenzen mit Paaren aus jeweils einem Savage-Modell-Umstand und einer Savage-Modell-Konsequenz, einfach eine Konjunktionsbildung versteckt: Daß die Fishburn-Modell-Konzequenz $c' = \langle w, c\rangle$ eintritt, heißt nichts anderes, als daß der Grundmodell-Umstand w vorliegt und die Grundmodell-Konsequenz c eintritt. Die Definition von P' in Satz 5 verdeutlicht so sehr schön die spezielle Annahme des Grundmodells über die Glaubensdispositionen des Handelnden, daß sich die handlungsbedingten Wahrscheinlichkeitsmaße für Konsequenzen alle als Bildmaße *eines* Wahrscheinlichkeitsmaßes auf einer Menge möglicher Umstände darstellen lassen. Und die Definition von V' in Satz 5 bringt gerade die auf S. 46f. festgestellte Beschränkung des Grundmodells bezüglich der subjektiven Wertfunktion zum Ausdruck. Satz 5 weist also drei Verallgemeinerungen des Grundmodells durch die Konzeption von Fishburn auf: Weder brauchen sich die Konsequenzen eines Fishburn-Modells speziell als Paare von Savage-Modell-Umständen und Savage-Modell-Konsequenzen darstellen zu lassen, noch brauchen seine Wahrscheinlichkeitsfunktion und seine subjektive Wertfunktion von der in Satz 5 angenommenen speziellen Gestalt zu sein.

Nun ist allerdings, wie schon auf S. 51 angedeutet, die umgekehrte Einbettung ebenfalls möglich:

Satz 6: Sei $\langle C, H, P, V, U\rangle$ ein Fishburn-Modell. Sei W' die Menge aller Funktionen von H in C und $C' = C$. Sei für jedes $f \in H$ f' diejenige Funktion von W' in C', für die für alle $w \in W'$ $f'(w) = w(f)$, und sei $H' = (f' \mid f \in H)$. Sei ferner ein Wahrscheinlichkeitsmaß P' auf Pot(W') derart gewählt, daß für alle $f \in H$ und $c \in C$ $\sum_{w(f)=c} P'(w) = P_f(c)$. Sei schließlich $V' = V$ und $U'(f) = U(f)$ für alle $f \in H$. Dann ist $\langle W', C', H', P', V', U'\rangle$ ein Savage-Modell.

P' so zu wählen, wie in Satz 6 verlangt, ist immer möglich. Z.B. tut es das dadurch definierte P', daß für alle $w \in W'$ $P'(w) = \prod_{f \in H} P_f(w(f))$.

Satz 6 scheint also zu zeigen, daß Fishburn-Modelle doch keine Verallgemeinerung von Savage-Modellen liefern. Jedoch ist die von Satz 6 geleistete formale Einbettung keine, die man irgendwie natürlich nennen könnte. Denn man muß, wie wir auf S. 51 in der Diskussion des dritten Weges zur Behandlung des Prüfungsbeispiels sahen, die Funktionen, die in Satz 6 als Savage-Modell-Umstände auftreten, als Konjunktionen von subjunktiven Konditionalen auffassen: „Wenn X f_1 täte, so geschähe $w(f_1)$, und wenn X f_2 täte, so geschähe $w(f_2)$, und wenn ...“. Das Fishburn-Modell leistet also genau dann keine Verallgemeinerung des Grundmodells, wenn man bereit ist, im Grundmodell mit subjunktiven Konditionalen und Konjunktionen daraus samt subjektiven Wahrscheinlichkeiten dafür zu arbeiten.

Auf derart ungeklärte Dinge sollte man sich aber, wie gesagt, besser nicht einlassen. Daher können wir trotz Satz 6 bei unserer Beurteilung bleiben, daß Savage-Modelle Beschränkungen aufweisen, die in Fishburn-Modellen beseitigt sind. Insofern verdiente es die Konzeption von Fishburn viel eher, als der Kern der Entscheidungstheorie bezeichnet zu werden (wenngleich auch nicht als mehr, soweit wir sie hier geschildert haben). So wird denn auch das Modell, das wir in Kapitel 3 formulieren werden, im wesentlichen ein Ausbau der Konzeption von Fishburn sein.

Ein kleiner Punkt ist hier aber noch anzumerken: Zwar können in Fishburn-Modellen Umstände Savagescher Provenienz subjektive Werte tragen, was in Savage-Modellen nur bei unschöner Deutung der möglichen Konsequenzen ging. Wie steht es in dieser Hinsicht mit Handlungen, die ja nicht nur erwarteten subjektiven Wert, sondern zuweilen auch, ganz abgesehen von ihren Konsequenzen, eigenständigen subjektiven Wert haben? Das bietet Fishburn-Modellen keine Schwierigkeiten; man kann ja Handlungen als Konsequenzen des jeweiligen Modells aufnehmen. Mich stört daran nur etwas, daß Handlungen auf diese Weise in Fishburn-Modellen doppelt repräsentiert sind: einmal in der Menge H der Handlungen und zum anderen eben als Komponenten der Konsequenzen in C.

Bezüglich der Anwendungsfrage ist Fishburn-Modellen direkt ebenso wenig zu entnehmen wie Savage-Modellen. Allerdings wären hier die Ausführungen von Fishburn (1964), S. 403-408, einschlägig, auch wenn sie dort nicht so gedacht waren. Doch wollen wir das jetzt nicht darstellen, sondern erst in Abschnitt 3.5 wieder ausführlich auf dieses Thema zurückkommen.

Was die Explizitheitsfrage betrifft, so ist hier die Antwort ebenso mager wie beim Grundmodell. Und die Zusammensetzung komplexer Konsequenzen und Handlungen aus elementaren Bestandteilen ließe sich in Fishburn-Modellen ebenso leicht formal wiedergeben wie in Savage-Modellen. Doch taucht hier ein Problem auf, das sich beim Grundmodell nicht stellte:

Im Gegensatz zum Grundmodell enthält das Fishburn-Modell durch Elemente von H bedingte Wahrscheinlichkeiten. Daher stellt sich, will man die Elemente von H als Strategien auffassen, die Frage nach dem logischen Charakter von Strategien mit besonderer Schärfe. Wie Fishburn als bedingte Imperative darf man sie sicherlich nicht verstehen, denn durch Imperative bedingte Wahrscheinlichkeiten sind ähnliche Ungetüme wie Wahrscheinlichkeiten für Konjunktionen von subjunktiven Konditionalen. Der Defekt scheint schnell behoben: Man verstehe unter Strategien nicht bedingte Imperative, sondern einfach die entsprechenden indikativischen Formulierungen. Doch kommt nun erst das eigentliche Problem zum Vorschein: Die Konditionale, die in den Formulierungen der Strategien enthalten sind, darf man nicht als materiale Implikationen verstehen. Denn täte man es, so käme erstens der schon auf S. 30 festgestellte Zusammenhang zwischen Strategien und Beobachtungen nicht zum Ausdruck. Zweitens würden sich dann z.B. die beiden Strategien „wenn A, so tut X f; wenn nicht A, so tut X auch f“ und „wenn A, so tut X f; wenn nicht A , so tut X nicht f“ nicht gegenseitig ausschließen, obwohl man dies eigentlich erwartet hätte. Und drittens hätte die harmlose Annahme, daß ein Ereignis A von allen Strategien, die das Ausführen oder Unterlassen von f davon abhängig machen, ob A oder nicht A, probabilistisch unabhängig ist, eigenartige Folgen: Wie Sneed (1966), S. 230, zeigt, folgte dann nämlich, daß f und nicht-f jeweils 0,5 als Wahrscheinlichkeit haben. Jedoch wird sich auch derjenige, der Wahrscheinlichkeiten für Handlungen überhaupt als sinnvoll ansieht, durch die genannte Annahme nicht auf diese Wahrscheinlichkeitswerte festlegen lassen wollen.

Sneed schlägt vor, die fraglichen Konditionale stattdessen als kausale Implikationen aufzufassen. Einen Schatten auf diesen Vorschlag wirft die Tatsache, daß auch ihm zufolge die verschiedenen Strategien nicht disjunkt sind, obwohl Sneed dies nur wenige Zeilen vor seinem Vorschlag als erwünscht bezeichnet.[21] Außerdem ist die Hereinnahme von etwas philosophisch so Ungeklärtem wie Kausalaussagen in den Bereich der subjektiven Wahrscheinlichkeitsfunktion ein drastischer Schritt, der nur als letzte Notrettung vollzogen werden sollte. Was man sonst hier tun könnte, will ich nun nicht diskutieren. Es sei nur festgehalten, daß die durch Strategien bedingten Wahrscheinlichkeiten der Konzeption von Fishburn ein besonderes Problem präsentieren. In Abschnitt 4.3 (S. 172) werden wir sehen, daß in der Entscheidungstheorie durch Strategien bedingte Wahrscheinlichkeiten gar nicht benötigt werden.

---

[21] Sneed gibt in (1966), S. 281, an, wie er die vier möglichen Strategien, die das Ausführen oder Unterlassen der Handlung B vom Eintreten des Ereignisses X abhängig machen, formal repräsentieren würde. Dabei stellt er aber die ersten beiden möglichen Strategien schon durch disjunkte *und* erschöpfende Propositionen dar, so daß die Propositionen, die die restlichen Strategien repräsentieren, nur dann mit den ersten zwei Strategien unverträglich sein können, wenn sie mit der logisch falschen Proposition identisch sind.

2.5 Wahrscheinlichkeiten für Handlungen

Die bisherigen Ausführungen haben sich an ein bestimmtes entscheidungstheoretisches Prinzip gehalten, von dem die meisten Entscheidungstheoretiker stillschweigend ausgegangen sind – vermutlich weil sie es für selbstverständlich hielten. Da es aber nicht nur für unsere Beurteilung der Konzeption von Jeffrey und der von Luce und Krantz in den nächsten Abschnitten, sondern auch in den Abschnitten 3.2, 3.3 und 5.1 eine zentrale Rolle spielen wird, empfiehlt es sich, es explizit zu machen. Es besagt: *Ein adäquates quantitatives Entscheidungsmodell darf weder explizit noch implizit irgendwelche subjektive Wahrscheinlichkeiten für Handlungen enthalten* (außer vielleicht trivialen bedingten Wahrscheinlichkeiten wie $P(A \mid A) = 1$ für eine Handlung A oder $P(A' \mid A) = 0$ für zwei disjunkte Handlungen A und A'.

Wenn hierbei von Handlungen die Rede ist, so sind damit, das ist klar, diejenigen Dinge gemeint, die im betreffenden Entscheidungsmodell als Handlungen ausgezeichnet sind; und das sind immer die dem Entscheidenden noch offenstehenden Handlungen, also insbesondere nicht Handlungen anderer Personen und auch nicht seine eigenen vergangenen Handlungen; über die kann und wird er natürlich Vermutungen haben.

Wieso sollte man dieses Prinzip akzeptieren? Auf den ersten Blick scheint es nicht unbedingt überzeugend zu sein. Menschen haben und äußern häufig Überzeugungen über ihre eigenen zukünftigen Handlungen. Z.B. werde ich es für sehr unwahrscheinlich halten, daß ich im nächsten Winter meine kurzen Hosen außer Hauses tragen werde, oder ich werde einem Besucher, der mich gerade beim Nachdenken über meinen nächsten Zug in einer Fernschachpartie antrifft, eventuell mitteilen, daß ich vermutlich eine große Rochade machen werden.

Andererseits ist unbestreitbar etwas Eigenartiges an Wahrscheinlichkeiten für Handlungen. Wenn ein Freund mich fragt, ob ich heute abend zur Party komme, und ich „ja“ sage, so ist dieses und ähnliches nicht als Behauptung oder Voraussage, sondern als Ankündigung, Zusage oder womöglich Versprechen zu verstehen. Und wenn der Besucher darauf herumreitet, ob ich wirklich *glaube*, eine große Rochade zu machen, so werde ich schließlich unwirsch entgegnen, die Frage sei nicht, ob ich das glaube, sondern vielmehr, ob ich das wirklich wolle. Das heißt allgemein, daß es vielfach zumindest zweifelhaft ist, daß sich in Äußerungen über die eigenen zukünftigen Handlungen Glaubensdispositionen bezüglich dieser Handlungen manifestieren.

Handfester wird diese Eigenartigkeit durch die folgende Überlegung: Es ist allgemein akzeptiert, daß sich subjektive Wahrscheinlichkeiten für bestimmte Ereignisse in der Bereitschaft zu Wetten auf diese Ereignisse mit geeigneten Wettquotienten manifestieren. Auf Wahrscheinlichkeiten für Handlungen trifft dies jedoch nicht zu. Jemandes Bereitschaft, eine Wette auf eine Handlung von ihm zu akzeptieren,[22] hängt natürlich mitnichten vom Wettquotienten, sondern allein von seinem Gewinn, d.h. vom Einsatz seines Wettpartners ab. Wenn der Gewinn so groß ist, daß diese Handlung zusammen mit dem Gewinn allen Alternativen vorzuziehen ist, so wird er die Abmachung akzeptieren, sonst nicht. Sein eigener Einsatz bei der Wette tut überhaupt nichts zur Sache. Wenn z. B. jemand mir DM 30,- dafür bietet, daß ich einen bestimmten Film anschaue, so würde ich vermutlich akzeptieren, während ich bei DM 3,- noch nicht akzeptieren würde – beides unabhängig davon, wieviel der andere von mir bekommen sollte, falls ich diesen Film nicht anschaue.[23] Wer dennoch an Wahrscheinlichkeiten für Handlungen festhält, hätte also zumindest zu erläutern, wie sie sich sonst manifestieren.

Das bringt mich auf den eigentlichen, eher theoretischen Grund für das obige Prinzip, der, kurz gesagt, darin liegt, daß sich Wahrscheinlichkeiten für Handlungen nicht in diesen Handlungen selbst manifestieren können: Wenn wir eine Person mit einem Entscheidungsmodell beschreiben wollen, so wollen wir darin neben den ihr offenstehenden Handlungen vor allem ihre quantitativen Glaubens- und Wünschensdispositionen repräsentieren, *insoweit* sie für ihre Entscheidung relevant sind, sich darin niederschlagen und sich daraus (bzw. aus ihrer Präferenzordnung für ihre möglichen Handlungen bzw. aus ihren Entscheidungen in leicht variierten Situationen) wieder teilweise oder ganz rekonstruieren lassen. Subjektive Wahrscheinlichkeiten dieser Person für ihre eigenen Handlungen gehören dazu *nicht*. Denn ihre Entscheidung wird alleine davon abhängen, wie lieb ihr die verschiedenen offenstehenden Handlungen sind, und dies wird wiederum davon abhängen, wozu diese Handlungen ihrer Meinung nach möglicherweise führen und

---

[22] Hier von einer Wette zu reden, ist fast schon sprachwidrig; vielleicht sollte man besser von einer wettähnlichen Abmachung sprechen.

[23] In Wirlichkeit wäre natürlich ein solches Angebot so ungewöhnlich, daß ich irgendwelchen Ungedeih wittern und mich deshalb eventuell auch durch größere Beträge nicht in Versuchung führen lassen würde. Doch ändert das nichts am Argument.

wie lieb ihr das ist. Nirgendwo gehen dabei Wahrscheinlichkeiten für diese Handlungen ein; die Person wählt die ihr am besten erscheinende Handlung, was immer ihre Wahrscheinlichkeit wäre. Dementsprechend könnten diese Wahrscheinlichkeiten variieren, und trotzdem bliebe ihre Entscheidung (sowie ihre Präferenzordnung, ihre Entscheidung in einer leicht variierten Situation etc.) dieselbe, solange sich sonst nichts änderte. Daher hat es auch keinen Sinn, ihr mit einem Entscheidungsmodell irgendwelche Wahrscheinlichkeiten für ihre möglichen Handlungen zu unterstellen, d.h. wir sollten das genannte Prinzip akzeptieren.

Wie gesagt, haben wir uns bisher daran gehalten; es läßt sich leicht nachprüfen, daß sowohl Savage- wie Fishburn-Modelle nicht einmal versteckte Wahrscheinlichkeiten für Handlungen enthalten. Demnach ist auch dieses Prinzip mindestens so weitgehend akzeptiert wie die Konzeptionen von Savage und Fishburn. Legen wir also dieses Prinzip allem Folgenden zugrunde. Freilich: Ebenso wie wir denjenigen, der es nicht akzeptiert, einiges zu erklären aufgegeben haben, ist auch seine Annahme noch nicht völlig problemfrei; es bleibt noch zu klären, wie sich dieses Prinzip damit verträgt, daß wir doch häufig unsere eigenen zukünftigen Handlungen probabilistisch beurteilen und auch Grund haben, dies anderen zu unterstellen. Doch will ich dieses Thema erst in den Abschnitten 5.1 und 5.2 wieder aufgreifen, wenn wir die Auswirkungen unseres Prinzips besser überblicken können. Einstweilen sollte nur klar geworden sein, daß Wahrscheinlichkeiten für Handlungen, wenn sie existieren, nicht unter die unmittelbare Zuständigkeit der Entscheidungstheorie fallen.

Eine einzige Folgerung aus diesem Prinzip will ich aber gleich noch festhalten: Wenn in einem Entscheidungsmodell keine Wahrscheinlichkeiten für Handlungen auftauchen dürfen, so dürfen darin auch keine nicht-bedingten, absoluten Wahrscheinlichkeiten für handlungsabhängige Ereignisse enthalten sein. Denn wenn etwa das Ereignis B handlungsabhängig ist, d.h. wenn $P(B \mid A) \neq P(B \mid \overline{A})$ für eine Handlung A und ihr Komplement $\overline{A}$, und wenn außerdem $P(B)$ gegeben wäre, so könnte man $P(A)$ berechnen. Es gilt ja der elementare Sachverhalt:

$$P(A) = \frac{P(B) - P(B \mid \overline{A})}{P(B \mid A) - P(B \mid \overline{A})} \, .$$

Wenn wir also jemanden eine nicht bedingte Wahrscheinlichkeit für ein bestimmtes Ereignis unterstellen, so nehmen wir damit gleichzeitig an, daß dieses Ereignis seiner Meinung nach von seinen Handlungen probabilistisch unabhängig ist.[24]

[24] Genau diese Tatsache verwandten wir – damals noch ohne ausführliche Begründung –, als wir auf S. 46 der formalen Beschaffenheit des Grundmodells entnahmen, daß seine Umstände handlungsunabhöngig seien.

2.5 *Die Konzeption von Jeffrey*

Alle umgangssprachlichen Ausdrücke für Glaubens- und Wünschensdispositionen lassen sich auf die Form „*X* steht in dem einen oder anderen dispositionalen Verhältnis dazu, daß ...“ bringen. Der gesamte Bereich der Glaubens- und Wünschensdispositionen wird also durch propositionale Einstellungen abgedeckt. Dies stellten wir bereits im ersten Kapitel fest. Jeffrey (1965) macht diesen Sachverhalt zum Ausgangspunkt seiner Überlegungen.[25] Insbesondere legt er Wert auf die Tatsache, daß häufig – oder womöglich immer – ein und dieselbe Proposition Gegenstand sowohl probabilistischer als auch wertmäßiger Beurteilung ist (z.B. die Proposition, daß es morgen regnet[26]). Er empfindet daher die strenge Trennung von Glaubens- und Wünschensbereich, wie sie das Grundmodell vornahm, als unannehmbar. Und das Verfahren, ein und dieselbe Proposition als Gegenstand wertmäßiger Beurteilung den Konsequenzen zuzuschlagen und als Gegenstand von Wahrscheinlichkeitserwägungen zum Bestandteil möglicher Umstände zu machen, erscheint widersinnig.

Diesem Übelstand abzuhelfen, ist Jeffreys Hauptziel. Und der erste Schritt dazu ist sehr einfach und einleuchtend. Man nehme schlicht eine Menge von Propositionen als Definitionsbereich sowohl der subjektiven Wahrscheinlichkeits- wie der subjektiven Wertfunktion. Nun sind Propositionen logisch strukturiert, d.h. genauer, daß sie zumindest unter den aussagenlogischen Operationen abgeschlossen sind. Wir wollen also mit Jeffrey annehmen, daß diese Propositionenmenge eine Boolesche Algebra bildet. Dies kommt ja auch der subjektiven Wahrscheinlichkeitsfunktion entgegen, die ja immer auf einer solchen Algebra definiert ist. Jeffrey erläutert freilich Propositionen nur informell, ähnlich wie wir es in Abschnitt 1.1 getan haben; ihre genaue Beschaffenheit ist nicht Bestandteil seiner formalen Theorie. Da ferner jede Boolesche Algebra zu einer Mengenalgebra isomorph ist,[27] könne wir also hier im Endlichen die Propositionen ohne Bedenken durch die Potenzemenge irgendeiner Grundmenge *W* repräsentieren; wir gewinnen dadurch Anschluß an die bisherige Darstellung und machen damit gleichzeitig darauf auf-

---

[25] S. Jeffrey (1965), S. 48f.

[26] Sein Beispiel, S. 48.

[27] Dies lehrt das Repräsentationstheorem von Stone; s. Birkhof (1948), S. 159.

merksam, daß die Rede von Propositionen allein noch keinen erheblichen Unterschied nach sich zieht.

Handlungen werden dabei ebenfalls durch Propositionen erfaßt und sind daher auch Bestandteile von Pot(W). Was in früheren Konzeptionen noch sorgfältig getrennt wurde, ist jetzt alles in einer Menge zusammengemengt. Deshalb wird Jeffreys Konzeption auch häufig als holistisch bezeichnet. Festzuhalten ist dabei, daß die dem Entscheidenden offenstehenden Handlungen nicht beliebige, sondern natürlich disjunkte und in ihrer Gesamtheit erschöpfende Propositionen sind, d.h. eine Zerlegung von W bilden.

Die subjektive Wahrscheinlichkeitsfunktion hat dann selbstverständlich ein Wahrscheinlichkeitsmaß auf Pot(W) zu sein. Und die subjektive Wertfunktion ist, wie oben schon festgehalten, ebenfalls auf Pot(W) definiert – mit einer Ausnahme: Die leere Menge, die der logisch falschen Proposition entspricht, erhält keinen subjektiven Wert. Wozu auch? Das durch sie repräsentierte Ereignis kann ja ohnehin nicht eintreten.[28] Allerdings schaut nun die subjektive Wertfunktion anders aus als von früher her gewohnt:

In den vergangenen Konzeptionen war sie für Konsequenzen definiert, die man sich als gegenseitig ausschließende Ereignisse vorstellte. Jetzt können aber kompliziertere logische Verhältnisse vorliegen. Jeffrey nimmt daher an, daß in der subjektiven Wertfunktion bereits *erwartete subjektive* Werte zum Ausdruck kommen. Genauer: Nach Jeffrey ist der subjektive Wert einer Proposition ein gewichtetes Mittel aus den subjektiven Werten der Fälle, in denen diese Proposition wahr ist, wobei die Gewichte den subjektiven Wahrscheinlichkeiten dieser Fälle proportional sind. Formal ist dies in der folgenden zusammenfassenden Definition ausgedrückt, in der also eine Funktion V für nicht-erwartete, absolute subjektive Werte überflüssig ist:

Definition 5: $\langle W, H, P, U\rangle$ ist ein Jeffrey-Modell genau dann, wenn gilt:
(1) W ist ein leere, endliche Menge,
(2) H ist eine Zerlegung von W,
(3) P ist ein Wahrscheinlichkeitsmaß von Pot(W),

---

[28] Dies ist die Regelung, die Jeffrey in (1965) trifft.. Man kann's auch anders machen; vgl. etwa Jeffrey (1977). Daran allein hängt jedenfalls nichts von Bedeutung.

(4) U ist eine Funktion von Pot(W) \ {∅} in **R**, für die gilt: für $A, B \subseteq W$ mit $P(A \cap B) = 0$[29] und $P(A \cup B) \neq 0$ ist

$$U(A \cup B) \;=\; \frac{U(A)P(A) + U(B)P(B)}{P(A) + P(B)} \;=\; U(A)P(A \mid A \cup B) + U(B)P(B \mid A \cup B).$$

Die mit Jeffrey-Modellen verknüpfte empirische Behauptung[30] besagt dann: Wird X durch das Jeffrey-Modell $\langle W, H, P, U\rangle$ charakterisiert, so führt X ein $A \in H$ aus, für das $U(A)$ maximal ist.

Stellen Jeffrey-Modelle einen Fortschritt dar? Augenfällig ist zunächst die Umbenennung der Gegenstände der subjektiven Wahrscheinlichkeits- und der subjektiven Wertfunktion. Und diese ist auch nützlich; der neutrale Begriff der Proposition ist viel weniger dazu geeignet, falsche Assoziationen zu wecken, als der Begriff des möglichen Umstandes und insbesondere der möglichen Konsequenz. Dennoch ist es zunächst nichts weiter als eine Umbenennung, und deshalb sind wir bezüglich der Explizitheitsfrage nicht klüger geworden. Dies würden wir, glaube ich, wenn man die exakte Konstruktion von Propositionen aus den Sätzen einer Sprache zum Bestandteil der Theorie selbst machte. Und dies wollen wir im Abschnitt 3.1 im Anklang an Carnap (1971) machen.

War der letzte Punkt eher beckmesserisch, so scheint mit der nächste Punkt entscheidend zu sein. Denn Jeffrey-Modelle enthalten ganz offensichtlich Wahrscheinlichkeiten für Handlungen, und das sollten sie unserer Argumentation im letzten Abschnitt zufolge nicht tun.[31]

Jeffrey scheint irrtümlich angenommen zu haben, man entkäme den auf S. 47-52 aufgezeigten Beschränkungen des Grundmodells von Savage schon dadurch, daß man die Trennung von Glaubens- und Wünschensbereich aufhebt. In Wirklichkeit liegen diese Beschränkungen bereits in der Annahme nicht bedingter subjektiver Wahrscheinlichkeiten für Umstände und Ereignisse begründet. Und *diese* Annahme gilt es aufzugeben. Die genannte Trennung fällt dann quasi von selbst in sich zusammen, wie wir an Fishburns Konzeption gesehen haben, die ja fast so holistisch ist wie Jeffreys; in ihr tragen ja auch ein und dieselben Dinge, nämlich

---

[29] Ob dies oder $A \cap B = \emptyset$ angenommen wird, ist gleichgültig. Ich folge hier Jeffrey (1965), S. 70.

[30] Jeffrey selbst hätte freilich keine solche empirische Hypothese formuliert; er faßt seine Theorie vielmehr als eine Theorie rationalen Verhaltens auf.

[31] Vgl. allerdings die Verteidigung von Jeffrey (1977).

Konsequenzen, sowohl handlungsbedingte Wahrscheinlichkeiten wie subjektive Werte.

(Im übrigen: Falls man partout Wahrscheinlichkeiten für Handlungen irgendwie Sinn beimessen will, so müßte man wohl verlangen, daß die subjektive Wahrscheinlichkeit einer Handlung umso großer ist, je höher ihr (erwarteter) subjektiver Wert. Bei Jeffrey ist dies jedoch nicht der Fall. Bei ihm ist die Wahrscheinlichkeit einer Handlung desto größer, je kleiner der Abstand ihres subjektiven Wertes vom subjektiven Wert der logisch wahren Proposition W ist. Um ein Beispiel zu geben: Sei A eine Handlungsproposition aus Pot(W) und $\overline{A}$ die Komplement-Handlung von A. Sei $U(W) = 0$, $U(A) = 20$ und $U(\overline{A}) = -10$. Mit Definition 5, (4), erhalten wir dann $P(A) = 1/3$ und $P(\overline{A}) = 2/3$. Die schlechtere Handlung hat hier also die höhere Wahrscheinlichkeit. Und dies scheint unannehmbar.)

Läßt sich dieser Defekt reparieren? Gewiß. Allerdings nicht dadurch, daß wir die Handlungen aus Pot(W) herausnehmen und die Ereignisse in Pot(W) als handlungsunabhängig betrachten; denn dann hätten wir nur Handlungen und Umstände im Sinne Savages, aber nichts, was uns erwartete subjektive Werte für Handlungen liefern könnte. Vielmehr müssen wir die Wahrscheinlichkeiten für Handlungen und die nicht-bedingten Wahrscheinlichkeiten für handlungsunabhängige Ereignisse aus Jeffrey-Modellen entfernen. Ohne dies jetzt genau zu schildern – was wir im Abschnitt 5.2 tun, läßt sich direkt hier verwenden –, dürfte klar sein, daß dann gerade Fishburns Repräsentation subjektiver Glaubensgrade resultiert.

Betrachten wir schließlich die Darstellung subjektiver Wünschensgrade im Jeffrey-Modell. Da ist zunächst zu bemerken, daß die übliche subjektive Wertfunktion V in der Funktion U des Jeffrey-Modells implizit enthalten ist. Denn wenn wir V durch $V(w) = U(\{w\})$ für alle $w \in W$ definieren, so gilt gemäß Definition 5, (4), für alle $A \subseteq W$ $U(A) = \sum_{w \in A} V(w) \cdot P(w \mid A)$, d.h. U gibt gerade Erwartungswerte von V bzgl. P an.

So wäre es denn egal, ob man die Wünschensdispositionen des Handelnden wie Jeffrey durch eine Funktion U von Pot(W) \ ($\varnothing$) in $\mathbf{R}$, für die Definition 5, (4), gilt, oder durch eine Funktion V von W in $\mathbf{R}$ repräsentiert? Ich glaube nicht. Denn der eben festgestellte Defekt auf der Wahrscheinlichkeitsseite berührt auch die Funktion U, nicht aber die Funktion V. Ersetzt man nämlich das auf ganz Pot(W) definierte Wahrscheinlichkeitsmaß P durch etwas Komplizierteres, wie wir es vorhin verlangt haben, dann lassen sich erwartete subjektive Werte nicht mehr für alle

Propositionen berechnen. Das übliche Vorgehen scheint mir daher angemessener zu sein.

Dieses abstrakte Argument hat auch seine intuitiven Gegenstücke. Schon die Frage „Wie lieb ist es dir, daß entweder kaltes, trockenes oder warmes, aber regnerisches Wetter herrscht?" ist etwas merkwürdig, doch erscheint hier eine Antwort im Sinne von Definition 5, (4), nicht völlig abwegig. Sind die Disjunktionsglieder dagegen handlungsabhängig, so errechnet sich der subjektive Wert der Disjunktion, sofern er überhaupt Sinn ergibt, nicht nach Definition 5, (4). Habe ich z.B. zum Nachtisch lieber Eis als Obst, so ist es mir egal, ob auf der Menukarte unter „Dessert" nur „Eis" oder „Eis oder Obst" steht. Nach Jeffrey müßte aber der subjektive Wert von „Eis oder Obst" ein gewichtetes arithmetisches Mittel aus dem subjektiven Wert von Eis und dem von Obst, mithin im allgemeinen kleiner als der subjektive Wert von Eis sein. Pflegte das Restaurant die eigenartige Praxis, durch Würfelwurf zu bestimmen, welcher Gast welchen Nachtisch erhält, so wäre Definition 5, (4), auch in diesem Falle angemessen. Aber diese Praxis hat sich bisher nicht durchgesetzt – zum Glück.

Allgemein läßt sich sagen: Sind $A_1, \ldots, A_m$ disjunkte Handlungspropositionen und will man neben den $U(A_i)$ partout auch noch $U(\bigcup_{i=1}^{m} A_i)$ erklärt haben, so wäre $U(\bigcup_{i=1}^{m} A_i)$ durch $\max_{i \leq m} U(A_i)$ adäquater ausgedrückt als durch ein gewichtetes Mittel aus den $U(A_i)$ mit Gewichten, die man ohnehin nicht hat.

Zuletzt ist darauf hinzuweisen und als offenes Problem festzuhalten, daß sich dem Einbau von Strategien in Jeffreys Konzeption die gleichen Schwierigkeiten entgegenstellen, die sich auch schon bei Fishburns Konzeption ergaben. Denn aus denselben Gründen wie dort können wir hier Strategien nicht als Konjunktionen verschiedener materialer Implikationen auffassen. In der Tat ist ja der Aufsatz von Sneed (1966), dem ich diese Gründe entnahm, ursprünglich eine Kritik an Jeffrey.

Ich möchte aber doch noch betonen, daß diese ganze Kritik nur die hier dargestellten Jeffrey-Modelle und daher Jeffrey (1965) insgesamt nur zu einem geringen Teil trifft.

2.7 *Die Konzeption von Luce und Krantz*

Zuletzt wollen wir die von Luce und Krantz (1971) vorgeschlagene Variante der Entscheidungstheorie unter die Lupe nehmen. Da sie fast ausschließlich mit der – erfolgreichen – Lösung des Metrisierungsproblems für ihre Konzeption beschäftigt sind, erfahren wir hier bzgl. der Explizitheits- und der Anwendungsfrage nichts Neues, so daß wir uns wiederum auf die Allgemeinheitsfrage konzentrieren können.

Luce und Krantz beabsichtigen zwar ausdrücklich, die Anregungen Fishburns aufzunehmen,[32] indem sie, Fishburn (1964) und Pfanzagl (1967) folgend, eine, wie sie es nennen, bedingte Version der Entscheidungstheorie formulieren. Doch ähnelt ihre Version auf den ersten Blick (und, wie wir feststellen werden, nicht nur auf den ersten Blick) viel mehr der Savages. Dieser gegenüber enthält sie aber zwei Neuerungen. Die erste besteht in einer neuen formalen Repräsentation von Handlungen, und die zweite bringt eine allgemeinere Behandlung subjektiver Werte mit sich. Letztlich hängen natürlich beide Neuerungen miteinander zusammen, doch empfiehlt es sich für unsere Darstellung und Diskussion, auf die zweite einzugehen, nachdem wir die erste abgehandelt haben.

Der Ausgangspunkt von Luce und Krantz ist dem Fishburns ähnlich. Auch sie stoßen sich daran, daß die Umstände in Savage-Modellen handlungsunabhängig sind. Und die Ursache erblicken sie in der Tatsache, daß Handlungen in Savage-Modellen durch Funktionen repräsentiert werden, die für *alle* Umstände definiert sind. Dies berücksichtigt nicht, daß Handlungen zumindest teilweise bestimmen, welche Ereignisse eintreten. Wähle ich etwa die Eisenbahn als Fortbewegungsmittel, so kann ich schlechterdings keinen Autounfall haben. Die Rede von Umständen ist dann im Grunde unangemessen, da dies schon Handlungsunabhängigkeit mitbedeutet. Ich will daher in diesem Abschnitt stattdessen von Zuständen reden (was ja eigentlich die bessere, wenn auch nicht die eingebürgerte Übersetzung von Savages Terminus „world states" ist).

Luce und Krantz ändern demgemäß Savage-Modelle in der folgenden Weise ab: Zunächst nehmen sie wie Savage eine Menge W möglicher Zustände, eine Menge C möglicher Konsequenzen und ein Wahrscheinlichkeitsmaß P auf Pot(W)

[32] S. Luce, Krantz (1971), S. 253.

für die subjektiven Wahrscheinlichkeiten des Handelnden an. Ferner will ich unter vorläufiger Vernachlässigung ihrer zweiten Neuerung eine auf C definierte subjektive Wertfunktion V in ihr Modell aufnehmen.

Die entscheidende Abweichung gegenüber Savage liegt in der Darstellung der Handlungen. Um den Einfluß der Handlungen auf das tatsächliche Geschehen gerecht zu werden, nehmen Luce und Krantz an, daß jede Handlung gewissermaßen ihren eigenen Möglichkeitsraum schafft. Wird eine gewisse Handlung ausgeführt, so können sich nicht mehr alle, sondern nur noch ganz bestimmte Möglichkeiten aus dem Möglichkeitsraum W realisieren. Diese bilden zusammen ein Ereignis $A \subseteq W$, und daher läßt sich diese Handlung als eine Funktion f_A von der Menge A der verbliebenen Möglichkeiten in die Menge C der Konsequenzen repräsentieren. Da die außerhalb A liegenden Möglichkeiten durch die Handlung selbst schon ausgeschlossen sind, wäre es sinnlos, wenn die diese Handlung repräsentierende Funktion auch für sie definiert wäre. Vernünftigerweise ist noch zu verlangen, daß das Ereignis A nicht selbst von Wahrscheinlichkeit 0 ist.

Luce und Krantz nennen solche partielle Funktionen *bedingte Entscheidungen*. Es ist dabei nützlich, den Definitionsbereich einer bedingten Entscheidung als Index am Funktionssymbol festzuhalten. f_A steht also für eine Funktion von $A \subseteq W$ in C, g_B für eine Funktion von $B \subseteq W$ in C etc.

Die andersartige Repräsentation der möglichen Handlungen kommt auch in der Berechnung ihres erwarteten subjektiven Wertes zum Ausdruck: Da mit der Durchführung der bedingten Entscheidung f_A der Möglichkeitsraum auf A eingeschränkt ist, sind dann nur durch A bedingte Wahrscheinlichkeiten von Bedeutung. Der erwartete subjektive Wert einer bedingten Entscheidung errechnet sich daher nach der in der folgenden zusammenfassenden Definition angegebenen Formel.

Definition 6: $\langle W, C, H, P, V, U\rangle$ ist ein *spezielles Luce-Krantz-Modell*[33] genau dann, wenn gilt:

(1) W und C sind nicht leere, endliche Mengen,

(2) H ist eine nicht leere Menge von Funktionen in C, deren Definitionsbereiche Teilmengen von W sind, wobei für jedes $f_A \in H$ $P(A) \neq 0$,

(3) P ist ein Wahrscheinlichkeitsmaß auf Pot(W),

(4) V ist eine Funktion von C in $\mathbf{R}$,

[33] Wir haben ja bisher erst eine der beiden von Luce und Krantz eingeführten Neuerungen geschildert; daher das Attribut „speziell“.

(5) U ist die Funktion von H in R, für die für jedes $f_A \in H$

$$U(f_A) = \sum_{w \in A} V(f_A(w)) \cdot P(w \mid A).$$

Doch ist die Repräsentation von Handlungen durch partielle, also nicht unbedingt auf ganz W definierte Funktionen wirklich überzeugend? Ich glaube nicht. Die dafür gegebene Erläuterung sagte ja, daß eine Handlung dann durch eine Funktion f_A zu repräsentieren sei, wenn sie gerade mit den Zuständen in A und keinen anderen verträglich ist. Doch mit welchen möglichen Zuständen ist eine Handlung verträglich? Offensichtlich mit genau den Zuständen, in denen diese Handlung ausgeführt wird. Daher ist das von diesen Zuständen gebildete Ereignis gerade das Ereignis, das in der Ausführung dieser Handlung besteht. Kurz gesagt, drückt A also aus, daß f_A ausgeführt wird. A erhält jedoch in speziellen Luce-Krantz-Modellen einen Wahrscheinlichkeitswert. Damit stehen wir vor der unangenehmen Tatsache, daß spezielle Luce-Krantz-Modelle zumindest versteckt Wahrscheinlichkeiten für Handlungen enthalten.

Die Voraussetzungen dieses Argumentes lassen sich sogar noch abschwächen. Es hängt nicht wirklich davon ab, daß das Ereignis A gerade in der Ausführung der Handlung f_A besteht. Vielmehr reicht bereits die Tatsache, daß Pot(W) handlungsabhängige Ereignisse enthält; den Handlungen wird ja ein Einfluß auf die möglichen Zustände eingeräumt. Auch dann sollte, wie am Ende von Abschnitt 2.5 festgestellt, ein Entscheidungsmodell für diese Ereignisse keine nicht-bedingten Wahrscheinlichkeiten enthalten.

Luce und Krantz beseitigen also mit ihren bedingten Entscheidungen ebenso wenig wie Jeffrey mit seinen vereinheitlichten Definitionsbereichen die Beschränkungen des Grundmodells auf der Glaubensseite. Diese liegen eben, wie in Abschnitt 2.2 bereits festgestellt, in der Repräsentation quantitativer Glaubensgrade durch ein nicht-bedingtes Wahrscheinlichkeitsmaß, und daran haben weder Jeffrey noch Luce und Krantz gerüttelt. Für Luce-Krantz-Modelle hat dies allerdings nicht so schlimme Folgen wie für Jeffrey-Modelle, die ohne Änderung nicht haltbar waren. Denn die Luce-Krantz-Modelle lassen sich immer noch sinnvoll interpretieren:

Um unserem Prinzip aus Abschnitt 2.5 zu genügen, müssen wir dazu zunächst Zustände und die von ihnen gebildeten Ereignisse als handlungsunabhängig auffassen. Es hat dann auch keinen Sinn mehr, Handlungen durch partielle Funktionen zu repräsentieren. Denn Handlungen haben nun keinen Einfluß mehr auf die möglichen Zustände und müssen daher unter *allen* möglichen Zuständen bedacht, d.h. als auf ganz W definierte Funktionen repräsentiert werden. Wir scheinen damit freilich geradewegs auf Savages Position zurückgeworfen zu werden.

Tatsächlich ist die Situation aber noch unangenehmer. Denn wir müssen nach wie vor den bedingten Entscheidungen einen Sinn abgewinnen. Um nämlich mit ihrer Metrisierung zu Rande zu kommen, benötigen Luce und Krantz einen ganzen Haufen bedingter Entscheidungen, genauer gesagt, nehmen sie die folgende Struktur für H an: Bezeichne H_W die Menge aller auf ganz W definierten $f_W \in H$. Dann soll H genau aus allen Beschränkungen der Funktionen aus H_W auf Ereignisse $A \subseteq W$ mit $P(A) \neq 0$ bestehen.[34] Lassen sich aber bedingte Entscheidungen überhaupt noch sinnvoll interpretieren?

Ja, wenngleich nicht als echte Handlungen. Denn nach wie vor ist unangefochten, daß eine bedingte Entscheidung f_A ($A \neq W$) eine möglich Handlung auf einem eingeschränkten Möglichkeitsraum darstellt. Aufzugeben war nur die Vorstellung, daß diese Einschränkung von der Handlung selbst herrührt. Wodurch könnte diese Einschränkung aber sonst bedingt sein? Ganz einfach – durch die Erfahrung des Handelnden, daß das Ereignis A tatsächlich vorliegt. Nun hat der Handelnde aber diese Erfahrung nicht – jedenfalls noch nicht – gemacht. Die ganze Geschichte ist also folgendermaßen zu verstehen: Sei $f_A \in H$. Dann gibt es ein $f_W \in H$ mit $f_A \subseteq f_W$.[35]; dies ergab sich ja aus Luces und Krantz' Forderungen an H. f_W repräsentiert dann genau dieselbe Handlung unter der hypothetischen Annahme, daß der Entscheidende erfährt, daß A vorliegt. Luce und Krantz verlangen also vom Entscheidenden, daß er nicht nur die ihm offenstehenden Handlungen nach seinen Präferenzen ordnet, sondern in seine Präferenzordnung auch noch einbezieht, wie lieb ihm diese Handlungen wären, wenn er bestimmte Erfahrungen machte; und dabei muß er jede mögliche Erfahrung bedenken. Wie Luce und Krantz selbst feststellen,[36] lassen sich diese Präferenzen durch hypothetische Fragen ermitteln: „Wäre dir diese Handlung mit dem Wissen, daß A, oder jene Handlung mit dem Wissen, daß B, lieber?"

---

[34] Tatsächlich sind die strukturellen Annahmen über H, die Krantz et. al. (1971), S. 380, Axiome 1 und 9, machen, noch stärker. Für uns sind sie aber nur insoweit interessant, als sie die genannte Struktur implizieren.

[35] Funktionen sind dabei natürlich als Mengen von Paaren aufzufassen.

[36] Luce, Krantz (1971), S. 258.

Luce und Krantz sind sich sehr wohl im klaren darüber, daß sich in H eine ganze Menge solcher hypothetischer Handlungen befinden.[37] Aber sie nahmen zumindest von einigen partiellen Funktionen an, daß sie tatsächlich offenstehende Handlungen repräsentieren. Dies scheint mir nicht richtig zu sein; aufgrund von Abschnitt 2.5 können nur auf ganz W definierte Handlungen für echte Handlungen stehen, wie es in Savage-Modellen der Fall ist.

In der Tat sind wir bisher nur zu einer sinnvollen Interpretation spezieller Luce-Krantz-Modelle, aber noch zu keinerlei Fortschritt gegenüber Savage gelangt. Selbst die erwarteten subjektiven Werte bedingter Entscheidungen gemäß Definition 6, (5), bringen nichts Neues, da nach der Erfahrung, daß A, der erwartete subjektive Wert einer Handlung f_A auf der Grundlage von Savage-Modellen ebenfalls nach dieser Formel zu berechnen wäre.

Doch ist jetzt der Zeitpunkt gekommen, die zweite Neuerung von Luce und Krantz ins Feld zu führen. In spezielle Luce-Krantz-Modelle nahmen wir eine subjektive Wertfunktion auf, wie sie auch schon in Savage-Modellen vorkommt. Tatsächlich machen Luce und Krantz etwas anderes; ähnlich wie Jeffrey operieren sie direkt mit einer Funktion U, die erwartete subjektive Werte für bedingte Entscheidungen liefert und die daher eine zu Definition 5, (4), (S. 78) analoge Bedingung zu erfüllen hat:

Definition 7: $\langle W, C, H, P, U\rangle$ ist ein *Luce-Krantz-Modell* genau dann, wenn gilt[38]:

(1) W und C sind nicht leere, endliche Mengen,

(2) H ist eine nicht leere Menge von Funktionen in C, deren Definitionsbereiche Teilmengen von W sind, wobei für jedes $f_A \in H$ $P(A) \neq 0$,

(3) für jedes $f_A \in H$ gibt es ein $f_W \in H$ mit $f_A \subseteq f_W$, und mit $f_W \in H$ ist auch für jedes $A \subseteq W$ mit $P(A) \neq 0$ die Einschränkung von f_W auf A in H,

(4) P ist ein Wahrscheinlichkeitsmaß auf Pot(W),

(5) U ist eine Funktion von H in $\mathbf{R}$, für die gilt: wenn $f_A \cup g_B \in H$, wobei $A \cap B = \emptyset$ und $P(A) \neq 0 \neq P(B)$, dann ist

$$U(f_A \cup g_B) = U(f_A) \cdot P(A \mid A \cup B) + U(g_B) \cdot P(B \mid A \cup B).$$

---

[37] Vgl. Luce, Krantz (1971), S. 257f., und (1974).

[38] Es ist in den folgenden Bedingungen wesentlich, Funktionen als Mengen von Paaren aufzufassen.

Der erste Teil von Bedingung (3) spiegelt dabei lediglich unsere Interpretation bedingter Entscheidungen wider, während der zweite Teil, Luces und Krantz' Forderung, daß die offenstehenden Handlungen auch unter allen möglichen Erfahrungen zu betrachten seien[39], für das nachstehende Theorem nötig ist. Die mit Luce-Krantz-Modellen verknüpfte empirische Behauptung besagt: Wird X durch das Luce-Krantz-Modell $\langle W, C, H, P, U\rangle$ charakterisiert, so führt X eine Handlung $f_W \in H_W$ aus, für die gilt: für alle $g_W \in H_W$ ist $U(f_W) \geq U(g_W)$.

Nun interessiert natürlich die Frage, wessen Erwartungswerte die Funktion U von Luce-Krantz-Modellen ausdrückt und ob damit eine allgemeinere Behandlung subjektiver Werte geleistet wird.[40] Vollständig beantwortet wird diese Frage durch den

Satz 7: Sei $\langle W, C, H, P, U\rangle$ ein Luce-Krantz-Modell. Sei V eine auf $W \times C$ definierte Funktion, für die gilt: wenn es für $w \in W$ mit $P(w) \neq 0$ und $c \in C$ ein $f_{\{w\}} \in H$ gibt mit $f_{\{w\}}(w) = c$, dann ist $V(w, c) = U(f_{\{w\}})$. Genau dann gilt für alle $f_A \in H$:

$$U(f_A) = \sum_{w \in A} V(w, f_A(w)) \cdot P(w \mid A)\ .$$ [41]

Satz 7 sagt also gerade, daß in Luce-Krantz-Modellen implizit eine weitgehend eindeutig bestimmte, für Kombinationen aus Zuständen *und* Konsequenzen definierte subjektive Wertfunktion enthalten ist, bezüglich der U die Erwartungswerte bedingter Entscheidungen liefert. „Weitgehend eindeutig bestimmt" heißt dabei: für alle für realisierbar gehaltenen Kombinationen $\langle w, c\rangle$ eindeutig bestimmt. Die Festlegung von V in Satz 7 war ja nur bei solchen $\langle w, c\rangle$ beliebig, in denen entweder w selbst schon unmöglich, also $P(w) = 0$, war oder für die es keine Handlung gab, die unter w zu c führte.

Luce-Krantz-Modelle sind also allgemeiner als Savage-Modelle; sie heben die auf S. 47 notierte restriktive Annahme des Grundmodells auf, daß Zustände und Konsequenzen wertmäßig unabhängig sind. Freilich tritt diese Verallgemeinerung an einer zunächst unerwarteten Stelle auf: Luce und Krantz hofften, durch die Einführung bedingter Entscheidungen die Beschränkungen des Grundmodells auf der

---

[39] Vgl. S. 84.

[40] Diese Frage habe ich ausführlich und ohne die hier gegebene Beschränkung auf den endlichen Fall in Spohn (1977), Abschnitt 7, erörtert. Ich schildere hier die Quintessenz davon.

[41] Dies ist eine ziemlich direkte Folgerung aus Definition 7, (3) und (5).

*Glaubens*seite zu überwinden. Die bedingten Entscheidungen taten ihnen, wie wir sahen, diesen Gefallen nicht. Und nun stellte sich heraus, daß Luce-Krantz-Modelle gerade die Beschränkungen des Grundmodells auf der *Wünschens*seite beseitigen. Leider ist die Restriktivität bei der Repräsentation quantitativer Glaubensgrade auch in Luce-Krantz-Modellen ungebrochen.

2.8 Schlußbemerkung

Unsere Diskussion hat also ergeben, daß Fishburns Ansatz, jedenfalls hinsichtlich der zentralen Allgemeinheitsfrage, zu den befriedigendsten quantitativen Entscheidungsmodellen führt; und ich habe den Eindruck gewonnen, daß diese Einschätzung nicht sehr strittig ist. Dennoch spielen Fishburn-Modelle eine relativ geringe Rolle. Das dürfte seinen Grund darin haben, daß dort das von uns außer acht gelassene Metrisierungsproblem einen breiten Raum einnimmt und Fishburn-Modelle im Hinblick darauf eindeutig am schlechtesten abschneiden:

Savage gelang es in (1954) – und nicht nur in dieser Hinsicht war sein Werk bahnbrechend –, aus einer einzigen Präferenzrelation auf der Menge der Handlungen die quantitativen subjektiven Wahrscheinlichkeiten und subjektiven Werte herauszuholen. D.h. er bewies für hinreichend reichhaltiges W, C und H, daß es zu jeder transitiven und konnexen Relation $\preceq^w$ auf H, die natürlich noch einige weitere Eigenschaften haben muß, Funktionen P und V eines Savage-Modells existieren derart, daß für alle $f, g \in H$ $f \preceq^w g$ genau dann, wenn $U(f) \leq U(g)$, und obendrein bewies er, daß dadurch P eindeutig und V bis auf positiv lineare Transformationen eindeutig bestimmt ist. (Die durchsichtigste Darstellung dieses Beweises fand ich in Fishburn (1977), Kap. 14.)

Und ebenso bewiesen Luce und Krantz (in Krantz et. al. (1971), Abschnitte 8.2 und 8.3), daß es zu jeder noch weitere Bedingungen erfüllenden transitiven und konnexen Relation $\preceq^w$ auf einer genügend reichhaltigen Menge H bedingter Entscheidungen Funktionen P und U eines Luce-Krantz-Modells gibt derart, daß für alle $f_A, g_B \in H$ $f_A \preceq^w g_B$ genau dann, wenn $U(f_A) \leq U(g_B)$, wobei die Eindeutigkeitsaussagen dieselben sind wie bei Savege. Schließlich wurde auch im Rahmen der Konzeption Jeffreys (von Bolker (1966)) Ähnliches bewiesen, nämlich daß es zu jeder Präferenzrelation $\preceq^w$ zwischen Propositionen Funktionen P und U eines Jeffrey-Modells gibt derart, daß für je zwei Propositionen A und B $A \preceq^w B$ genau dann, wenn $U(A) \leq U(B)$, wobei diesmal die Eindeutigkeitsbehauptungen etwas komplizierter aussehen (vgl. dazu Jeffrey (1965), Kap. 8, oder auch Stegmüller (1973a), S. 353ff.). All diese Theoreme muten sehr erstaunlich an, und sie erfordern auch einen gehörigen mathematischen Aufwand.

Für Fishburn-Modelle hingegen existiert ein vergleichbares Theorem nicht. Der Grund dafür ist leicht einzusehen. Denn bei Fishburn-Modellen wäre neben der subjektiven Wertfunktion nicht nur ein Wahrscheinlichkeitsmaß, sondern gleich eine ganze Familie von Wahrscheinlichkeitsmaßen, zwischen denen keinerlei Zusammenhang zu bestehen braucht, aus einer Präferenzrelation zu konstruieren. Und das ist natürlich eine noch wesentlich schwierigere und vermutlich überhaupt nicht zu lösende Aufgabe.[42]

Doch von unserem in Abschnitt 2.1 eingenommenen und, wie versprochen, in Abschnitt 5.3 noch etwas ausführlicher begründeten Standpunkt aus braucht uns dies nicht zu beeindrucken. Im Bewußtsein, daß Fishburn-Modelle in Bezug auf die Allgemeinheitsfrage im wesentlichen optimal sind, wollen wir daher, von der Konzeption Fishburns ausgehend, im nächsten Kapitel die in diesem Kapitel offen gebliebenen Probleme und im vierten Kapitel dynamische Fragestellungen angehen.

[42] Neuerdings existiert sogar für Fishburn-Modelle ein Metrisierungsresultat von Balch, Fishburn (1974), das jedoch „extraneous probabilities“, d.h. Lotterien zur Hilfe nimmt und so von wesentlich stärkeren Voraussetzungen ausgeht; vgl. dazu Abschnitt 5.3.

KAPITEL 3

Details zur Konzeption Fishburns

Unser Programm für dieses Kapitel ist, genauer gesagt, dies: In Abschnitt 3.1 wollen wir, um in der Explizitheitsfrage zu einer gewissen formalen Klärung zu kommen, die genaue Konstruktion von Propositionen auf sprachlicher Grundlage angeben, dabei gleichzeitig zeitliche Relationen formal berücksichtigen und dann auf dieser Basis das Fishburn-Modell neu formulieren. Danach wenden wir uns der Anwendungsfrage zu. Dazu ist es nützlich, sowohl den Glaubensbereich wie den Wünschensbereich betreffende Unabhängigkeitsrelationen erst einmal ganz allgemein zu untersuchen. So befaßt sich der Abschnitt 3.2 mit probabilistischer Unabhängigkeit, was – aus Gründen, die dort klar werden – eine Untersuchung kausaler Unabhängigkeit nach sich zieht (Abschnitt 3.3). Abschnitt 3.4 ist wertmäßiger Unabhängigkeit gewidmet. Mit Hilfe dieser Unabhängigkeitsrelationen versuchen wir dann im Abschnitt 3.5 eine Antwort auf die Anwendungsfrage. Abschnitt 3.6 schließlich knüpft an Abschnitt 2.3 an; wir gehen hier also noch einmal der dort tatsächlich behandelten Frage der Verkleinerbarkeit oder Reduktion von Entscheidungsproblemen nach.

3.1 Spielräume, Propositionen und ein erstes Entscheidungsmodell

Die Grundentitäten einer Entscheidungssituation – was früher als Umstand, Konsequenz, Handlung, Ereignis etc. in die verschiedenen Modelle einging – sollen nun also durch Propositionen repräsentiert werden. Damit da nicht einfach dieselben vagen Dinge mit einem neuen Namen belegt werden, wollen wir die Verbindung von Propositionen zu ihren sprachlichen Grundlagen genau beschreiben. Dabei können wir uns im wesentlichen auf Carnap (1971) oder auf dessen Darstellung in Stegmüller (1973a) stützen. Einige kleine Abweichungen sind allerdings nötig; welche das sind, wird im Text klar werden. Sie resultieren daraus, daß sich das Interesse in der Entscheidungstheorie auf andere Dinge konzentriert als in der induktiven Logik.

Der erste Schritt zur Konstruktion von Propositionen besteht in der Annahme eines Individuenbereiches *Ind*.[1] Im Gegensatz zu Carnap wollen wir aber, da wir hier nur den endlichen Fall behandeln, annehmen, daß wir es nur mit endlich vielen Individuen zu tun haben. Außerdem ist zu verlangen, daß sich der Handlende unter diesen Individuen befindet. Eine weitere Abweichung ergibt sich daraus, daß wir von vornherein zeitliche Relationen berücksichtigen wollen. Daher beziehen wir noch eine weitere Menge *T* von Individuen – Zeitindizes –, die ebenfalls endlich sein soll, in unsere Darstellung ein. Es ist dann natürlich zu verlangen, daß *Ind* selbst keine Zeitpunkte, -intervalle oder dergleichen enthält. Carnaps einheitliche Menge von Individuen zersplittert hier also in die spezielle Menge *T* der Zeitindizes und die Menge *Ind* sonstiger Individuen.

Ein Wort dazu, wie diese Zeitindizes zu verstehen sind: Man kann die Elemente von *T* einmal so auffassen, daß sie für Zeitintervalle stehen, die, im allgemeinen wenigstens, eine Zerlegung eines Ausschnitts der reellen Zeitachse bilden. (Ließe man zu, daß *T* überabzählbar ist, so könnte man sogar die einzelnen Zeitpunkte als Elemente von *T* ansehen). Man braucht allerdings unter diesen Zeitindizes nicht unbedingt konkrete Zeitintervalle oder -punkte zu verstehen. Denn in vielen Entscheidungsproblemen kommt es nicht auf die einzelne Minute oder Sekunde an, sondern nur darauf, daß die Dinge in der richtigen zeitlichen Ordnung geschehen.

---

[1] Vgl. Stegmüller (1973a), S. 418.

In solchen Fällen brauchen die Elemente von T nur diese zeitliche Ordnung widerzuspiegeln. In der Tat benötigen wir noch eine Ordnungsrelation $\leq$ für T, aber auch nur sie; selbst wenn die Zeitindizes für Zeitintervalle stehen, wollen wir auf die metrischen Eigenschaften der Zeitintervalle nicht eingehen.

Als nächstes nehmen wir eine endliche Menge $\mathcal{F}$ von Attributfamilien an.[2] Jede Familie $F \in \mathcal{F}$ besteht dabei aus einer endlichen Menge sich gegenseitig ausschließender Attribute, die in ihrer Gesamtheit das abdecken, was Carnap eine Modalität nennt. Modalitäten sind z.B. Farbe, Größe, Alter, aber auch Art der Tätigkeit, Kinoprogramm, Wanderroute, Ertrag pro Hektar, Wetterzustand, etc. Attributfamilien zerlegen Modalitäten mehr oder weniger detailliert. Das Alter von Menschen kann ich z.B. in Jahrzehnten, Jahren, Tagen oder noch genauer angeben. Den Wetterzustand kann ich einfach als schön, mäßig oder schlecht klassifizieren, ich kann ihn aber auch in mehrere Modalitäten auflösen: Temperatur, Luftdruck, Luftfeuchtigkeit, Grad der Bewölkung, etc.

Zu betonen ist, daß die Attribute einer Familie auch mehrstellig sein können. In unserem Kontext müssen sie es sogar sein; denn eine Stelle eines jeden Attributes ist von vornherein für einen Zeitindex reserviert. Natürlich müssen dabei alle Attribute einer Familie gleiche Stellenzahl haben. Die Stellenzahl der Attribute der Familie $F \in \mathcal{F}$ sei $n_F + 1$. (n_F Stellen sind für Individuen aus *Ind*, und die letzte Stelle ist für Zeitindizes.)

Zu diesem begrifflichen System können wir nun das sprachliche Gegenstück aufbauen.[3] Wir können Individuenkonstanten $a_1, a_2, \ldots$ für die Individuen aus *Ind* und $t_1, t_2, \ldots$ für die Zeitindizes aus T einführen. Wir können Prädikatkonstanten geeigneter Stellenzahl für die Attribute der verschiedenen Familien einführen, außerdem eine zweistellige Prädikatkonstante, um die zeitliche Ordnungsrelation auszudrücken. Schließlich können wie auch noch Individuenvariablen, Junktoren und Quantoren in die Sprache einbauen und so logisch zusammengesetzte Ausdrücke erklären. Auf diese Weise landen wir im wesentlichen bei einer zweisortigen Prädikatenlogik erster Stufe, bei der wir sogar auf die Quantifikation verzichten könnten, da der Gegenstandsbereich $Ind \cup T$ endlich ist. Hätten wir es mit abzählbar unendlich vielen Gegenständen zu tun, so wäre die Quantifikation wesentlich. Und wäre $Ind \cup T$ oder auch einige der Attributfamilien überabzählbar, so würde dies die sprachlichen Mittel der Prädikatenlogik erster Stufe sprengen. Das sprachliche

---

[2] Vgl. Stegmüller (1973a), S. 419.

[3] Vgl. Stegmüller (1973a), S. 446f.

Gegenstück zum Begriffssystem müßte sich dann auf logisch stärkere Systeme stützen.

Propositionen könnte man dann als Mengen logisch äquivalenter Sätze dieser Sprache auffassen. Ein allerdings erst im unendlichen Fall allgemeinerer Propositionsbegriff ergibt sich freilich, wenn man unter Propositionen bestimmte Mengen semantischer Bewertungen dieser Sprache versteht. Da sich die Attribute einer Familie aus $\mathcal{F}$ gegenseitig ausschließen, kann man dabei eine Bewertung dieser Sprache als eine Funktion ansehen, die jeder Familie $F \in \mathcal{F}$, jedem n_F-Tupel a von Individuen aus *Ind* und jedem Zeitindex $t \in T$ ein Attribut $p \in F$ zuordnet. Eine solche Funktion regelt die Wahrheitswerte aller Atomsätze und somit aller Sätze dieser Sprache. Die Bewertungsfunktionen, nach denen ein bestimmter Atomsatz wahr ist, bilden dann zusammen eine Elementarproposition. Und Propositionen schließlich sind Mengen solcher Bewertungsfunktionen, die wir durch Anwendungen der mengentheoretischen Operationen der Komplement-, Vereinigungs- und Durchschnittsbildung aus den Elementarpropositionen erhalten. Diese mengentheoretischen Operationen entsprechen ja gerade der aussagenlogischen Negation, Adjunktion und Konjunktion für Sätze.

Beschreiben wir diese letzten Schritte ausführlicher und formal präziser: Unsere Schilderung ergab, daß nur das begriffliche System im Sinne von Stegmüller (1973a), S. 446, wirklich wesentlich ist: d.h. die Menge *Ind* von Individuen, die Menge T von Zeitindizes mit der Ordnungsrelation $\leq$ auf T und die Menge $\mathcal{F}$ von Attributfamilien F der Stellenzahl $n_F + 1$. Wir können also unsere Aufmerksamkeit auf dieses Begriffssystem beschränken.

Um dafür den Begriff der Proposition zu definieren, führen wir zunächst den sich als zentral erweisenden Begriff des Spielraums ein. Ein *Spielraum* oder, wie wir auch zuweilen sagen werden, ein *Faktor* ist ein Tripel $\langle F, a, t\rangle$, wobei F eine Attributfamilie aus $\mathcal{F}$, a ein n_F-Tupel von Individuen aus *Ind* und $t \in T$ ist.

Ein *Modell* – ganz im Sinne Carnaps (1971), S. 55 – ist dann eine Funktion, die jedem Spielraum $\langle F, a, t\rangle$ ein Element von F zuordnet. Ein solches Modell gibt also an, wie sich jeder Spielraum realisiert. Die Menge aller Modelle bildet den *Möglichkeitsraum W*.

Für jeden Spielraum $\langle F, a, t\rangle$ und jedes Attribut $p \in F$ nennen wir $\{w \mid w \in W, w(\langle F, a, t\rangle) = p\}$ eine *Elementarproposition*. Natürlich sind Elementarpropositionen *Propositionen*. Und wenn A und B Propositionen sind, so sind auch $W \setminus A$, $A \cap B$ und $A \cup B$ *Propositionen*. Man kann dies auch so ausdrücken, daß die Propositionen gerade die von den Elementarpropositionen erzeugte Algebra über W bilden. In dem hier gegebenen endlichen Fall ist die Menge aller Propositionen natürlich gerade Pot(W).

Überlegen wir uns nun, wie man in diesem Rahmen die möglichen Handlungen des Entscheidenden X unterbringen könnte. Das Begriffssystem, von dem wir ausgegangen sind, gibt dazu einige Anhaltspunkte. Man könnte gewisse Attributfamilien aus $\mathcal{F}$ als Handlungsattributfamilien auszeichnen. Freilich repräsentiert nicht jede Realisierung eines mit einer solchen Attributfamilie gebildeten Spielraums eine mögliche Handlung von X. Man wird allenfalls dann einen Spielraum $\langle F, a, t\rangle$ als Handlungsspielraum von X bezeichnen können, wenn F eine Handlungsattributfamilie ist und dann X an einer ganz bestimmten Stelle im n_F-Tupel a auftaucht. Freilich basiert schon die Unterscheidung zwischen Handlungs- und sonstigen Attributfamilien zur Gänze auf unserem intuitiven Verständnis von „Handlung“. Zur Klärung des Handlungsbegriffs tragen wir also nach wie vor nichts bei. Klar ist nur, daß die Handlungsspielräume eine Teilmenge der Menge aller Spielräume bilden. Vor einer abschließenden Definition ist noch zu bedenken, daß wir nicht von einem festen Begriffssystem ausgehen können; welche Spielräume in Betracht zu ziehen sind, hängt ganz von der jeweils gegebenen Entscheidungssituation ab. In der Tat ist es schon nicht sinnvoll, einer Entscheidungssituation alle von einem Begriffssystem erzeugten Spielräume zugrunde zu legen. Wenn Frau Schmidtlein den Kauf eines teuren Pelzmantels erwägt, so ist für sie vielleicht relevant, ob ihr der Pelzmantel steht, ob Herr Schmidtlein auf die Stute Mariechen gesetzt hat und ob Mariechen das Pferderennen gewinnt. Ganz unwichtig ist dagegen, ob Herr Schmidtlein auf Frau Schmidtlein gesetzt hat oder ob Mariechen der Pelzmantel steht. Natürlich könnte man auch solche Spielräume in die Entscheidungssituation einbeziehen, aber man sollte sich nicht von vornherein darauf festlegen, dies immer tun zu müssen. Da wir später ohnehin wenig Veranlassung haben werden, auf das jeweilige Begriffssystem zu rekurrieren, will ich nun ganz von ihm abstrahieren und nur die Struktur von Spielräumen und Propositionen charakterisieren. Da wir häufig von der zeitlichen Ordnung der Dinge absehen können, soll dies zunächst atemporal geschehen[4]:

Definition 8a: i ist ein *Spielraum* oder *Faktor* genau dann, wenn i eine endliche, mindestens zwei-elementige Menge ist.

[4] Definitionen und Sätze seien vom letzten Kapitel her durchlaufend numeriert.

Definition 8b: $\langle I, I^h\rangle$ ist ein *atemporales Entscheidungsfeld*[5] genau dann, wenn I eine endliche, nicht leere Menge von Spielräumen und I^h eine Teilmenge von I ist.[6] Die Elemente von I^h heißen *Handlungsspielräume* oder *-faktoren.* , die von $I \setminus I^h$, das wir auch mit I^g bezeichnen, *Geschehensspielräume* oder *-faktoren*. Als Variablen für die Elemente von I werden wir „i", „j", „k" und „l" verwenden, für Teilmengen von I „J", für Teilmengen von I^h „H" und für Teilmengen von I^g „K", „L", „M" und „N", jeweils mit oder ohne Indizes.

Relativ zu einem atemporalen Entscheidungsfeld $\langle I, I^h\rangle$ sei dann weiter definiert:

Definition 8c: W sei die Menge aller Funktionen auf I, die jedem Spielraum $i \in I$ ein Element von i zuordnet. Wir bezeichnen W als *Möglichkeitsraum*.

Definition 8d: Für jedes $i \in I$ und $p \in i$ heiße $E = \{w \mid w \in W, w(i) = p\}$ eine *Elementarproposition über i*. Für $J \subseteq I$ sei dann $\mathcal{A}(J)$ die von allen Elementarpropositionen über allen $i \in J$ erzeugte Algebra über W. Statt $\mathcal{A}(I)$ schreiben wir auch einfach $\mathcal{A}$ und statt $\mathcal{A}(\{i\})$ einfach $\mathcal{A}(i)$. Die Elemente von $\mathcal{A}(J)$ nennen wir *Propositionen* (*über J*) oder auch, wie es üblich ist, *Ereignisse* (*über J*). Die Elemente von $\mathcal{A}(I^h)$ heißen speziell *Handlungspropositionen*. Als Variablen für Propositionen allgemein verwenden wir „A", „B", „C", „D" und „E", für Handlungspropositionen „F" und „G", jeweils mit oder ohne Indizes.[7]

Definition 8e: Für $J \subseteq I$ gelte $A \in \mathcal{Z}(J)$ genau dann, wenn es zu jedem $i \in J$ eine Elementarproposition E_i über i gibt derart, daß $A = \bigcap_{i \in J} E_i$ Die Elemente von $\mathcal{Z}(J)$ nennen wir *Zustände von J*. Wiederum schreiben wir einfach $\mathcal{Z}$ statt $\mathcal{Z}(I)$ und für $i \in I$ $\mathcal{Z}(i)$ statt $\mathcal{Z}(\{i\})$.

Für $H \subseteq I^h$ bezeichnen wir die Elemente von $\mathcal{Z}(H)$ auch als *Handlungsverläufe in H*; falls $H = I^h$, reden wir von *Handlungsverläufen* schlechthin, und falls $H = \{i\}$ für ein $i \in I^h$, auch einfach von *Handlungen*. Die Menge aller Handlungsverläufe, also $\mathcal{Z}(I^h)$, bezeichnen wir auch mit $\mathcal{H}$.

---

[5] Diesen meines Erachtens recht passenden Ausdruck fand ich bei Gäfgen (1974), S. 199.

[6] Es ist also auch zugelassen, daß I^h leer ist, da sich dadurch einige der kommenden Sätze eleganter formulieren lassen. Natürlich wird man zur Repräsentation von Entscheidungssituationen immer ein nicht leeres I^h benötigen.

[7] Daß wir „F" auch für Attributfamilien benutzen, kann keine Verwirrung stiften, da in den folgenden Abschnitten von Attributfamilien nicht mehr die Rede sein wird.

Die Zustände von I seien auch *Gesamtzustände* und die Zustände von echten Teilmengen von I auch *Teilzustände* genannt.

Definition 8f: Für $H \subseteq I^h$ und $F, G \in \mathcal{H}$ gelte $F \approx_H G$ genau dann, wenn für jedes $i \in H$ und jede Elementarproposition E_i über i mit $F \subseteq E_i$ auch $G \subseteq E_i$ gilt. Wir sagen dann, daß die Handlungsverläufe F und G *in H gleich* sind.

Wie schon an der Vielfalt der Bezeichnungen deutlich wird, werden Zustände in allem Folgenden eine wesentliche Rolle spielen. Es sei darauf hingewiesen, daß $\mathcal{Z}(\emptyset) = \{W\}$ und daß hier im Endlichen immer $\mathcal{Z}(J) \subseteq \mathcal{A}(J)$ gilt; dabei enthält $\mathcal{Z}(J)$ gerade die (nach $\emptyset$) logisch schärfsten Propositionen aus $\mathcal{A}(J)$ oder, verbandstheoretisch gesprochen, die Atome von $\mathcal{A}(J)$.

Wie erwähnt, sind die Definitionen 8c-f (und auch die Definition von I^g) relativ zu einem Entscheidungsfeld. Es dürfte im folgenden immer klar sein, wenn wir all die obigen Bezeichnungen verwenden, relativ zu welchem Entscheidungsfeld sie zu verstehen sind. Und wenn je von mehreren Entscheidungsfeldern auf einmal die Rede sein sollte, etwa von $\langle I_1, I_1^h \rangle, ..., \langle I_n, I_n^h \rangle$, so ist ja $\langle \bigcup_{m=1}^{n} I_m, \bigcup_{m=1}^{n} I_m^h \rangle$ wiederum ein Entscheidungsfeld, und unsere Bezeichnungen sind dann relativ zu diesem umfassendsten Entscheidungsfeld erklärt.

Mit der folgenden Definition werden atemporale Entscheidungsfelder zeitlich strukturiert:

Definition 9a: $\langle I, I^h, T, \leq, s \rangle$ ist ein *temporales Entscheidungsfeld* genau dann wenn gilt:

(1) $\langle I, I^h \rangle$ ist ein atemporales Entscheidungsfeld,
(2) T ist eine nicht leere, endliche Menge,
(3) $\leq$ ist eine transitive, konnexe und antisymmetrische Relation, d.h. eine Ordnungsrelation auf T,
(4) s ist eine Funktion von I in T.

T stellt natürlich die Menge der Zeitindizes dar, die man in den verschiedenen, auf S. 92f. angegebenen Weisen verstehen kann. $\leq$ entspricht der Höchstens-so-spät-wie-Relation; ich werde mir gestatten, für $t, t' \in T$ auch die sich selbst erläuternden Wendungen „$t \geq t'$", „$t < t'$", „$t > t'$", und „$t = t'$" zu verwenden. Die

Funktion s schließlich weist jedem Spielraum $i \in T$ das Zeitintervall bzw. den Platz s_i in der zeitlichen Ordnung zu, in dem er sich realisiert. Wegen unserer Abstraktion vom Begriffssystem müssen wir die Funktion s extra anführen. Koppelten wir die Spielräume aus I wieder an ein Begriffssystem an, indem wir sie als Tripel $\langle F, a, t\rangle$ der geschilderten Art darstellen, so müßte natürlich immer $s_{\langle F,a,t\rangle} = t$ gelten.

Zuletzt sei noch definiert:

Definition 9b: Sei $\langle I, I^h, T, \leq, s\rangle$ ein temporales Entscheidungsfeld. Dann sei für $J \subseteq I$ und $t \in T$ $J_{\leq t} = \{i \mid i \in J \text{ und } s_i \leq t\}$. Analog seinen $J_{\geq t}$, $J_{\lhd t}$, $J_{>t}$ und $J_{=t}$ definiert.

Damit haben wir erst einmal alles Nötige beieinander. Ich möchte dabei betonen, daß in diesem und dem nächsten Kapitel immer und neben der noch einzuführenden nur die eben erklärte Notation verwandt wird.

Man wird einwenden, daß wir ja nun doch wieder bei irgendwelchen abstrakten Mengen gelandet seien; in der Tat wird der Bewanderte dahinter einfach die übliche wahrscheinlichkeitstheoretische Begriffsbildung aufs Endliche gestutzt und in etwas anderem Gewande erkennen. Das war auch so beabsichtigt; diese Begriffsbildung hat sich ja ausgezeichnet bewährt. Nur schien es mir im Hinblick auf die Explizitheitsfrage nützlich, im Rahmen einer Untersuchung wie dieser nicht nur von Propositionen zu reden, sondern auch einige Seiten darauf zu verwenden, wie sich die sonst nicht näher spezifizierten Mengen und die als Mengen aufgefaßten Propositionen nach Carnap aus Sprachen oder Begriffssystemen heraus konstruieren lassen. Die damit erzielte Klärung in der Explizitheitsfrage ist allerdings bescheiden. Die Unterscheidung zwischen Handlungs- und Geschehensspielräumen bleibt unanalysiert, und die mit dem Propositionsbegriff verknüpften sprachphilosophischen Schwierigkeiten, die wir, wie am Ende von Abschnitt 1.1 festgestellt, außer acht lassen wollen, werden dadurch nicht im mindestens verringert. (Vgl. auch die nächste Bemerkung.)

Mit Hilfe von Entscheidungsfeldern ist es nun auch möglich, Entscheidungssituationen eine extensive Analyse angedeihen zu lassen; wir können ja für jede relevante Einzelheit einen gesonderten Spielraum einführen. Natürlich hätten wir uns bereits im zweiten Kapitel die Mühe machen können, die dort auftretenden Mengen W, C und H in ähnlicher Weise aufzuschlüsseln; nur war das da noch nicht nötig. Völlig gesichert ist eine extensive Analyse allerdings immer noch nicht, selbst dann nicht, wenn man wieder direkt mit Modalitäten in Carnaps Sinn und mit Attributfamilien arbeitet. Denn durch eine extensive Analyse soll eine Entscheidungssituation ja in ihre logisch einfachen Bestandteile zerlegt werden – wobei das „logisch einfach" nicht relativ zu einer irgendwie gewählten formalen Sprache zu verstehen ist, sondern in einem absoluten Sinne gemeint ist. Dieser Sinn ist gewiß kein scharfer; vielfach wird sich eine logisch einfach anmutende Modalität in mehrere Modalitäten auflösen lassen, z.B. die Farbmodalität in Farbton, Sättigung und Intensität.[8] Eine Farbklassifizierung à la Goodman in „grot", „rün" etc. wird man aber sicherlich als eine logisch zusammengesetzte empfinden. Jedenfalls sind die Spielräume von Entscheidungsfeldern in diesem vagen absoluten Sinn als logisch einfach gedacht. Dies ist insbesondere für die Handlungsspielräume zu betonen, da damit ausgeschlossen ist, sie als Spielräume von Strategien aufzufassen. Dennoch ist festzuhalten, daß wir uns hier auf eine weitere intuitive, vortheoretische Unterscheidung stützen: die zwischen logisch einfachen und logisch zusammengesetzten Spielräumen.

Begeben wir uns nun an eine Reformulierung des Fishburn-Modells auf der neuen Basis. Gehen wir also, Zeitliches erst einmal außer acht lassend, davon aus, daß das atemporale Entscheidungsfeld $\langle I, I^h \rangle$ alle relevanten Faktoren einer bestimmten Entscheidungssituation, in der sich X befindet, abdeckt. Auf das „relevant" kommt es dabei noch nicht an; was es hier heißen könnte, betrifft die Abgrenzung einzelner Entscheidungssituationen und damit die Anwendungsfrage, die wir erst in den nächsten Abschnitten behandeln.

Wie sind nun die quantitativen Glaubensdispositionen von X in dieser Situation zu repräsentieren? Natürlich dürfen wir sie nicht durch ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$ darstellen. Einem ähnlichen Vorgehen verdankte ja die Theorie von Jeffrey ihre Unzulänglichkeit und die Konzeption von Luce und Krantz, daß sie die Theorie von Savage nicht an der ursprünglich beabsichtigten Stelle weiter entwickelte. Vielmehr müssen wir gemäß Abschnitt 2.5 peinlichst vermeiden, X irgendwelche subjektive Wahrscheinlichkeiten für seine eigenen Handlungen zuzuschreiben, und dürfen daher, eben Fishburn folgend, nur handlungsbedingte Wahrscheinlichkeiten für Propositionen, die über Handlungen nichts aussagen, zulassen. Genauer: Da kein Handlungsspielraum von vornherein als für die Wahrscheinlichkeit von Ereignissen irrelevant ausgeschaltet werden kann, dürfen nur Wahrscheinlichkeiten für Geschehenspropositionen betrachtet werden, die durch ganze Hand-

[8] Dieses Beispiel wählte Carnap (1971), S. 71.

lungsverläufe bedingt sind. Wir müssen also, formal ausgedrückt, annehmen, daß X für jeden Handlungsverlauf $F \in \mathcal{H}$ ein gesondertes subjektives Wahrscheinlichkeitsmaß auf $\mathcal{A}(I^g)$ hat. Tatsächlich werden wir aber X für jeden Handlungsverlauf $F \in \mathcal{H}$ ein Wahrscheinlichkeitsmaß auf ganz $\mathcal{A}$ zuschreiben, das F die Wahrscheinlichkeit 1 gibt. Dadurch werden spätere Formulierungen entschieden einfacher, und es ist klar, daß wir mit dieser trivialen Beifügung auf keinen Fall irgendwelche echte Handlungswahrscheinlichkeiten hereinschmuggeln.

Man mag sich fragen, ob die Handlungsabhängigkeit von Wahrscheinlichkeiten nun nicht zu kräftig ausgefallen ist. Wie weit sie sich zurücknehmen läßt, insbesondere wenn man die zeitliche Ordnung berücksichtigt, werden wir in den nächsten Abschnitten diskutieren. Einstweilen bleibt uns aber nichts anderes übrig, als von der eben angenommenen, extremen Handlungsabhängigkeit von Wahrscheinlichkeiten auszugehen. Definieren wir also:

Definition 10: Sei $\langle I, I^h\rangle$ ein atemporales Entscheidungsfeld. Dann sei P eine *Wahrscheinlichkeitsfamilie für* $\langle I, I^h\rangle$ genau dann, wenn P eine Funktion auf $\mathcal{H}$ ist derart, daß für jedes $F \in \mathcal{H}$ P_F ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$ mit $P_F(F) = 1$ ist.[9]

Was die Repräsentation der quantitativen Wünschensgrade von X betrifft, so befinden wir uns da in bester Übereinstimmung mit allen im zweiten Kapitel geschilderten Konzeptionen, wenn wir die subjektive Wertfunktion als eine reell-wertige Funktion auf der Menge $\mathcal{Z}$ der Gesamtzustände ansetzen.[10] Halten wir also fest:

Definition 11: Sei $\langle I, I^h\rangle$ ein atemporales Entscheidungsfeld. Dann sei V eine *subjektive Wertfunktion für I* genau dann, wenn V eine Funktion von $\mathcal{Z}(I)$ in $\mathbf{R}$ ist.

In Abschnitt 3.4 werden wir noch etwas ausführlicher auf subjektive Wertfunktionen zu sprechen kommen. Damit können wir zusammenfassen:

---

[9] $\langle W, \mathcal{A}, \mathcal{H}, P\rangle$ ist also ein Spezialfall dessen, was Renyi (1973), S. 57, eine bedingte Wahrscheinlichkeitsalgebra nennt und unter diesem Namen untersucht. Auch ist darauf hinzuweisen, daß Wahrscheinlichkeitsfamilien mit Markoffschen Kernen formal praktisch identisch sind; zu Markoffschen Kernen vgl. etwa Bauer (1968), §56. Man beachte schließlich, daß im Fall, daß $I^h = \emptyset$, P nur noch aus einem einzigen Wahrscheinlichkeitsmaß besteht; in diesem Fall ist ja $\mathcal{H} = \mathcal{Z}(\emptyset) = \{W\}$.

[10] Man bedenke dabei, daß $w \in W$ genau dann, wenn $\{w\} \in \mathcal{Z}$.

Definition 12: $\langle I, I^h, P, V, U\rangle$ ist ein *atemporales E1-Modell* genau dann, wenn gilt:

(1) $\langle I, I^h\rangle$ ist ein atemporales Entscheidungsfeld,

(2) P ist eine Wahrscheinlichkeitsfamilie für $\langle I, I^h\rangle$,

(3) V ist eine subjektive Wertfunktion für I,

(4) U ist die für alle $F \in \mathcal{H}$ durch $U(F) = \sum_{A \in Z} V(A) \cdot P_F(A)$ definierte Funktion von $\mathcal{H}$ in $\mathbf{R}$.

Zur Berücksichtigung von Zeitlichem sind zwei Zusätze nötig: Erstens ist zu bedenken, daß ein Entscheidungsmodell den handelnden X ja immer zu einem bestimmten Zeitpunkt t charakterisieren soll. Dies ist aber nur möglich, wenn I^h aus sich erst (relativ zu t) in der Zukunft realisierenden Handlungsspielräumen besteht; nur diese stehen X ja noch offen. Selbstverständlich kann man auch – und dies ist sogar die natürlichere Vorgehensweise – die Menge I^h zeitunabhängig vorgeben und dann $I^h_{>t}$ als die zu t relevante Handlungsspielraummenge nehmen.

Zweitens ist darauf hinzuweisen, daß sich bei Einbeziehung der zeitlichen Struktur eine gewisse Bedingung an Wahrscheinlichkeitsfamilien als sinnvoll herausstellen wird. Dies können wir jedoch erst im Abschnitt 3.3 erläutern, wo in Definition 17 (S. 115) Wahrscheinlichkeitsfamilien für temporale Entscheidungsfelder erklärt werden. Definieren wir also im Vorgriff darauf:

Definition 13: $\langle I, I^h, T, \leq, s, P, V, U\rangle$ ist ein *temporales E1-Modell für t* genau dann, wenn gilt:

(1) $\langle I, I^h, T, \leq, s\rangle$ ist ein temporales Entscheidungsfeld,

(2) $\langle I, I^h, P, V, U\rangle$ ist ein atemporales E1-Modell,

(3) $t \in T$ und $I^h \subseteq I_{>t}$,

(4) P ist eine Wahrscheinlichkeitsfamilie für $\langle I, I^h, T, \leq, s\rangle$.

Häufig wird die durch T, $\leq$ und s festgelegte zeitliche Struktur schon im Kontext vorgegeben sein, so daß wir dann bei der Formulierung von temporalen E1-Modellen auf die explizite Ausführung von T, $\leq$ und s verzichten können und werden. Ebenso werden wir uns des öfteren die Attribute „temporal“ und „atemporal“ sparen können.

Die mit E1-Modellen verbundene empirische Behauptung lautet: Wird X zum Zeitpunkt t durch das temporale E1-Modell $\langle I, I^h, T, \leq, s, P, V, U\rangle$ für t charakterisiert, so fängt X an, einen Handlungsverlauf $F \in \mathcal{H}$ zu realisieren, für den $U(F)$ maximal ist.

Mehr als daß X anfängt, einen Handlungsverlauf mit maximalem erwarteten subjektiven Wert zu realisieren, darf man sicherlich nicht behaupten; die Situation kann sich für X ja schon im nächsten Moment geändert haben. Aber auch so ist dies gewiß keine These, mit der man sich zufrieden geben kann; sie fällt sogar hinter die Thesen der Theorien aus dem zweiten Kapitel zurück. Denn es geht ja um die Ermittlung einer optimalen Strategie für X oder wenigstens einer optimalen nächsten Handlung, die freilich nichts anderes ist als die erste Handlung einer optimalen Strategie; und in den Thesen früherer Theorien wurde dies geleistet, da dort immer als Strategie interpretierbar war, was als Handlung bezeichnet wurde. Hier wird dagegen nur ein Handlungsverlauf als optimal ausgezeichnet, dessen zeitlich erste Handlung nicht einmal die erste Handlung einer optimalen Strategie zu sein braucht. Vertagen wir dieses Thema auf die Abschnitte 4.3 und 4.4, wo wir uns ausführlich mit Strategien beschäftigen.

Es dürfte klar sein, daß E1-Modelle ebenso adäquat und allgemein wie Fishburn-Modelle und daher allgemeiner als Savage- und Luce-Krantz-Modelle und adäquater als Jeffrey-Modelle sind. Dies kann man natürlich auch formal präzise durch den Beweis entsprechender Einbettungstheoreme zeigen; doch wollen wir es uns schenken, diese Übungsaufgabe jetzt vorzuexerzieren. Zwei kleine Unterschiede gegenüber Fishburn-Modellen bestehen allerdings. Erstens haben wir uns auf die etwas gefährlichen, durch Strategien bedingten Wahrscheinlichkeiten (vgl. S. 69) nicht eingelassen, weil wir uns eben noch gar nicht auf Strategien eingelassen haben. Und zweitens haben wir nebenbei den kleinen, auf S. 69f. erwähnten Schönheitsfehler ausgemerzt; Handlungen können in E1-Modellen als Bestandteile von Gesamtzuständen eigenständigen subjektiven Wert haben, ohne darin doppelt repräsentiert zu sein.

3.2 Probabilistische Unabhängigkeit

Im letzten Abschnitt haben wir im Rahmen der Explizitheitsfrage allgemein die formale Repräsentation der Gegenstände von (quantitativen) Glaubens- und Wünschensdispotitionen etwas genauer ausgeführt. Im Kontext der Anwendungsfrage interessiert uns nun, welche dieser Gegenstände wir für die Beschreibung einer bestimmten Entscheidungssituation benötigen oder, anders ausgedrückt, wie sich die für eine bestimmte Entscheidungssituation relevanten Spielräume auszeichnen lassen. Eines dürfte dabei von vornherein einleuchten, nämlich daß die Relevanz oder Irrelevanz gewisser Spielräume keine objektiv vorgegebene Angelegenheit ist, sondern sich nach den subjektiven Glaubens- und Wünschensdispositionen des Entscheidenden richtet. Dies wurde auch im Abschnitt 2.3 bereits deutlich. Dort wurde auch intuitiv plausibel, daß probabilistische Unabhängigkeit bei der Abgrenzung einzelner Entscheidungssituationen eine wichtige Rolle spielt, und in den Sätzen 1-4 (S. 58ff.) trat sie auch formal, wenngleich, wie sich dann herausstellte, nicht in der ihr zugedachten Funktion auf. Wenden wir uns ihr daher nun erst einmal ganz allgemein zu.

Wie in der Wahrscheinlichkeitstheorie konzentriert sich unser Interesse nicht auf die Unabhängigkeit zwischen einzelnen Ereignissen oder Propositionen, sondern auf die Unabhängigkeit zwischen Ereignisalgebren bzw. Zufallsvariablen, d.h. – in unserem Rahmen – zwischen Spielräumen und Mengen von Spielräumen. Und da ist dem Bekannten im Grunde nichts hinzuzufügen. Aus zwei Gründen müssen wir uns hier trotzdem etwas darüber verbreiten; erstens weil wir es hier mit Unabhängigkeit nicht gemäß einzelner Wahrscheinlichkeitsmaße, sondern gemäß den Nonstandard-Dingen zu tun haben, die wir als Wahrscheinlichkeitsfamilien definiert haben; und zweitens weil wir insbesondere im Hinblick auf die Abschnitte 3.3 und 3.6 auch an bedingter probabilistischer Unabhängigkeit interessiert sind, die in den Lehrbüchern nur sehr kurz oder gar nicht angeschnitten wird.

Gehen wir also von einem (atemporalen) Entscheidungsfeld $\langle I, I^h \rangle$ und einer Wahrscheinlichkeitsfamilie P für $\langle I, I^h \rangle$ aus, um zu erklären, was es für $J_1, J_2, J_3 \subseteq I$ heißt, daß J_1 von J_2 gemäß P durch J_3 bedingt unabhängig ist. Diesen Sachverhalt wollen wir durch „$J_1 \perp J_2 / J_3$“ symbolisieren.[11] Nicht bedingte Unabhängigkeit wird dann durch „$J_1 \perp J_2 / \emptyset$“ oder kurz durch „$J_1 \perp J_2$“ ausgedrückt. Schließlich wollen wir statt „$\{i\} \perp \{j\} / \{k\}$“ einfach „$i \perp j / k$“ schreiben.

Betrachten wir zunächst einmal nur Geschehensspielräume aus I^g. Hätten wir es nur mit einem Wahrscheinlichkeitsmaß Q auf $\mathcal{A}(I^g)$ zu tun, so lautete die übliche Definition, daß $K, L \subseteq I^g$ genau dann gemäß Q voneinander unabhängig sind, wenn die entsprechenden Algebren $\mathcal{A}(K)$ und $\mathcal{A}(L)$ voneinander unabhängig sind, wenn also für alle $A \in \mathcal{A}(K)$ und $B \in \mathcal{A}(L)$

$$Q(A \cap B) = Q(A) \cdot Q(B).$$

Und daß K und L gemäß Q durch $M \subseteq I^g$ bedingt unabhängig sind, hieße dann, daß für alle $A \in \mathcal{A}(K)$, $B \in \mathcal{A}(L)$ und $C \in \mathcal{Z}(M)$ mit $Q(C) \neq 0$

$$Q(A \cap B \mid C) = Q(A \mid C) \cdot Q(B \mid C)$$

gilt, daß also jedes Ereignis in $\mathcal{A}(K)$ von jedem Ereignis in $\mathcal{A}(L)$ unter jedem möglichen Zustand von M unabhängig ist. Nun haben wir es aber mit der Wahrscheinlichkeitsfamilie P zu tun, und (bedingte) Unabhängigkeit gemäß P kann dann sinnvollerweise nur heißen, daß die fraglichen Gleichungen eben für alle Maße P_F ($F \in \mathcal{H}$) gelten.

Wenn wir nun die Handlungsspielräume hinnehmen, so ist zuerst zu klären, wo in einer Aussage der Form $J_1 \perp J_2 / J_3$ überhaupt Handlungsspielräume auftauchen dürfen. Es sind da zwei Punkte zu beachten: Erstens ist es sinnvoll zu verlangen, daß die bedingende Menge J_3 keine Handlungsspielräume enthält, da wir es ohnehin mit lauter durch Handlungsverläufe bedingten Wahrscheinlichkeiten zu tun haben; diese noch einmal durch Handlungen zu bedingen, ergäbe keinen Sinn. Zweitens – und dies ist noch viel wichtiger – ist es nicht sinnvoll, einen Handlungspielraum als von irgendetwas probabilistisch abhängig oder unabhängig zu betrachten. Dies ist eine unmittelbare Folgerung unseres Prinzips aus Abschnitt 2.5, daß unser Entscheidender X seine eigenen Handlungen in keiner Weise probabilistisch beurteilt. Natürlich mag ein Handlungspielraum von einem Faktor in dem Sinne unabhängig sein, daß dieser Faktor keinerlei Einfluß auf Xs Entscheidung hat, wie er diesen Handlungsspielraum realisiert; doch ist diese Sorte Unabhängig-

[11] Die Definition von $\perp$ wird also relativ zu P sein. Es ist jedoch nicht nötig, dies explizit zu machen, da bei jeder künftigen Verwendung von „$\perp$“ klar sein wird, relativ zu welcher Wahrscheinlichkeitsfamilie sie zu verstehen sein wird.

keit nicht probabilistische Unabhängigkeit (wir explizieren sie in Abschnitt 3.5). Dies bedeutet, daß nur eine der beiden Mengen J_1 und J_2 Handlungsspielräume enthalten darf. Wir wollen festlegen, daß J_2 diejenige ist; damit gleichen wir die Aussage „$J_1 \perp J_2$" der umgangssprachlichen Wendung „J_1 ist von J_2 unabhängig" an. Man beachte, daß unsere Unabhängigkeitsrelation also (in den ersten zwei Stellen) nicht symmetrisch ist. Doch das ist nicht schlimm; die Wendung „ist unabhängig von" war noch nie symmetrisch. Zusammengefaßt heißt dies, daß die Aussage „$J_1 \perp J_2 / J_3$" nur sinnvoll ist, wenn J_1 und J_2 Teilmengen von I^g sind.

Daß $K \subseteq I^g$ von der Menge $H \subseteq I^h$ von Handlungsspielräumen gemäß P unabhängig ist, kann dann eigentlich nur heißen, daß für alle Handlungsverläufe F, $G \in \mathcal{H}$ und alle $A \in \mathcal{A}(K)$ $P_F(A) = P_G(A)$, sofern sich F und G allenfalls auf H unterscheiden, sofern also (mit Definition 8f, S. 97) $F \approx_{I^h \setminus H} G$.

Geht es um die durch ein $L \subseteq I^g$ bedingte Unabhängigkeit Ks von H, so müssen wir natürlich wiederum zu den durch die Zustände von L bedingten Wahrscheinlichkeiten übergehen. Und daß $K \subseteq I^g$ von einer beliebigen Menge $J \subseteq I$ unabhängig ist, kann schließlich nur heißen, daß K sowohl vom Handlungsteil wie vom Geschehensteil von J, d.h. von $J \cap I^h$ und von $J \cap I^g$ unabhängig ist. Nimmt man all dies zusammen, so ergibt sich

Definition 14: Sei $\langle I, I^h \rangle$ ein Entscheidungsfeld und P eine Wahrscheinlichkeitsfamilie für $\langle I, I^h \rangle$. Dann gelte für alle $J \subseteq I$ und alle $K, L \subseteq I^g$ $K \perp J / L$ genau dann, wenn für alle $A \in \mathcal{A}(K)$, $B \in \mathcal{A}(J \cap I^g)$, $F, G \in \mathcal{H}$ mit $F \approx_{I^h \setminus J} G$ und $C \in \mathcal{Z}(L)$ mit $P_F(C) \neq 0$ gilt:

$$P_F(A \cap B \mid C) = P_F(A \mid C) \cdot P_F(B \mid C)$$

und

$$P_F(A \mid C) = P_G(A \mid C).$$

Die wichtigsten Eigenschaften der so definierten *bedingten probabilistischen Unabhängigkeit* sind zusammengestellt in

Satz 8: Für alle $J, J' \subseteq I$, $H \subseteq I^h$ und $K, K', L, M, M', N \subseteq I^g$ gilt:

(a) $\varnothing \perp J / M$,

(b) wenn $K \perp L / M$, dann $L \perp K / M$,

(c) wenn $K \perp J / M$, $K' \subseteq K \cup M$, $J' \subseteq J \cup M$, und $M \subseteq M' \subseteq M \cup K \cup (J \cap I^g)$, dann $K' \perp J' / M'$,

(d) wenn $K \perp J \,/\, M$ und $K \perp H \,/\, M$, dann $K \perp J \cup H \,/\, M$,
(e) wenn $K \perp J \,/\, L \cup M$ und $K \perp L \,/\, M$, dann $K \perp J \cup L \,/\, M$,
(f) wenn $K \perp J \,/\, L \cup M$ und $L \perp J \,/\, M$, dann $K \cup L \perp J \,/\, M$,
(g) wenn $K \perp H \,/\, L$, und $K \perp H \,/\, M$, dann $K \perp H \,/\, L \cap M$,
(h) wenn $K \cup L \perp J \,/\, M$ und $K \cup M \perp J \,/\, L$, dann $K \cup L \cup M \perp J \,/\, L \cap M$,
(i) wenn $K \perp L \,/\, M$, dann $(K \cap L) \setminus M \perp I^g$,
(j) wenn $K \perp L \,/\, M$ und $N \perp I^g$, dann $K \cup N \perp L \,/\, M$.[12]

Für die nicht-bedingte Unabhängigkeit folgt daraus unmittelbar:

Satz 9: Für alle $J, J' \subseteq I$, $H \subseteq I^h$ und $K, K', L, N \subseteq I^g$ gilt:
(a) $\varnothing \perp J$,
(b) wenn $K \perp L$, dann $L \perp K$,
(c) wenn $K \perp J$, $K' \subseteq K$ und $J' \subseteq J$, dann $K' \perp J'$,
(d) wenn $K \perp J$ und $K \perp H$, dann $K \perp J \cup H$,
(e) wenn $K \cup L \perp J$ und $K \perp L$, dann $K \perp J \cup L$,
(f) wenn $K \perp J \cup L$ und $L \perp J$, dann $K \cup L \perp J$,
(g) wenn $K \perp L$, dann $K \cap L \perp I^g$,
(h) wenn $K \perp J$ und $N \perp I^g$, dann $K \cup N \perp L$.

Vor einer Diskussion dieser Sätze ist noch eine weitere technische Beobachtung zu machen, die sich für die Lösung der Anwendungsfrage als bedeutsam erweisen wird. Ähnlich wie im üblichen wahrscheinlichkeitstheoretischen Rahmen läßt sich nämlich I mit Hilfe von $\perp$ in voneinander unabhängige Spielraummengen zerlegen. Definieren wir sicherheitshalber zunächst einige Begriffe über Zerlegungen im allgemeinen und dann die von $\perp$ induzierten Zerlegungen von I.

Definition 15:
(a) Sei A eine nicht leere Menge. Dann sei $\{A_1, ..., A_n\}$ eine (endliche) *Zerlegung* von A genau dann, wenn $A_p \neq \varnothing$ und $A_p \cap A_q = \varnothing$ für $p \neq q$ ($p, q = 1, ..., n$) und $\bigcup_{p=1}^{n} A_p = A$.

[12] Für einen Beweis dieses Satzes auch für den unendlichen Fall s. Spohn (1980).

(b) Seien $\{A_1, ..., A_n\}$ und $\{B_1, ..., B_m\}$ Zerlegungen von A. Dann heißt $\{A_1, ..., A_n\}$ *feiner als* $\{B_1, ..., B_m\}$ bzw. $\{B_1, ..., B_m\}$ eine *Vergröberung von* $\{A_1, ..., A_n\}$ genau dann, wenn es zu jedem A_q $(q = 1, ..., n)$ ein B_p $(p = 1, ..., n)$ gibt, so daß $A_q \subseteq B_p$.
(c) Seien $\{A_1, ..., A_n\}$ und $\{B_1, ..., B_m\}$ Zerlegungen von A. Dann heißt $\{A_q \cap B_p \mid q = 1, ..., n; p = 1, ..., m\} \setminus \{\emptyset\}$ das *Produkt von* $\{A_1, ..., A_n\}$ *und* $\{B_1, ..., B_m\}$.

Definition 16: $\langle I, I^h\rangle$ ein Entscheidungsfeld und P eine Wahrscheinlichkeitsfamilie für $\langle I, I^h\rangle$. Dann heißt $\{J_1, ..., J_n\}$ eine *$\perp$-Zerlegung von I* genau dann, wenn $\{J_1, ..., J_n\}$ eine Zerlegung von I ist und wenn für alle $r = 1, ..., n$ $J_r \setminus I^h \perp I \setminus J_r$ gilt.

Es gilt dann der leicht zu beweisende

Satz 10:
(a) $\{I\}$ ist (trivialerweise) eine $\perp$-Zerlegung von I,
(b) mit $\{J_1, ..., J_n\}$ ist auch jede Vergröberung von $\{J_1, ..., J_n\}$ eine $\perp$-Zerlegung von I,
(c) sind $\{J_1, ..., J_n\}$ und $\{J_1', ... J_m'\}$ $\perp$-Zerlegungen von I, so auch ihr Produkt,
(d) es gibt eine feinste $\perp$-Zerlegung von I.

Der Leser wird schon ahnen, auf welche Weise $\perp$-Zerlegungen zur Abgrenzung einzelner Entscheidungssituationen herangezogen werden. Doch können wir dies erst ausführen, wenn wir die entsprechenden Dinge auf der Wünschensseite von Entscheidungsmodellen bereitgestellt haben. Überlegen wir uns aber trotzdem schon – unter der Annahme, daß wir von der Glaubensseite her nur $\perp$-Zerlegungen zur Isolierung einzelner Entscheidungssituationen benötigen –, wie befriedigende Vehikel $\perp$-Zerlegungen dazu sind:

Da ist zunächst festzustellen, daß $\perp$ selbst und somit auch $\perp$-Zerlegungen nur über eine Wahrscheinlichkeitsfamilie P definiert sind. Das bedeutet, daß man zur Loslösung begrenzter Entscheidungssituationen aus einer umfassenderen schon alle subjektiven Wahrscheinlichkeiten des Handelnden für diese umfassendere Situation voraussetzen muß – ein Tatbestand, den wir schon auf S. 54 mit Savage als unangenehm empfanden.

Diese starke Voraussetzung ließe sich beträchtlich mildern, wenn man eine rein qualitative Charakterisierung probabilistischer Unabhängigkeit hätte. Denn dann könnte man dem Handelnden zunächst eine solch rein qualitativ charakterisierte (bedingte oder nicht bedingte) Unabhängigkeitsrelation zuschreiben, und so zu den gewünschten Zerlegungen gelangen, ohne auf irgendetwas Quantitatives zurückgreifen zu müssen. Will man dem Handelnden dann noch eine Wahrscheinlichkeitsfamilie zuschreiben, so wäre von dieser eben zu verlangen, daß sie mit der vorgegebenen Unabhängigkeitsrelation übereinstimmt. Dies ist ja gerade der Weg, der Savage vorschwebte.[13]

Freilich bedarf es zur Durchführung dieses Programms einer vollständigen qualitativen Charakterisierung probabilistischer Unabhängigkeit. In der Tat bin ich davon überzeugt[14], daß Satz 8 so etwas liefert, d.h. genauer, daß folgendes gilt: Zu jeder dreistelligen Relation $\perp$ mit dem in Satz 8 angegebenen Eigenschaften gibt es eine Wahrscheinlichkeitsfamilie P derart, daß $\perp$ gerade die von P gemäß Definition 14 induzierte bedingte probabilistische Unabhängigkeitsrelation ist, vorausgesetzt die Spielräume in I sind groß genug. Und ebenso bin ich davon überzeugt, daß Satz 9 im selben Sinne eine vollständige Charakterisierung nicht-bedingter Unabhängigkeit gibt. Bewiesen ist das meines Wissens allerdings nicht. Der Zusatz, daß die Spielräume in I groß genug zu sein haben, ist dabei wesentlich; wie leicht nachzuprüfen ist, gilt jedenfalls noch das folgende Gesetz: Wenn $K \subseteq I^g$, $J \subseteq I$, $i \in I^g$, $K \perp J$ und $K \perp J / \{i\}$ und wenn der Spielraum i genau zwei Elemente hat, dann gilt $K \cup \{i\} \perp J$ oder $K \perp J \cup \{i\}$.

Nehmen wir nun einmal an, diese Vollständigkeitsvermutungen seien richtig. Damit kommt meines Erachtens nämlich erst die eigentliche Misere des geschilderten Weges zu $\perp$-Zerlegungen zum Vorschein. Denn Relationen wie in den Sätzen 8 und 9 geschildert ermangelt es völlig an intuitiven Gegenstücken, einfach deswegen, weil sie relativ komplizierten Gesetzen unterliegen. Das liegt zum Teil daran, daß für die probabilistische Unabhängigkeit von Handlungsspielräumen andere Gesetze gelten als für die von Geschehensspielräumen; insoweit kann man unseren Nonstandard-Rahmen für diese Kompliziertheit verantwortlich machen. Aber selbst wenn wir uns auf Geschehensspielräume beschränken, d.h. $I^h = \emptyset$ setzen würden, so wären die dann geltenden Gesetze immer noch unschön genug. Für entscheidend halte ich dabei vor allem, daß für die probabilistische Unabhängigkeit nicht gilt, was für jeden umgangssprachlichen Unabhängigkeitsbegriff zu gelten scheint: nämlich daß, wenn etwas von einer Sache und von einer anderen Sache unabhängig ist, es dann auch von beiden Sachen zusammen unabhängig ist,

[13] S. S. 62.

[14] Aufgrund meiner vergeblichen Beweisanstrengungen.

und umgekehrt, daß, wenn eine Sache von etwas abhängig ist und ebenso eine zweite, dann auch beide Sachen zusammen davon unabhängig sind. Und gerade darin liegt die formale Einfachheit umgangssprachlicher Unabhängigkeitsbegriffe begründet.[15] An diesem Sachverhalt ändert sich auch nichts, wenn die obigen Vollständigkeitsvermutungen falsch sind. Denn die dann bestehenden weiteren Gesetze machen probabilistische Unabhängigkeit, so viel ist sicher, allenfalls noch komplizierter und nicht entscheidend einfacher.

Dieser Mangel an intuitiven Gegenstücken macht es außerdem sehr schwierig, jemandem eine probabilistische Unabhängigkeitsrelation ohne Rückgriff auf seine subjektiven Wahrscheinlichkeiten zuzuschreiben; seine Äußerungen, was er wovon als unabhängig ansehe, geben dafür jedenfalls keine direkten Anhaltspunkte, da eben sein Unabhängigkeitsbegriff nicht der probabilistische sein wird. Es empfiehlt sich also, nach einem anderen Weg zu $\perp$-Zerlegungen Ausschau zu halten, der engeren Kontakt zum intuitiv Vorgegebenen hält. Und zumindest der Startpunkt eines solchen Weges scheint klar zu sein. Denn wenn von Unabhängigkeit innerhalb des Glaubensbereichs, also nach Maßgabe unserer empirischen Überzeugungen die Rede ist, so ist damit (neben logischer Unabhängigkeit, die uns hier natürlich nichts hilft) gewöhnlich so etwas wie kausale Unabhängigkeit gemeint. Jedenfalls ist unsere Neigung ungeheuer stark, Wendungen wie „ist unabhängig von" oder „hängt ab von" mit Worten zu paraphrasieren, die noch eindeutiger kausalen Sinn haben, z.B. als „hat keinen Einfluß auf" „wirkt sich aus auf", „beeinflußt" etc. Ob dieser Weg auch zum gewünschten Ziel führt, werden wir im nächsten Abschnitt sehen.

---

[15] Daß dieses und ähnliches für probabilistische Unabhängigkeit nicht gilt, hat wissenschaftstheoretischen Untersuchungen über Bestätigung und Relevanz regelmäßig Kopfzerbrechen bereitet. S. etwa Salmon (1975).

3.3 Kausale Unabhängigkeit und Abschirmbarkeit

Es geht also im folgenden um eine Explikation kausaler Unabhängigkeit; und damit irgendeine Chance auf eine Verbindung mit den $\perp$-Zerlegungen besteht, muß sich diese Explikation in einem probabilistischen Rahmen bewegen. Dies scheint freilich von vornherein ein schiefer Ansatz zu sein. Kausalität ist doch wohl unlösbar mit Determinismus verknüpft. Will man Kausalbegriffen zu Leibe rücken, so hätte dies also über die geltenden Naturgesetze zu geschehen, die ja erst Kausalzusammenhänge stiften. Wirklich?

Nein; das Problem der Kausalität ist eines, das man von sehr vielen verschiedenen Seiten anpacken kann.[16] Insbesondere sind die Zusammenhänge zwischen Kausalem und Probabilistischem bzw. die Vermutungen darüber so alt wie ausgeprägt. Spätestens seit Humes *Enquiry Concerning Human Understanding* sind sie nicht mehr aus der Welt zu schaffen. Auch sind hier einige objektivistische Wahrscheinlichkeitsinterpretationen ins Feld zu führen, die die Sache mit ihrer Annahme kurzschlossen, daß kausale Unabhängigkeit probabilistische Unabhängigkeit impliziere[17]; diese Annahme ermöglichte es ihnen, alle wahrscheinlichkeitstheoretischen Theoreme, die probabilistische Unabhängigkeitsvoraussetzungen machen, insbesondere also das für objektivistische Interpretationen so wesentliche Gesetz der großen Zahlen, anzuwenden, ohne vorher erst die Gültigkeit dieser Voraussetzungen statistisch rechtfertigen zu müssen. Und wenn man sich etwas in der vor allem in den Sozialwissenschaften genutzten multivariaten Analyse umtut, so merkt man, daß sich die Leute dort zwar sehr vorsichtig darüber auslassen, inwieweit ihre Methoden dazu taugen, Kausalzusammenhänge aufzuspüren, daß dies aber nichtsdestotrotz eines der Dinge ist, auf die sie gerne hinausmöchten. In der Wissenschaftstheorie schließlich ist z.B. der Versuch von Reichenbach (1956), Abschnitte 22 und 23, prominent, den kausalen Begriff der Abschirmbarkeit (dessen wir uns hier annehmen wollen) und allgemein Kausalordnungen probabilistisch zu charakterisieren (um daraus dann die zeitliche Ordnung zu entnehmen).

Neuerdings bemüht man sich sogar, Kausalbegriffe mittels probabilistischer Begriffe zu *definieren*. Am nachdrücklichsten tut dies Suppes (1970), der meint:

---

[16] S. etwa den vergleichenden Überblick von Domotor (1972).

[17] Vgl. etwa v. Kutschera (1969), S. 87.

> „From the standpoint of the philosophical analysis of causality probably the most confusing episode in the history of thought was the reign of Newtonian mechanics, from the beginning of the eighteenth century until the end of the nineteenth. An apparent universality and certainty of this mechanics led Kant and other philosophers into a mistaken notion of causality." (S. 6)

und daher von der Betrachtung ausgehend,

> „that the everyday concept of causality is not sharply deterministic in character"(S. 7)

zu zeigen versucht,

> „in what way causal notions of a probabilistic sort occupy an intuitive and natural place in scientific work" (S. 11).

Dies mündet dann in eine probabilistische Explikation der Wendung, daß ein Ereignis A eine *Ursache* für ein Ereignis B ist. Wir sind hier etwas weniger anspruchsvoll, da wir nur auf die kausale Abhängigkeit zwischen Spielräumen oder Faktoren aus sind; ansonsten lehnen wir uns aber eng an Suppes (1970) an, schon in unserer Vorgehensweise, die darin bestehen wird, möglichst viele probabilistische Argumente für und wider kausale Unabhängigkeit zu finden. Gehen uns diese Argumente aus, so können wir hoffen, alles über kausale Unabhängigkeit gesagt zu haben, was sich mit probabilistischen Mitteln darüber sagen läßt. Vorher sind aber noch einige Vorbemerkungen nötig.

Zunächst einmal ist klar, daß wir in unserem Rahmen Wahrscheinlichkeit subjektivistisch als Glaubensgrad einer bestimmten Person interpretieren; was wir dann explizieren, ist natürlich ebenfalls subjektivistisch zu interpretieren, nämlich als die Kausalvorstellungen dieser Person. Dennoch ist dieser Abschnitt, wenn er überhaupt interessant ist, auch für den Objektivisten interessant. Er kann ja aus dem Kontext der Entscheidungstheorie heraustreten, indem er $I^h = \varnothing$ setzt, und das eine Wahrscheinlichkeitsmaß, aus dem eine Wahrscheinlichkeitsfamilie dann noch besteht, objektivistisch interpretieren. So aufgefaßt, zielt unsere Explikation dann auf objektive kausale Unabhängigkeit ab. Diese Variabilität kann uns ja nur recht sein.

Zweitens ist klar, daß wir jetzt eine zeitliche Struktur voraussetzen müssen, wir haben hier nicht den Ehrgeiz, die zeitliche Ordnung erst aus der kausalen Ordnung abzuleiten. Legen wir also diesem Abschnitt durchweg das temporale Entscheidungsfeld $\langle I, I^h, T, \leq, s\rangle$ zugrunde.

Drittens kommt wieder einmal unser Prinzip aus Abschnitt 2.5 zur Anwendung. Denn ebenso wie es keinen Sinn ergab, von Handlungsspielräumen zu sagen, sie seien von etwas *probabilistisch* abhängig oder unabhängig, hat es diesem Prinzip zufolge keinen Sinn, von ihnen zu sagen, sie seien von etwas *kausal* abhängig oder unabhängig. Ein Entscheidender sieht seine eigenen Handlungen, deren Ausführung er erwägt, nicht als von bestimmten Ereignissen *verursacht* an; oder zumindest können wir ihm das von der Entscheidungstheorie her nicht unterstellen. Wir stellen daher die Frage, ob $J \subseteq I$ von $J' \subseteq I$ kausal abhängt oder nicht, nur, wenn $J \subseteq I^g$.

Schließlich ist gar nicht von vornherein klar, daß sich jede Wahrscheinlichkeitsfamilie für $\langle I, I^h\rangle$ für eine sinnvolle kausale Unabhängigkeitsrelation hergibt. Unsere Diskussion könnte ja ergeben, daß eine Wahrscheinlichkeitsfamilie bestimmten Bedingungen zu genügen hat, damit sich relativ zu ihr kausale Unabhängigkeit sinnvoll erklären läßt. Ob so ein Ergebnis interessant oder fatal wäre, braucht uns jetzt aber nicht zu kümmern, denn tatsächlich werden wir auf solche Bedingungen nicht stoßen – mit einer Ausnahme. Sie hängt allerdings vor allem schon damit, daß wir nun Zeitliches berücksichtigen, und weniger mit Kausalität zusammen und wird auch nur im Rahmen einer subjektivistischen Wahrscheinlichkeitsinterpretation, wo $I^h \neq \emptyset$, wichtig. Zu ihr führt folgende Überlegung:

Die verrückten Möglichkeiten, über die die Physiker anscheinend spekulieren, einmal außer acht lassend, gehen wir von der selbstverständlichen und praktisch unerschütterlichen Annahme aus, daß i von j *kausal* unabhängig ist, sofern $s_i < s_j$ ($i \in I^g, j \in I$). Dabei kann natürlich trotzdem nicht $i \perp j$ gelten, d.h. i von j *probabilistisch* abhängen; der Faktor j kann ja durchaus zusätzliche Information über den früheren Faktor i liefern. Problematisch wird das nur, wenn j ein Handlungsspielraum ist. Dann folgt nämlich, daß man keine von den eigenen zukünftigen Handlungen unabhängige epistemische Einschätzung der Vergangenheit, d.h. hier, keine nicht-bedingten Wahrscheinlichkeiten für vergangene Ereignisse haben kann (wobei wir ja annehmen können, daß s_i bereits der Vergangenheit angehört; s_j muß dagegen noch in der Zukunft liegen, da j ein offenstehender Handlungsspielraum sein soll). Denn beides zusammen, nicht-bedingte Wahrscheinlichkeiten für die Zustände von i und von den Handlungen aus $\mathcal{Z}(j)$ abhängige Wahrscheinlichkeiten für die Zustände von i (aufgrund derer erst nicht $i \perp j$ gilt), verletzt, wie wir uns im

Abschnitt 2.5 bereits überlegten, unser Verbot von Wahrscheinlichkeiten für Handlungen. Diese Folgerung ist aber, wie ich meine, völlig unakzeptabel; eine handlungsunabhängige Einschätzung der Vergangenheit scheint mir immer möglich zu sein. Hätten wir *nur* handlungsabhängige Wahrscheinlichkeiten für Vergangenes, so käme das sehr der Überzeugung gleich, rückwirkend noch auf die Vergangenheit Einfluß nehmen zu können. Und sobald wir uns eine handlungsunabhängige Einschätzung der Vergangenheit zugestehen, so müssen wir auch einräumen – dies schlossen wir gerade mit Abschnitt 2.5 –, daß wir *nur* sie haben. Schließlich ist noch zu bedenken, daß es für diese Überlegung nicht auf die Annahme, daß s_i schon tatsächlich vergangen ist, sondern nur darauf ankam, daß s_i irgendwann einmal in der Vergangenheit liegt, während s_j dann immer noch Zukunft ist. Wir haben dann das Ergebnis, daß innerhalb eines zeitlichen Rahmens sinnvollerweise nur Wahrscheinlichkeitsfamilien zugelassen werden können, nach denen für $i \in I^g$ und $j \in I^h$ mit $s_i < s_j$ $i \perp j$ bzw. allgemein für alle $t \in T$ $I^g_{\leq t} \perp I^h_{>t}$ gilt.

Dieses Ergebnis, das in Abschnitt 5.1 noch ganz wichtig werden wird, ist auch intuitiv und ohne Berufung auf Abschnitt 2.5 plausibel, auch wenn der erste Anschein wie im folgenden Beispiel dagegen spricht: *X* wird doch, so mag man vermuten, im Jahr 1978 die Wahrscheinlichkeit dafür, daß er am 15.9.1982 aus seiner Stelle entlassen wird, unter der Bedingung, daß er sich im Winter 82/83 ein neues Auto anschafft, niedriger einschätzen als unter der Bedingung. daß er sich in diesem Zeitraum kein neues Auto kauft; und man mag hinzusetzen, daß dies eine durch und durch gewöhnliche Sachlage sei. In der Tat. Bloß entkräftet sie unser Ergebnis nicht. *X*s Wahrscheinlichkeiten werden nur dann so aussehen, wenn er über seine Handlungen zu theoretisieren beginnt und sich z.B. überlegt, daß es doch reichlich unwahrscheinlich sein, daß er sich als frischgebackener Arbeitsloser ein neues Auto kauft. Doch fällt solches Theoretisieren über die eigenen Handlungen zumindest nicht unter die unmittelbare Zuständigkeit der Entscheidungstheorie.[18] Und wenn der Autokauf zum Winter 82/83 schon 1978 zur Entscheidung anstünde (etwa wegen 4-5-jähriger Lieferfristen), so würden *X*s Wahrscheinlichkeiten natürlich nicht wie eben beschrieben aussehen. Vielmehr wird seine Wahrscheinlichkeit dafür, entlassen zu werden, lediglich von seinem *vorherigen* Bemühen um die Stelle abhängen; und er wird sich überlegen, ob selbst bei intensivsten Bemühungen das Risiko für eine größere finanzielle Belastung nicht doch noch zu groß ist. (Möglicherweise wird er auch Auswirkungen von seiner Unterzeichnung

[18] Wie mehrfach angekündigt, führen wir dies in Abschnitt 5.2 noch weiter aus.

eines Kaufvertrages im Jahre 1978 auf seine Entlassung im Jahre 1982 sehen, etwa dergestalt, daß er so an der Butter zu sparen beginnt, daaß er aufgrund körperlicher Entkräftung beruflich versagt.) In keinem Fall aber wird der Autokauf selbst für die Entlassung probabilistisch relevant.

Fassen wir also zusammen, damit gleichzeitig die angekündigte notwendige Ergänzung zu Definition 13 (S. 102) liefernd:

Definition 17: Sei $\langle I, I^h, T, \leq, s\rangle$ ein temporales Entscheidungsfeld. Dann sei P eine Wahrscheinlichkeitsfamilie für $\langle I, I^h, T, \leq, s\rangle$ genau dann, wenn gilt:

(a) P ist eine Wahrscheinlichkeitsfamilie für $\langle I, I^h\rangle$

(b) für alle $t \in T$ gilt $I^g_{\leq t} \perp I^h_{>t}$.

Wir unterscheiden also Wahrscheinlichkeitsfamilien für atemporale und für temporale Entscheidungsfelder und führen die nun folgende probabilistische Explikation kausaler Unabhängigkeit relativ zu einer fest angenommenen Wahrscheinlichkeitsfamilie P für das bereits vorgegebene temporale Entscheidungsfeld durch.

Dazu führen wir erst einmal eine Notation ein: Daß $K \subseteq I^g$ von $J \subseteq I$ relativ zu P kausal unabhängig ist, wollten wir durch „$K \perp_c J$“ symbolisieren; wiederum schreiben wir einfach „$i \perp_c j$“ statt „$\{i\} \perp_c \{j\}$“. Ferner will ich eine Maßnahme ergreifen, mit der ich hoffe, die Lesbarkeit der folgenden, etwas unverdaulichen Seiten zu verbessern: Es soll im Zusammenhang mit einer Aussage der Form $K \perp J / L$ der Ausdruck „[t]“ für die Menge $I^g_{\leq t} \setminus (K \cup J)$ stehen ($t \in T$). Z.B. heißt also „$i \perp j$ / [s_j]“ dasselbe wie „$i \perp j / I^g_{\leq s_i} \setminus \{i, j\}$“. Solange man einen Ausdruck der Form [t] immer nur auf die probabilistische Unabhängigkeitsaussage bezieht, in der er vorkommt oder von der im jeweiligen Moment die Rede ist, können aus dieser Notation eigentlich keine Mißverständnisse erwachsen.

Damit schließlich die Sache nicht zu undurchschaubar wird, wollen wir vorerst zwei Vereinfachungen vornehmen. Erstens wollen wir die Diskussion zunächst auf kausale Unabhängigkeit zwischen einzelnen Faktoren beschränken, und zweitens wollen wir vorläufig annehmen, daß jeder Zeitindex höchstens einem Faktor zugeordnet ist; dadurch können wir die Frage verschieben, wie es mit der kausalen Abhängigkeit zwischen Faktoren mit gleichem Zeitindex steht.

Sei i also ein Faktor I^g und j ein Faktor in I, so daß $s_i > s_j$. Wann würden wir dann sagen, daß $i \perp_c j$ gilt? Vielleicht, wenn $i \perp j$? Doch ist dieser erste Tip unzulänglich. Es könnte ja ein $k \in I^g$ mit $s_k < s_j$ geben derart, daß nicht $i \perp j / k$. In diesem Fall würden wir j einen Einfluß auf i nicht absprechen wollen; vielmehr werden uns dadurch, daß die durch Zustände von k bedingten Wahrscheinlichkeiten unterschiedlich ausfallen, Einflüsse von j auf i nahegelegt; und daß sich diese Einflüsse nicht mehr in den zur Feststellung von $i \perp j$ benötigten absoluten Wahrscheinlichkeiten niederschlagen, würden wir damit erklären, daß sich diese Einflüsse bei einer Mittelung gegenseitig aufgehoben haben müssen: nämlich bei der Mittelung der durch die Zustände von k bedingten Wahrscheinlichkeiten mit den Wahrscheinlichkeiten für diese Zustände von k, die ja gemäß der Formel von der totalen Wahrscheinlichkeit[19] gerade diese absoluten Wahrscheinlichkeiten erzeugt. Kurz gesagt, würden wir also nicht auf das Fehlen jeglicher Einflüsse, sonder vielmehr auf das Vorhandensein versteckter Einflüsse von j auf i schließen.

Ist die kausale Unabhängigkeit von i gegenüber j wenigstens dann gesichert wenn nicht nur $i \perp j$, sondern auch $i \perp j \,/\, k$ für jedes $k \in [s_j] = I^g_{<s_j}$ gilt? Mitnichten. Es könnte ja immer noch zwei Faktoren $k, l \in [s_j]$ geben derart, daß nicht $i \perp j \,/\, \{k, l\}$. Wie zuvor würden wir dann einen Einfluß von j auf i nicht leugnen, sondern vielmehr von einem noch besser versteckten Einfluß reden. Setzt man diese Überlegung fort, so sind wir anscheinend zu der Folgerung genötigt, daß $i \perp_c j$ allenfalls dann gilt, wenn $i \perp j \,/\, L$ für alle Teilmengen L von $[s_j]$.

Diese Bedingung scheint freilich etwas stark geraten zu sein. Überlegen wir uns daher den entgegengesetzten Fall, wann wir sagen würden, daß nicht $i \perp_c j$. Der erste Tip „wenn nicht $i \perp j$" liegt wiederum schief. Es könnte ja ein $k \in I^g$ mit $s_k < s_j$ geben derart, daß $i \perp j \,/\, k$. In diesem Fall wäre der durch die Nicht-Gültigkeit von $i \perp j$ nahegelegte Einfluß von j auf i als reiner Scheineinfluß zu werten, und der Schein ließe sich dadurch erklären, daß die Einflüsse, die k anscheinend auf i und j ausübt[20], die Wahrscheinlichkeiten für die Zustände von $\{i, j\}$ so gestalten, daß der Eindruck entsteht, j beeinflusse i.

Damit nicht $i \perp_c j$ gilt, ist also nicht nur zu verlangen, daß nicht $i \perp j$, sondern auch, daß nicht $i \perp j \,/\, k$ für jedes $k \in [s_j]$. Doch reicht auch das wiederum nicht hin; es könnte immer noch zwei Spielräume $k, l \in [s_j]$ mit $i \perp j \,/\, \{k, l\}$ geben, und der vermutete Einfluß von j auf i wäre wieder als Scheineinfluß entlarvt. Wie zuvor landen wir anscheinend bei der Bedingung, daß nicht $i \perp_c j$ allenfalls dann gilt, wenn nicht $i \perp j \,/\, L$ für alle $L \subseteq [s_j]$.

---

[19] S. etwa Stegmüller (1973a), S. 156, und hier S. 170.

[20] Man beachte, daß aus Satz 8 (S. 107) folgt: wenn $i \perp j \,/\, k$ und nicht $i \perp j$, dann weder $i \perp k$ noch $j \perp k$.

Was lehren diese beiden Überlegungen? Vor allem dieses: Was immer durch die Gültigkeit bzw. Nicht-Gültigkeit von $i \perp j \,/\, L$ für ein $L \subseteq [s_j]$ über die kausale Unabhängigkeit von i gegenüber j nahegelegt wird, es läßt sich durch die Nicht-Gültigkeit bzw. Gültigkeit von $i \perp j \,/\, M$ für ein größeres $M \subseteq [s_j]$ entkräften. Mit anderen Worten: Für die Beurteilung von $i \perp_c j$ ist die Gültigkeit bzw. Nicht-Gültigkeit von $i \perp j \,/\, L$ solange irrelevant, wie es ein M mit $L \subset M \subseteq [s_j]$ gibt. Das heißt schließlich: Dafür, ob $i \perp_c j$ gilt oder nicht, ist allein entscheidend, ob $i \perp j \,/\, [s_j]$ gilt oder nicht.

Damit sind wir freilich noch nicht fertig. In unseren bisherigen Überlegungen haben wir nur die Faktoren vor j berücksichtigt. Doch haben auch die Spielräume, die zeitlich zwischen j und i liegen, ein Wörtchen mitzureden; sie gilt es nun einzubeziehen. Nehmen wir also an, daß $i \perp j \,/\, [s_j]$ gilt, was uns vermuten läßt, daß $i \perp_c j$. Es könnte dann, das überrascht uns mittlerweile nicht mehr, ein $k \in I^g$ mit $s_j < s_k < s_i$ geben derart, daß nicht $i \perp j \,/\, [s_j] \cup \{k\}$, und wie zuvor würden wir versteckte Einflüsse von j auf i wittern; die Situation hat sich um kein Jota geändert. Anscheinend ist $i \perp_c j$ erst wieder garantiert, wenn $i \perp j \,/\, L$ für jedes L mit $[s_j] \subseteq L \subseteq [s_i]$.

Allmählich gewinnen wir Routine. Doch müssen wir uns jetzt unvermutet auf einen neuen Fall einstellen. Denn nehmen wir an, daß nicht $i \perp j \,/\, [s_j]$, was nahelegt, daß nicht $i \perp_c j$, und nehmen wir weiter an, daß $i \perp j \,/\, [s_j] \cup \{k\}$ für ein $k \in I^g$ mit $s_j < s_k < s_i$. Würden wir dann wie zuvor unseren Verdacht, daß nicht $i \perp_c j$ zurückziehen? Nein, vielmehr würden wir sagen, daß da zwar ein Einfluß von j auf i sei und daß dieser aber nur indirekt sei und von dem Faktor k abgeschirmt würde. Wir würden diesen Einfluß also nicht als scheinbaren, sondern als abschirmbaren ansehen.

Damit haben wir ein erstes Ergebnis: Wenn nicht $i \perp j \,/\, L$ für ein L mit $[s_j] \subseteq L \subseteq [s_i]$ gilt, so ist dies ein unwiderlegbares Argument gegen $i \perp_c j$, da eine eventuelle durch eine größere Faktorenmenge bedingte probabilistische Unabhängigkeit von i und j lediglich eine abschirmbare und nicht eine scheinbar kausale Abhängigkeit anzeigt. Es gilt also tatsächlich

(3.1) $i \perp_c j$ genau dann, wenn $i \perp j \,/\, L$ für jedes L mit $[s_j] \subseteq L \subseteq [s_i]$.

Doch sind wir mittlerweile bei einem neuen Thema angelangt, nämlich bei der Abschirmbarkeit von i gegenüber j, die wir durch „$i \perp_a j$“ symbolisieren. Eine Situation, wie wir sie im vorletzten Absatz beschrieben haben, beweist aus den bekannten Gründen noch nicht, daß $i \perp_a j$: Es könnte einen weiteren Faktor l mit $s_j <$

$s_l < s_i$ geben derart, daß nicht $i \perp j \,/\, [s_j] \cup \{k, l\}$. In diesem Fall würden wir vielleicht immer noch behaupten, daß k den Einfluß von j auf i abschirmt; aber offensichtlich tut es das nicht sehr wirksam, denn bei Hinzunahme von l bricht der Einfluß von j auf i wieder durch. Der von k errichtete Schirm ist also nicht stabil, und es wäre dann unangemessen zu sagen, i sei vor j abschirmbar. Wiederum sind wir zur Betrachtung immer größerer bedingender Mengen und damit zu dem Schluß benötigt, daß der einzige garantiert stabile Schirm aus allen zwischen j und i liegenden Faktoren besteht. Unser zweites Resultat besagt somit, daß

(3.2) $i \perp_a j$ genau dann, wenn $i \perp j \,/\, [s_i]$.

Man beachte, daß gemäß (3.1) und (3.2) mit $i \perp_c j$ auch $i \perp_a j$ gilt, wie erwünscht. Wenn kein Einfluß von j auf i besteht, so sollte erst recht kein direkter, unabschirmbarer Einfluß von j auf i bestehen. $i \perp_c j$ ist natürlich stärker als $i \perp_a j$. Man könnte sagen, daß kausale Unabhängigkeit dann vorliegt, wenn bereits der leere Schirm stabil ist.

Genaugenommen haben wir noch nicht alle möglichen probabilistischen Argumente für und wider kausale Unabhängigkeit und Abschirmbarkeit erwogen; wir haben die Faktoren, die später sind als i, völlig vernachlässigt. Doch zu Recht. Denn wenn nicht $i \perp j \,/\, L$ für ein L gälte, das keine Teilmenge von $[s_i]$ ist, so würden wir lediglich schließen, daß die Faktoren in L, die später sind als i, weitere Informationen über i und j liefern, aber wir würden in keinem Fall unsere Absichten über die kausalen Beziehungen zwischen i und j ändern. Es bleibt daher bei (3.1) und (3.2).

Halten wir inne für einen kurzen Vergleich von (3.1) und (3.2) mit Suppes (1970). Wir finden bei ihm die gleichen Argumente dafür, daß ein Einfluß nur scheinbar bzw. indirekt ist (vg. dort S. 21ff. und S. 28ff.); ich brauchte sie nur meinem Rahmen anzupassen. Die andere Seite der Medaille, nämlich die Existenz versteckter Einflüsse scheint Suppes jedoch übersehen zu haben. Und es war gerade das Wechselspiel zwischen Argumenten für scheinbare und abschirmbare Einflüsse einerseits und für versteckte Einflüsse andererseits, das uns zu immer größeren bedingenden Mengen und damit zu (3.1) und (3.2) führte. Wahrscheinlich ließen sich dieselben Überlegungen auch innerhalb Suppes' Rahmen durchführen. Im übrigen korrespondieren unsere Ergebnisse ganz gut mit der üblichen Praxis zur Aufdeckung von Abhängigkeiten zwischen zwei Faktoren, einen der beiden

Faktoren unter Konstanthaltung aller übrigen als relevant vermuteten Faktoren zu variieren und die Auswirkungen davon auf den anderen Faktor zu beobachten.

Nun müssen wir noch unsere zwei vereinfachenden Annahmen zurücknehmen und (3.1) und (3.2) entsprechend verallgemeinern. Lassen wir also zuerst zu, daß mehrere Faktoren denselben Zeitindex tragen. Es stellen sich uns dann zwei Probleme. Das geringere davon ist dieses: Darf ein Faktor von sich selbst kausal abhängen? Das ist eine merkwürdige Frage, und ich glaube, die einfachste Methode, mit ihr fertig zu werden, ist, sie für sinnlos zu erklären. Wir legen also fest, daß die Aussage „$K \perp_c J$" nur sinnvoll ist, wenn $K \cap J = \varnothing$; dasselbe gilt natürlich für $\perp_a$. Man könnte freilich ebenso gut andere Regelungen treffen, es hängt wirklich nichts davon ab.

Die Hauptfrage ist natürlich: Sollten wir zulassen, daß ein Faktor von einem anderen mit gleichem Zeitindex beeinflußt wird? Suppes (1970) beantwortet diese Frage negativ[21], und im Gefolge der relativistischen Physik liegt dies besonders nahe. Aber wir müssen der modernen Physik nicht unbedingt gerecht werden. Im übrigen hätte die klassische Physik eher mit „ja" geantwortet. Auch ist zu bedenken, daß die Zeitindizes für Zeitintervalle stehen können; in der Mehrzahl der Anwendungen dürften sie es tatsächlich tun. Es kann also zwei Faktoren derselbe Zeitindex zugewiesen werden, obwohl sie nicht strikt simultan sind. In einem solchen Fall scheint mir eine liberale Position angemessen zu sein, d.h. in Unkenntnis der genauen zeitlichen Ordnung würde ich es zulassen, daß jeder der beiden Faktoren den anderen beeinflußt. Tatsächlich scheint es mir keinen zwingenden Grund für die eine oder andere Regelung zu geben. Ich gestatte mir daher, die technisch günstigere zu wählen, und dies ist die liberale Position. Genauer: Für $s_i \geq s_j$ ziehe ich die Faktoren k mit $s_k \leq s_j$ zur Beurteilung eines Einflusses von j auf i als scheinbarem und die Faktoren k mit $s_j < s_k \leq s_i$ zur Beurteilung eines Einflusses von j auf i als abschirmbaren heran. Und wenn wir uns die genaue Festlegung der Notation $[t]$ vergegenwärtigen, so zeigt sich, daß dann für $i \in I^g$ und $j \in I$ mit $i \neq j$ und $s_i \geq s_j$ (3.1) und (3.2) auch in diesem allgemeineren Fall angemessen sind.

Ich muß freilich gestehen, daß diese Regelung für die am Ende dieses Abschnitts formulierten Theoreme entscheidend ist. Ohne sie (oder eine ähnliche Regelung) gälten diese Theoreme nur, wenn verschiedene Faktoren verschiedene Zeitindizes tragen.

---

[21] Was auch daran liegt, daß er Ereignisse als instantan annimmt, Spielräumen also echte Zeitpunkte und keine Zeitintervalle zuordnet; vgl. Suppes (1970), S. 12.

Im letzten Schritt unserer Analyse gilt es, die Relationen $\perp_c$ und $\perp_a$ allgemein für *Mengen* von Faktoren zu erklären. Behandeln wir also zuerst die kausale Unabhängigkeit eingeschränkt auf den Spezialfall, daß ein einzelner Faktor von einer Faktorenmenge unabhängig ist. Konzentrieren wir uns dabei, um die Sache noch weiter zu vereinfachen, auf die Aussage „$i \perp_c \{j, k\}$". Da gibt es nur dann etwas zu überlegen, wenn $s_j, s_k \leq s_i$ und wenn $s_j \neq s_k$; denn wenn $s_j = s_k$. so gelten genau die früheren Überlegungen für $\{j, k\}$ anstelle von j. Nehmen wir also an, daß $s_k < s_j$. Es bieten sich dann zwei Möglichkeiten zur Verallgemeinerung von (3.1) auf diesen Fall an: nämlich daß $i \perp_c \{j, k\}$ heißt, daß $i \perp \{j, k\} / L$ entweder (a) für alle L mit $[s_k] \subseteq L \subseteq [s_i]$ oder (b) für alle L mit $[s_j] \subseteq L \subseteq [s_i]$. Ungeschickterweise eignet sich keine der beiden Möglichkeiten.

(a) ist zu stark. Denn nehmen wir an, daß nicht $i \perp \{j, k\} / [s_k]$ gilt. Würden wir das als schlüssige Widerlegung von $i \perp_c \{j, k\}$ akzeptieren? Nein; der zu vermutende Einfluß von j und k auf i mag allein von j ausgehen und sich dann durch die Verstärkung der bedingenden Menge immer noch als bloß scheinbarer erweisen lassen. Nur wenn auch nicht $i \perp k / [s_k]$ gilt, würden wir schließen, daß der Einfluß von j und k auf i zumindest zum Teil von k ausgeht und daß sich dieser Teil nur mehr abschirmen, aber nicht mehr als scheinbarer erweisen läßt.

Und (b) ist zu schwach. Denn nehmen wir an, daß nicht $i \perp k / [s_k]$. Nach unseren früheren Überlegungen besteht dann ein Einfluß von k auf i, und (b) schließt diesen Teil des Einflusses von j auf k und i nicht aus, sondern garantiert nur dessen Abschirmbarkeit.

Diese Gedanken zeigen deutlich, wie (a) abzuschwächen und (b) zu verstärken ist: nämlich so, daß $i \perp_c \{j, k\}$ genau dann gilt, wenn $i \perp \{j, k\} / L$ für jedes L mit $[s_j] \subseteq L \subseteq [s_i]$ (was mit (b) identisch ist) *und* wenn $i \perp k / L$ für jedes L mit $[s_k] \subseteq L \subseteq [s_i]$ (was schwächer als (a) ist). Verallgemeinert auf beliebige Faktorenmengen heiß dies: Für alle $i \in I^g$ und $J \subseteq I$ gilt

(3.3) $i \perp_c J$ genau dann, wenn für alle $t \leq s_i$: $i \perp J_{\trianglelefteq t} / L$ für jedes L mit $[t] \subseteq L \subseteq [s_i]$.

Der umgekehrte Fall der kausalen Unabhängigkeit einer Faktorenmenge von einem einzelnen Faktor ist ganz ähnlich. Betrachten wir wiederum die einfache Aussage „$\{k, i\} \perp_c j$", wobei wir ohne Beschränkung der Allgemeinheit annehmen können, daß $s_k > s_i \geq s_j$. Wie zuvor liegen zwei Möglichkeiten zur Verallgemeinerung von (3.1) besonders nahe: nämlich $\{k, i\} \perp_c j$ zu verstehen als $\{k, i\} \perp j / L$

entweder (a) für alle L mit $[s_j] \subseteq L \subseteq [s_k]$ oder (b) für alle L mit $[s_j] \subseteq L \subseteq [s_i]$. Doch ist (a) zu stark, weil es über i eine Aussage bedingter probabilistischer Unabhängigkeit macht, in der die bedingende Menge Faktoren enthält, die zeitlich später sind als i. Und (b) ist zu schwach, weil es einem eventuell vorhandenen, versteckten Einfluß von j auf k nicht alle Möglichkeiten versperrt, sich zu zeigen. Wiederum ist also (b) *und* die abgeschwächte Form von (a), daß $k \perp j \,/\, L$ für alle L mit $[s_j] \subseteq L \subseteq [s_k]$ zu verlangen. In voller Allgemeinheit bedeutet dies: Für alle $k \in I^g$ und $j \in I$ gilt

(3.4) $K \perp_c j$ genau dann, wenn für alle $t' \geq s_j$: $K_{\geq t'} \perp j \,/\, L$ für alle L mit $[s_j] \subseteq L \subseteq [t']$.

Wenn wir schließlich (3.3) und (3.4) kombinieren, so haben unsere Anstrengungen ein erstes Resultat gezeugt:

Definition 18: Sei $\langle I, I^h, T, \leq, s\rangle$ ein temporales Entscheidungsfeld, P eine Wahrscheinlichkeitsfamilie für $\langle I, I^h, T, \leq, s\rangle$, $K \subseteq I^g$ und $J \subseteq I$ mit $K \cap J = \varnothing$. Dann heißt *K von J kausal unabhängig* (relativ zu P), d.h. $K \perp_c J$, genau dann, wenn für alle $t, t' \in T$ mit $t \leq t'$ gilt:

$$K_{\geq t'} \perp J_{\leq t'} \,/\, L \text{ für alle } L \text{ mit } [t] \subseteq L \subseteq [t']$$

(d.h. für alle L mit $I^g_{\leq t} \setminus (K_{\geq t'} \cup J_{\leq t}) \subseteq L \subseteq I^g_{\leq t} \setminus (K_{\geq t'} \cup J_{\leq t})$).

Zur Verallgemeinerung der Abschirmbarkeitsaussage (3.2) können wir ganz ähnlich vorgehen: Der eine Schritt ist schnell getan und liefert für $i \in I^h$ und $J \subseteq I$:

(3.5) $i \perp_a J$ genau dann, wenn $i \perp I_{\leq s_j} \,/\, [s_i]$.

Wir haben hier einfach $I_{>s_j}$ vernachlässigt und den größtmöglichen Schirm angewandt, wie wir es auch in (3.1) getan haben, die früheren Überlegungen lassen sich hier direkt übertragen.

Der andere Schritt geht nicht so glatt. Betrachten wir also erst wieder die einfache Aussage „$\{k, i\} \perp_a j$“, wobei $s_k, s_i \geq s_j$. Wenn $s_k = s_i$, dann ist (3.2) mit $\{k, i\}$ anstelle von i immer noch angemessen; den früheren Überlegungen ist da nichts hinzuzufügen. Wenn jedoch $s_k > s_i$, so legen sich uns wieder zwei Verallgemeinerungen von (3.2) nahe: (a) $\{k, i\} \perp_a j$ genau dann, wenn $\{k, i\} \perp j \,/\, [s_k]$, und (b) $\{k,$

$i\} \perp_a j$ genau dann, wenn $\{k, i\} \perp j$ / $[s_i]$. Abermals sind beide ungeeignet. (a) verlangt, daß i von j bedingt durch (relativ zu i) zukünftige Faktoren probabilistisch unabhängig ist, und (b) errichtet nur einen möglicherweise unstabilen Schirm für k. Da wir für jeden zeitlichen Schnitt durch die abschirmende Menge $\{k, i\}$ einen maximalen Schirm verwenden müssen, besteht die einzige Lösung dieser Schwierigkeit darin zu verlangen, daß $\{k, i\} \perp_a j$ genau dann, wenn $i \perp j$ / $[s_i]$ und $k \perp j$ / $[s_k]$, oder verallgemeinert auf beliebige $K \subseteq I^g$, daß

(3.6) $K \perp_a j$ genau dann, wenn für alle $t \geq s_j$ $K_{=t} \perp j$ / $[t]$.

Die Kombination von (3.5) und (3.6) führt schließlich zu unserem zweiten Resultat:

Definition 19: Sei $\langle I, I^h, T, \leq, s\rangle$ ein temporales Entscheidungsfeld, P eine Wahrscheinlichkeitsfamilie für $\langle I, I^h, T, \leq, s\rangle$, $K \subseteq I^g$ und $J \subseteq I$ mit $K \cap J = \emptyset$. Dann sei *K vor J abschirmbar* (relativ zu P), d.h. $K \perp_a J$, genau dann, wenn für alle $t \in T$

$$K_{=t} \perp J_{\leq t} \,/\, [t]$$

(d.h. ausführlich $K_{=t} \perp J_{\leq t}$ / $I^g_{\leq t} \setminus (K_{=t} \cup J_{\leq t})$).

Ich fürchte, die Endprodukte unserer Explikation sind nun doch so kompliziert geraten, daß sie für den Leser intuitiv kaum noch durchschaubar und daher intuitiv wenig überzeugend sind. Mir ging es freilich ebenso; die Kompliziertheit scheint unvermeidlich. So habe ich eben verschiedene Explikationen ausprobiert und davon diejenige hier vorgelegt, die mir schließlich am überzeugendsten erschien – auch deswegen, weil sich für die so definierten Begriffe wünschenswerte Theoreme beweisen lassen (die allerdings auch bei leicht veränderten Explikationen gelten würden).

Das erste dieser Theoreme nennt die wichtigsten Eigenschaften von $\perp_c$ und $\perp_a$, wie sie in den Definitionen 18 und 19 definiert sind:

Satz 11: Für alle $K, K' \subseteq I^g$ und $J, J' \subseteq I$ gilt:
(a) wenn für alle $k \in K$ und $j \in J$ $s_k < s_j$, dann $K \perp_c J$,[22]
(b) wenn $K \perp_c J$, $K' \subseteq K$ und $J' \subseteq J$, dann $K' \perp_c J'$,

[22] Man beachte, daß daraus sowohl $K \perp_c \emptyset$ wie $\emptyset \perp_c J$ folgt.

(c) wenn $K \perp_c J$ und $K' \perp_c J$, dann $K \cup K' \perp_c J$,
(d) wenn $K \perp_c J$ und $K \perp_c J'$, dann $K \perp_c J \cup J'$.
Dasselbe gilt für $\perp_a$ anstelle von $\perp_c$.[23]

In der Tat glaube ich, daß Satz 11 in dem Sinne, in dem die Sätze 8 und 9 (S. 107), wie ich annahm, alles über das bedingte und das nicht-bedingte $\perp$ sagen, alles enthält, was sich über $\perp_c$ und über $\perp_a$ jeweils für sich sagen läßt. Doch ist dies ebenso wenig bewiesen wie die Vermutung über $\perp$. Die offenkundigsten Zusammenhänge zwischen $\perp_c$ und $\perp_a$ liefert der

Satz 12: Für alle $K \subseteq I^g$ und $J \subseteq I$ gilt:
(a) wenn $K \perp_c J$, dann $K \perp_a J$,
(b) wenn $K \perp_a J$, und wenn für alle $i \in I^g \setminus (K \cup J)$, $j \in J$ und $k \in K$ $s_j \leq s_k$ und entweder $s_i \leq s_j$ oder $s_k < s_i$, dann $K \perp_c J$.[24]

Wir werden später noch einen weiteren Zusammenhang zwischen $\perp_c$ und $\perp_a$ kennenlernen.

Sind diese Sätze so wünschenswert wie behauptet? Ich glaube schon. Satz 11(a) und (b) ist natürlich sowohl für $\perp_c$ wie für $\perp_a$ unverzichtbar: und von welcher Wichtigkeit Satz 11(c) und (d) für $\perp_c$ und $\perp_a$ ist, hatten wir schon am Ende von Abschnitt 3.2 festgestellt. Gerade deswegen sind $\perp_c$ und $\perp_a$ in ihren Eigenschaften ebenso einfach wie ihre intuitiven Gegenstücke. Satz 12(a) ist wiederum obligatorisch, und Satz 12(b) läßt sich in einer intuitiv recht überzeugenden Weise lesen: Wenn sich K vor J abschirmen läßt und wenn es in dem von K und J besetzten Zeitintervall keine Faktoren gibt außer denen von K und J selbst, dann muß es gerade der leere Schirm sein, der K vor J abschirmt, und wie auf S. 118 bereits bemerkt, bedeutet die Stabilität des leeren Schirms kausale Unabhängigkeit.

Waren aber nicht noch mehr Eigenschaften zu erwarten als die aus den Sätzen 11 und 12? Man mag vielleicht eine schwache Version des Kausalprinzips vermissen, d.h. die Eigenschaft, daß es zu jedem Faktor $i \in I^g$ einen Faktor aus I gibt, vom dem i kausal abhängt. Doch darf man dies sicherlich nicht allgemein erwarten. Die Faktoren in I können ja im Prinzip völlig willkürlich gewählt werden; es braucht

---

[23] Ein Beweis dieses Satzes und seiner Verallgemeinerung auf den unendlichen Fall findet sich in Spohn (1980).
[24] Ein Beweis dieses Satzes findet sich in Spohn (1980).

keinerlei Beziehung zwischen ihnen zu bestehen. Man könnte allenfalls verlangen, daß diese Eigenschaft in allen natürlichen Anwendungen oder auch in der universalen Anwendung, die alle existierenden Spielräume einbezieht, gilt; aber dem wollen wir jetzt nicht nachgehen.

Wahrscheinlich wird man jedoch die Transitivität der kausalen Abhängigkeit vermissen. (Ich wüßte nicht, was man sonst noch vermissen könnte.) zumindest wird man folgendes fordern wollen: Wenn sich i vor j und j vor k nicht abschirmen läßt, so hängt i von k kausal ab. Doch gilt selbst diese schwache Transitivität nicht (wie einfache numerische Beispiele zeigen), und sie sollte meines Erachtens auch nicht allgemein gelten. Nehmen wir z.B. an, daß die Faktoren in I den Wetterzustand an einem bestimmten Ort zu verschiedenen Tagen beschreiben. Bei einer hinreichend groben Wetterbeschreibung – und über die sind wir bis heute nicht hinausgelangt – gibt es dann bestimmt eine Zahl n derart, daß der kausale Einfluß des Wetters an einem Tag nach $2n$ Tagen völlig verschwunden und von den in unserer Beschreibung nicht erfaßten Zufälligkeiten total zugedeckt ist. Dabei können wir n so wählen, daß nach n Tagen noch ein Einfluß bemerkbar ist. (Als meteorologischer Laie vermute ich, daß $n = 7$ geeignet sein müßte.) Nehmen wir nun genauer an, daß die Faktoren in I den Wetterzustand jedes n-ten Tages beschreiben. Dann läßt sich offensichtlich jeder Faktor in I vor seinem unmittelbaren Vorgänger nicht abschirmen, und trotzdem hat kein Faktor Einfluß auf den übernächsten. Selbst die oben formulierte schwache Transitivität sollte also nicht innerhalb aller Faktorenmengen gelten. Wiederum mag man vielleicht sagen, daß sie in allen natürlichen Anwendungen gelten sollte; doch das steht auf einem anderen Blatt.

Es gäbe hier noch viele interessante Dinge zu tun. Z.B. könnte man den von uns nur informell verwandten Begriff des Schirmes präzise definieren, stabile und unstabile Schirme unterscheiden und ihre Eigenschaften untersuchen. Oder man könnte erkunden, unter welchen Bedingungen man wie weit die zeitliche Ordnung aus der kausalen rekonstruieren kann. Man kann auch das Vorhaben Suppes', den Begriff der Ursache zu explizieren, wieder angehen oder umgekehrt in Analogie zu Suppes (1970), S. 27 und S. 33, so etwas wie kausale Fast-Unabhängigkeit und Fast-Abschirmbarkeit definieren und untersuchen. Vor allem sollte man diese Ausführungen im Rahmen einer in diesem Abschnitt ja zugelassenen objektivistischen Wahrscheinlichkeitsinterpretation auf konkrete wissenschaftliche Beispiele aus der Theorie der stochastischen Prozesse, der Physik, der Statistik etc. anwenden und dann testen.

So verlockend dies alles wäre, es wird doch langsam Zeit, daß wir von unserem Ausflug zu unserem Ausgangspunkt, den $\perp$-Zerlegungen zurückkehren. Es liegt da zunächst nahe, ein kausales Analogon zu den $\perp$-Zerlegungen zu definieren, und wie das zu geschehen hat, dürfte klar sein:

Definition 20: $\{J_1, ..., J_n\}$ ist eine $\perp_c$- bzw. eine $\perp_a$-*Zerlegung von I* genau dann, wenn $\{J_1, ..., J_n\}$ eine Zerlegung von I ist und wenn für alle $r = 1, ..., n$ $J_r \cap I^g \perp_c I \setminus J_r$ bzw. $J_r \cap I^g \perp_a I \setminus J_r$ gilt.

Zwischen den Gliedern einer $\perp_c$-Zerlegung besteht dann keinerlei kausaler Zusammenhang. In der Tat könnte man definieren, daß zwei Faktoren i und j aus I genau dann *kausal zusammenhängen*, wenn es Faktoren $k_1,..., k_m$ gibt derart, daß $k_1 = i$, $k_m = j$ und für alle $r = 1, ..., n\text{-}1$ nicht $k_{r+1} \perp_c k_r$ oder nicht $k_r \perp_c k_{r+1}$. Offensichtlich ist diese Relation des kausalen Zusammenhangs eine Äquivalenzrelation in I, und mit Hilfe von Satz 11 ist leicht zu zeigen, daß die Äquivalenzklassen dieser Relation eine $\perp_c$-Zerlegung, und zwar die feinste $\perp_c$-Zerlegung, von I bilden.

Natürlich sind wir nun gespannt, was sich über das Verhältnis von $\perp$-, $\perp_c$-, und $\perp_a$-Zerlegung sagen läßt. Eine erschöpfende Antwort gibt

Satz 13: Für alle $J_1, ..., J_n$ sind die drei folgenden Aussagen äquivalent:
(a) $\{J_1, ..., J_n\}$ ist eine $\perp$-Zerlegung von I;
(b) $\{J_1, ..., J_n\}$ ist eine $\perp_c$-Zerlegung von I;
(c) $\{J_1, ..., J_n\}$ ist eine $\perp_a$-Zerlegung von I.[25]

Satz 10 (S. 108) gilt also auch für $\perp_c$- und $\perp_a$-Zerlegungen von I.

Die Äquivalenz von (b) und (c) durfte im übrigen erwartet werden. Sie besagt ja mit anderen Worten, daß es keinen Unterschied macht, ob man den oben definierten kausalen Zusammenhang mit Hilfe von $\perp_c$ (wie wir es oben taten) oder mit Hilfe von $\perp_a$ definiert; und es wäre schon ein Manko gewesen, wenn kausaler Zusammenhang über kleine direkte Schritte nicht ebenso weit gereicht hätte wie über größere indirekte Schritte. Diese Tatsache versöhnt vielleicht auch ein wenig mit

[25] Ein Beweis dieses Satzes findet sich wiederum in Spohn (1980). Dort ist er auch für unendlich große I bewiesen, vorausgesetzt, die Zeit ist zwar nicht unbedingt endlich, aber nach wie vor diskret.

der obigen Feststellung, daß kausale Abhängigkeit selbst in einem schwachen Sinne nicht transitiv ist.

Aber natürlich waren wir hauptsächlich auf die Äquivalenz von (a) und (b) aus. Sie leuchtet zwar intuitiv nicht unmittelbar ein, ist aber desto willkommener. Willkommen einmal aus allgemeinen Gründen: Sie stellt einen sehr einfachen Zusammenhang zwischen kausalen und probabilistischen Begriffen her, der ihren durch die vorausgegangenen Ausführungen nicht im geringsten gestützten, aber häufig zu beobachtenden sorglosen Austausch teils erklärt und teils rechtfertigt. Willkommen aber vor allem in unserem speziellen Kontext: Denn nun können wir damit beginnen, einem Individuum eine kausale Unabhängigkeitsrelation zuzuschreiben – inwiefern sich solche Relationen dazu besser eignen als probabilistische, hatten wir ja schon am Ende von Abschnitt 3.2 begründet. Diese kausale Unabhängigkeitsrelation zerlegt gemäß Definition 20 das Entscheidungsfeld des Individuums in kausal zusammenhängende Teile, und diese bilden nach Satz 13 gerade eine $\perp$-Zerlegung.

Ganz abgerundet ist die Sache freilich noch nicht; es fehlt nach wie vor eine rein qualitative Charakterisierung der fraglichen Unabhängigkeitsrelationen. Aber wir wissen nun immerhin, daß wir statt probabilistischer ebensogut kausale Unabhängigkeit qualitativ charakterisieren dürfen, da beiden demselben entscheidungstheoretischen Zweck dient.

3.4 Wertmässige Unabhängigkeit

Nachdem wir gerade eine etwas schwierige Wegstrecke hinter uns gebracht haben, befinden wir uns für den Rest des Kapitels wieder in ruhigen Gewässern. Wir sind ja immer noch hinter einer Antwort auf die Anwendungsfrage her, und dabei wird nicht nur probabilistische bzw., wie sich herausstellte, kausale Unabhängigkeit, sondern auch wertmäßige Unabhängigkeit wichtig – der wir übrigens schon begegneten, wie wir die Beschränkung von Savage-Modellen auf der Wünschensseite und deren Beseitigung in Luce-Krantz-Modellen schilderten.

Doch verdient wertmäßige Unabhängigkeit auch aus einem anderen Grunde unser Interesse. Dieser Grund wird klar, wenn wir uns eine weitere Abschweifung gestatten und etwas eingehender überlegen, welchen Dingen sinnvollerweise subjektiver Wert zukommen kann; gemäß E1-Modellen waren es ja nur die Zustände der Gesamtspielraummenge *I*, und diese Annahme sollte nicht ungeprüft hingenommen werden.

Eine Vorentscheidung in dieser Frage fiel schon im ersten Kapitel; dort legten wir uns mit gutem Grund darauf fest, als Gegenstände von G1aubens- und Wünschensdispositionen Propositionen oder, wie wir nun auch sagen, Ereignisse zu nehmen. Eine weitere Einschränkung bezüglich der quantitativen Wünschensdispositionen legte sich bei der Diskussion der Konzeption von Jeffrey in Abschnitt 2.6 nahe. Jeffrey wollte ja jeder Proposition – außer der logisch falschen – einen subjektiven Wert zuordnen, der sich, wie sich zeigte, bei den meisten Propositionen allenfalls als ein erwarteter subjektiver Wert verstehen ließ. Doch scheiterte dieses Vorhaben aus den auf S. 80 geschilderten Gründen, und daher empfahl sich das übliche Vorgehen, mit der subjektiven Wertfunktion nur nicht-erwartete oder, wie ich auch sagen will, absolute subjektive Werte auszudrücken und damit dann lediglich für Handlungen bzw. Handlungsverläufe erwartete subjektive Werte zu berechnen. Man wird einwerfen, absolute subjektive Werte seien ohnehin eine Fiktion. Dafür lassen sich gerade Savages „small worlds" anführen: Was sich in einem Entscheidungsmodell als sichere Konsequenz mit absolutem subjektivem Wert darstellt, entpuppt sich in einem umfassenderen Entscheidungsmodell als unsichere Konsequenz mit bloß erwartetem subjektivem Wert. Doch geht es gar nicht um nicht-erwartete subjektive Werte in irgendeinem absolutem

Sinn. Die Diskussion von Jeffreys Konzeption zeigte nur, daß man die subjektive Wertfunktion so konzipieren sollte, daß sie sich *relativ* zu dem Entscheidungsmodell, in das sie eingeht, als eine Funktion verstehen läßt, die nicht-erwartete, absolute subjektive Werte angibt. Und gerade so war ja die subjektive Wertfunktion etwa bei Savage und Fishburn konzipiert.

Der Gegenstandsbereich quantitativer Wünschensdispositionen umfaßt damit nicht mehr alle Ereignisse, sondern allenfalls noch Zustände. Sollte man ihn nicht gleich auf Zustände der Gesamtspielraummenge einschränken, d.h. sollte man subjektive Wertfunktionen nicht doch gerade gemäß Definition 11 (S. 101) konzipieren? Dies schiene mir voreilig; auch Teilzustände können einen absoluten subjektiven Wert tragen. Dies zeigt sich insbesondere an Zuständen, gegenüber denen man indifferent ist. So könnte mir ganz egal sein, welche Oper im Opernhaus gerade gespielt wird; alle Opern könnten mir gleich greulich sein und hätten demnach den gleichen niedrigen subjektiven Wert für mich. Und diese Aussage ist auch dann sinnvoll, wenn das Opernprogramm nur ein Spielraum unter vielen ist, die in eine Entscheidungssituation, in der ich mich befinde, eingehen. Doch ist Indifferenz nicht wesentlich. Ich kann z.B. Hotelzimmer mit Bad in einem gewissen Grad schätzen und gegen Hotelzimmer ohne Bad eine gewisse Abneigung hegen, unabhängig davon, wie lieb oder unlieb mir die anderen möglichen Zustände, z.B. die verschiedenen Zimmerpreise, sind, die für meine Entscheidung auch noch relevant sind.

Freilich braucht es nicht für jeden Zustand einen absoluten subjektiven Wert zu geben: Ob ich eine Frau der Mühe für wert halte, sie mit den Blicken zu verfolgen, hängt von vielen Faktoren ab: Figur, Gesicht, Kleidung, Gang, Gestik und Mimik etc.[26] Eine gute Figur kann in ihrem Eindruck von einem unbeholfenen Gang völlig verhunzt werden, ebenso wie eine mäßige Figur sich durch raffinierte Kleidung mächtig herausputzen läßt. Die Zustände jedes einzelnen Faktors haben so keinen absoluten subjektiven Wert für mich, sondern erst die Zustände aller Faktoren zusammengenommen. (Das Entscheidungsproblem, wo ich meinen Blick hinwenden soll, involviert natürlich noch unzählige weitere Faktoren.)

Manchen Zuständen kommt also ein absoluter subjektiver Wert zu, anderen nicht. Wonach richtet sich das? Die Beispiele legen das folgende Definitions-

---

[26] Daß man möglicherweise alles in einem einzigen Faktor, nämlich wie sexy sie ist, zusammenfassen könnte, tut dem Beispiel keinen Abbruch.

prinzip für subjektive Wertfunktionen nahe: *Ein Zustand erhält genau dann einen absoluten subjektiven Wert, wenn in ihm alle für ihn subjektiv-wertmäßig relevanten Details spezifiziert sind.* Dies scheint auch allgemein plausibel: Wird jemand gefragt, wie lieb ihm ein gewisser Zustand sei, in dem freilich nicht alle dem Befragten wertmäßig relevant erscheinenden Details aufgeführt sind, so stehen ihm zwei Reaktionsmöglichkeiten offen: Er kann die Frage entweder als eine Frage nach dem absoluten subjektiven Wert dieses Zustandes auffassen; dann wird er mit den Achseln zucken und sagen, das hänge doch noch davon ab, wie jene weiteren Details aussehen. Oder er wird, sofern das Sinn ergibt, die Frage als eine nach dem erwarteten subjektiven Wert dieses Zustandes verstehen, zu dessen Beurteilung er bedenken muß, mit welchen Wahrscheinlichkeiten die fehlenden Details so oder so ausfallen. In beiden Fällen hat nur der fragliche Zustand samt den fehlenden Details einen absoluten subjektiven Wert.

Wenn wir nun aber dieses Definitionsprinzip für subjektive Wertfunktionen formal präzisieren wollen, so benötigen wir – dies ist der angekündigte zweite Grund – eine formale Charakterisierung wertmäßiger Unabhängigkeit.

Das Standardverfahren, nach dem sich wertmäßige Unabhängigkeit in der Additivität der subjektiven Wertfunktion dokumentiert, ist schnell auf unseren Rahmen übertragen; dabei können wir Zeitliches außer acht lassen, und auch die Handlungsspielräume werden keine Sonderrolle einnehmen. Definieren wir also:

Definition 21: Sei $<I, I^h>$ ein atemporales Entscheidungsfeld und V eine subjektive Wertfunktion für I. Sei $K, I \subseteq I$ mit $K \cap L = \varnothing$. Dann heißt *K von L wertmäßig unabhängig* (relativ zu V), symbolisiert durch „$K \perp_w L$“, genau dann, wenn es K' und L' mit $K \subseteq K'$, $L \subseteq L'$, $K' \cap L' = \varnothing$ und $K' \cup L' = I$, eine Funktion V_1 von $Z(K')$ in $\mathbf{R}$ und eine Funktion V_2 von $Z(L')$ in $\mathbf{R}$ gibt derart, daß für alle $A \in Z(K\mathrm{K})$ und $B \in Z(L')$

$$V(A \cap B) = V_1(A) + V_2(B).$$

Schließen wir daran gleich die Definition der noch wichtig werdenden, von $\perp_w$ induzierten Zerlegungen von I an:

Definition 22: $\{J_1, ..., J_n\}$ heißt eine $\perp_w$-*Zerlegung von I* genau dann, wenn $\{J_1, ..., J_n\}$ eine Zerlegung von I ist und wenn für alle $r = 1, ..., n$ $J_r \perp_w I \setminus J_r$ gilt.

Alles Wissenswerte über die so definierten Dinge enthalten die folgenden Sätze:

Satz 14: Für alle $K, K', L, L' \subseteq I$ gilt:
(a) $\varnothing \perp_w K$,
(b) wenn $K \perp_w L$, dann $L \perp_w K$,
(c) wenn $K \perp_w L$, $K' \subseteq K$ und $L' \subseteq L$, dann $K' \perp_w L'$,
(d) wenn $K \perp_w L$, und $K' \perp_w L$, dann $K \cup K' \perp_w L$.

Satz 15:
(a) $\{I\}$ ist (trivialerweise) eine $\perp_w$-Zerlegung von I,
(b) mit $\{J_1, ..., J_n\}$ ist auch jede Vergröberung von $\{J_1, ..., J_n\}$ eine $\perp_w$-Zerlegung von I,
(c) sind $\{J_1, ..., J_n\}$ und $\{J_1',..., J_m'\}$ $\perp_w$-Zerlegungen von I, so auch ihr Produkt,
(d) es gibt eine feinste $\perp_w$-Zerlegung von I,
(e) sei $\{J_1, ..., J_n\}$ die feinste $\perp_w$-Zerlegung von I; dann gilt $K \perp_w L$ genau dann, wenn es kein $i \in K, j \in L$ und $r = 1, ..., n$ gibt mit $i, j \in J_r$.

Dazu erst einige eher technische Bemerkungen: Die Sätze 14 und 15 sind denkbar einfach zu beweisen; man scheut sich schon fast, sie als Sätze zu deklarieren. Aufgrund von Satz 15(e) leuchtet insbesondere ein, daß Satz 14 eine vollständige Charakterisierung von $\perp_w$ liefert, daß es also zu jeder zweistelligen Relation mit den in Satz 14 aufgeführten Eigenschaften eine subjektive Wertfunktion V für I gibt derart, daß diese Relation gerade die in Definition 21 aus V definierte Relation ist. Subjektive Wertfunktionen sind halt mathematisch wesentlich simplere Gebilde als Wahrscheinlichkeitsmaße oder gar -familien.

Ferner möchte ich auf zwei Verallgemeinerungsmöglichkeiten hinweisen, die aber später nicht wichtig werden. Erstens könnte man im Gegensatz zu Definition 21 zulassen, daß ein(e) Spielraum(menge) von sich selbst wertmäßig unabhängig ist. Das sind gerade die Spielräume, bei denen es einem völlig egal ist, wie sie sich realisieren. Formal ließe sich das in Definition 21 dadurch berücksichtigen, daß man die Bedingung „$K \cap L = \varnothing$“ streicht und für K' und L' verlangt, daß $K \setminus L \subseteq K'$, $L \setminus K \subseteq L'$, $K' \cap L' = \varnothing$ und $K' \cup L' = I \setminus (K \cap L)$. Es gälte dann zusätzlich nur: wenn $K \perp_w K$, dann $K \perp_w I$.

Zweitens läßt sich auch bedingte wertmäßige Unabhängigkeit definieren. Man könnte für paarweise disjunkte $K, L, M \subseteq I$ festlegen, daß $K \perp_w L / M$ genau dann, wenn es K' und L' mit $K \subseteq K'$, $L \subseteq L'$, $K' \cap L' = \varnothing$ und $K' \cup L' = I \setminus M$, V_1 auf $Z(K' \cup M)$ und V_2 auf $Z(L' \cup M)$ gibt derart, daß für alle $A \in Z(K')$, $B \in Z(L')$ und $C \in Z(M)$ $V(A \cap B \cap C) = V_1(A \cap C) + V(B \cap C)$. Es gälte dann z.B.: wenn $K \perp_w L / M$ und $M \subseteq M' \subseteq I \setminus (K \cup L)$, dann $K \perp_w L / M'$; und wenn $K \perp_w L \cup M_1 / M_2 \cup N$ und $K \perp_w L \cup M_2 / M_1 \cup N$, dann $K \perp_w L \cup M_1 \cup M_2 / N$. Insgesamt wird's also auch nicht sehr kompliziert.

Prüfen wir noch, wie intuitiv angemessen die in Satz 14 gegebene vollständige Charakterisierung wertmäßiger Unabhängigkeit ist. Nun, Satz 14(a), (c) und (d) sind ebenso unentbehrlich wie die entsprechenden Eigenschaften kausaler Unabhängigkeit; daß da so eine starke formale Ähnlichkeit besteht, scheint mir sehr wichtig. Was Satz 14(b), die Symmetrie wertmäßiger Unabhängigkeit, betrifft, so bin ich mir ihrer aufgrund von Gegenbeispielen wie dem folgenden weniger sicher: Ein Ehepaar sei gerade mit Nachwuchs gesegnet und daher mit der Namensgebung beschäftigt. Wenn es nun bei der Auswahl des Rufnamens – natürlich unter Berücksichtigung des Geschlechts des Kindes – allein nach dem Wohlklang der verschiedenen, sich anbietenden Namen geht, weitere Vornamen aber auf den Rufnamen abstimmt, so wäre für dieses Ehepaar der Rufname von den weiteren Vornamen des Kindes wertmäßig unabhängig, das Umgekehrte gälte aber nicht. Doch sind solche Fälle selten, und ich bin mir nicht sicher, ob sie wirklich überzeugende Gegenbeispiele darstellen.

Nutzen wir nun die wertmäßige Unabhängigkeit an den angesprochenen Stellen: Im Hinblick auf eine (im nächsten Abschnitt ausgeführte) Lösung der Anwendungsfrage liefert sie gemäß Definition 22 die dazu benötigten $\perp_w$-Zerlegungen (aus denen sie sich nach Satz 15(e) sogar wieder zurückgewinnen läßt). Und was wir im Abschnitt 3.3 zwar auf die (hoffentlich) richtige Bahn, aber nicht zu Ende brachten, fällt uns hier in den Schoß: Aufgrund der vollständigen Charakterisierung wertmäßiger Unabhängigkeit in Satz 14 können wir einem Individuum zunächst einmal eine rein qualitativ charakterisierte wertmäßige Unabhängigkeitsrelation zuschreiben und damit ohne Rückgriff auf Quantitatives einzelne Entscheidungssituationen abgrenzen; und es reicht, wenn wir ihm erst danach auch noch eine quantitative subjektive Wertfunktion in Übereinstimmung mit der bereits gegebenen Unabhängigkeitsrelation zuschreiben.

Außerdem können wir jetzt etwas präziser sagen, was die Gegenstände subjektiver Wertfunktionen sind: Nehmen wir an, $\perp_w$ sei die rein qualitativ charakterisierte wertmäßige Unabhängigkeitsrelation von X. Dann läßt sich die subjektive Wertfunktion von X genau dann für die Zustände einer Spielraummenge J erklären, wenn $J \perp_w I \setminus J$.[27] Denn das sind gerade die Zustände, die unserem Definitionsprinzip genügen, d.h. in denen alle für sie jeweils wertmäßig relevanten Details spezifiziert sind.

Technische Konsequenzen, etwa dergestalt, neben den in Definition 11 (S. 101) definierten subjektiven Wertfunktionen für I nun auch noch für Teilzustände definierte subjektive Wertfunktion einzuführen, brauchen wir daraus freilich nicht zu ziehen, und zwar aufgrund des folgenden, wohlbekannten und leicht zu beweisenden Sachverhalts. Definieren wir zunächst:

Definition 23: Sei V eine subjektive Wertfunktion für I und $\{J_1, ..., J_n\}$ eine Zerlegung von I. Dann heiße $\{V_1, ..., V_n\}$ eine *additive Zerlegung* von V bzgl. $\{J_1, ..., J_n\}$ genau dann, wenn für alle $r = 1, ..., n$ V_r eine subjektive Wertfunktion für J_r ist und wenn für alle $A_r \in \mathcal{Z}(J_r)$ $(r = 1, ..., n)$

$$(3.7) \qquad V(\bigcap_{r=1}^{n} A_r) = \sum_{r=1}^{n} V_r(A_r) \; .$$

Es gilt dann der

Satz 16: Sei V eine subjektive Wertfunktion für I und $\{J_1, ..., J_n\}$ eine Zerlegung von I. Dann gibt es genau dann eine additive Zerlegung von V bzgl. $\{J_1, ..., J_n\}$ wenn $\{J_1, ..., J_n\}$ eine $\perp_w$-Zerlegung von I ist; und wenn $\{V_1, ..., V_n\}$ und $\{V_1', ..., V_n'\}$ additive Zerlegungen von V bzgl. $\{J_1, ..., J_n\}$ sind, so gibt es reelle Zahlen z_r $(r = 1, ...,n)$ mit $\sum_{r=1}^{n} z_r = 0$, so daß für alle $A_r \in \mathcal{Z}(J_r)$ $V_r'(A_r) = V_r(A_r) + z_r$ $(r = 1, ..., n)$.

Wenn also $(J, I \setminus J)$ eine $\perp_w$-Zerlegung von I bilden muß, damit die Zustände von J subjektiven Wert tragen können, so folgt aus diesem Satz, daß sich die subjektiven Werte der Zustände von J aus den subjektiven Werten der Gesamt-

[27] Und wenn $J \neq \varnothing$; denn $\mathcal{Z}(\varnothing)$ enthält als einziges Element das sichere Ereignis W, und das ist kein Zustand im intuitiven Sinne, dem ein absoluter subjektiver Wert zukommt.

zustände bis auf eine additive Konstante eindeutig entnehmen lassen. Dabei braucht uns diese additive Konstante nicht zu stören, werden subjektive Werte doch ohnehin nur auf einer Intervallskala gemessen. Insofern können wir es im folgenden bei den nur für Gesamtzustände definierten subjektiven Wertfunktionen belassen.

Daß unsere Überlegungen über die Träger subjektiver Werte keine technischen Neuerungen nach sich zogen, liegt allerdings vor allem daran, daß wir hier wertmäßige Unabhängigkeit nur insoweit untersucht haben, als sie sich in subjektiven Wertfunktionen für Gesamtzustände niederschlägt und daher auch wieder aus ihnen herausholen läßt. Mir scheint es freilich nicht selbstverständlich zu sein, daß sich *jegliche* wertmäßige Unabhängigkeit in den subjektiven Werten der Gesamtzustände widerspiegeln muß. Ginge man davon ab, so ließe sich an dem hier wiedergegebenen Standardbild einiges ändern. Man könnte dann z.B. auch nicht-symmetrische wertmäßige Unabhängigkeitsrelationen in Betracht ziehen; unser Definitionsprinzip erzwänge dann tatsächlich technische Konsequenzen; etc. Doch wollen wir alledem jetzt nicht nachgehen.

3.5 Zerlegungen von Entscheidungsmodellen

Nachdem wir die Anwendungsfrage als Aufhänger für die vorangegangenen Abschnitte verwandt haben, wollen wir uns ihr nun endgültig zuwenden. Das Problem, um das es geht, ist, genau formuliert, dies: Wie soll man aus der Unmenge an Spielräumen, die sich formulieren und im Prinzip alle in Entscheidungsfelder aufnehmen lassen, diejenigen Spielräume heraussuchen und zu einem Entscheidungsfeld zusammenfassen, die man zur Repräsentation gerade einer einzelnen, selbständigen Entscheidungssituation benötigt? Die Überlegung, die dieses Problem mit Hilfe probabilistischer bzw. kausaler und wertmäßiger Unabhängigkeitsrelationen und den zugehörigen $\perp$- bzw. $\perp_c$- und $\perp_w$-Zerlegungen angeht und die unausgesprochenen bereits hinter unseren Ausführungen in den zwei vorangegangenen Abschnitten stand, ist sehr einfach. Gesucht wird ja diejenige Spielraummenge – sie heiße I_0 –, die wir zur Repräsentation einer Entscheidungssituation benötigen, in der es für unser Subjekt *X* darum geht, eine geeignete Handlung aus einem bestimmten Handlungsspielraum – nennen wir ihn *i* – auszuwählen, und die erhalten wir nun so:

Natürlich muß $i \in I_0$ gelten. Doch auch alle Spielräume, die von *i* kausal abhängen, müssen in I_0 enthalten sein; ferner alle Spielräume, die von diesen Spielräumen kausal abhängen oder von denen diese Spielräume kausal abhängen – kausale Abhängigkeit immer als subjektive kausale Abhängigkeit nach Maßgabe von *X* verstanden. Etc. D.h. dasjenige Element der feinsten $\perp_w$-Zerlegung der Menge aller denkbarer Spielräume, das *i* enthält – nennen wir es K_0 –, muß Teilmenge von I_0 sein; dieses Element enthält ja gerade alle Spielräume, die direkt oder indirekt mit *i* probabilistisch oder kausal zusammenhängen. Damit haben wir I_0 aber noch nicht bestimmt. Denn I_0 muß außerdem alle Spielräume umfassen, die mit *i* wertmäßig zusammenhängen, und ferner wiederum alle Spielräume, die mit diesen Spielräumen wertmäßig zusammenhängen, etc. D.h. dasjenige Element der feinsten $\perp_w$-Zerlegung der Menge aller denkbaren Spielräume, das *i* enthält – nennen wir es L_0 – muß ebenfalls Teilmenge von I_0 sein. Doch auch dies reicht noch nicht. Mit den Elementen von K_0 können neue Spielräume wertmäßig zusammenhängen; ebenso können mit den Elementen von L_0 weitere Spielräume kausal verbunden sein; und diese sind ebenfalls zu I_0 zu rechnen. Und so weiter. Wir müssen also verlangen, daß I_0 eine sowohl bezüglich kausaler wie bezüglich wertmäßiger Abhängigkeit abgeschlossene Menge von Spielräumen ist, deren Elemente direkt oder indirekt mit i kausal oder wertmäßig zusammenhängen. Um dies formal auszudrücken, benötigen wir die folgende Definition:

Definition 24: Sei $\langle I, I^h, P, V, U\rangle$ ein atemporales E1-Modell. Dann heiße $\{J_1, ..., J_n\}$ eine $\perp_{cw}$-Zerlegung von I genau dann, wenn $\{J_1, ..., J_n\}$ eine $\perp_c$-Zerlegung und eine $\perp_w$-Zerlegung von I ist.

Elemente von $\perp_{cw}$-Zerlegungen sind also gerade bezüglich kausaler und wertmäßiger Abhängigkeit abgeschlossene Spielraummengen. Aufgrund der Sätze 10 (S. 108) und 15 (S. 131) ist klar, daß es immer eine feinste $\perp_{cw}$-Zerlegung von I gibt und daß genau die Vergröberungen dieser feinsten $\perp_{cw}$-Zerlegung die $\perp_{cw}$-Zerlegungen von I sind.

Das Ergebnis unserer Diskussion ist also, daß I_0 gerade dasjenige Element der feinsten $\perp_{cw}$-Zerlegung der Menge aller denkbaren Spielräume sein muß, das i enthält. Kleiner darf I_0 nicht sein, größer braucht es nicht zu sein. Wir benötigen genau die Menge I_0, um die eine Entscheidungssituation unseres Subjektes X zu repräsentieren, in der es, unter anderem, um den speziellen Handlungsspielraum i geht.

Natürlich können wir von i abstrahieren. Was bleibt, ist die Feststellung, daß minimale, gegenüber kausaler bzw. probabilistischer und wertmäßiger Abhängigkeit abgeschlossene Spielraummengen gerade die Mengen sind, mit denen sich einzelne, isolierbare Entscheidungssituationen erfassen lassen. Daher die folgende

Definition 25: $\langle I, I^h, P, V, U\rangle$ ist ein *zusammenhängendes atemporales E1-Modell* genau dann, wenn gilt.

(1) $\langle I, I^h, P, V, U\rangle$ ist ein atemporales E1-Modell,

(2) $\{I\}$ ist die feinste (und damit die einzige) $\perp_{cw}$-Zerlegung von I.

Für temporale E1-Modelle lautet die Definition natürlich ebenso.

Damit scheint mir expliziert zu sein, was unter einem Entscheidungsproblem oder einer Entscheidungssituation zu verstehen ist, was also die Gegenstände der Entscheidungstheorie sind. Diese Klärung ist freilich eine theoretische in dem Sinne, daß man zur Abgrenzung einzelner Entscheidungssituationen wesentlich auf die Glaubens- und Wünschensdispositionen des Entscheidenden Bezug nehmen muß. Doch scheint mir dieser Bezug unvermeidlich zu sein. Welche Spielräume bei einem Entscheidungsproblem eines Subjektes eine Rolle spielen, ist eben nicht objektiv vorgegeben, sondern hängt von seinen Glaubens- und Wünschensdispositionen ab. Wären nun die kausale und die wertmäßige Unabhängigkeitsrelation des Subjektes rein qualitativ und ohne Rückgriff auf seine quantitativen Glaubens- und Wünschensgrade zugänglich, so fiele dieser Bezug außerdem recht milde aus. Bezüglich wertmäßiger Unabhängigkeit war dies einfach, bezüglich der kausalen ist es aber noch offen.

Eine in der Theorie selbst liegende und insofern vielleicht noch bessere Rechtfertigung für unsere Beantwortung der Anwendungsfrage liefert die Tatsache, daß sich jedes Entscheidungsmodell in lauter zusammenhängende Entscheidungsmodelle zerlegen und aus ihnen auch wieder zusammensetzen läßt, wobei innerhalb dieser zusammenhängenden Modelle getrennt entschieden werden kann, d.h. daß ihre optimalen Handlungsverläufe zusammen gerade einen optimalen Handlungsverlauf des Ausgangsmodells ergeben. Dies ist der Inhalt der folgenden Sätze:

Satz 17: $\langle I, I^h, P, V, U\rangle$ ein E1-Modell, $\{J_1, ..., J_n\}$ eine $\perp_{\mathrm{cw}}$-Zerlegung von I und $H_r = J_r \cap I^h$ $(r = 1, ..., n)$. Sei für alle $r = 1, ..., n$ $F_r \in \mathcal{Z}(H_r)$ und $A \in \mathcal{A}(J_r)$ $P^r_{F_r}(A) = P_F(A)$ für ein $F \in \mathcal{Z}(I^h)$ mit $F \subseteq F_r$. Dadurch ist für alle $r = 1, ..., n$ $P^r_{F_r}$ wohldefiniert und P^r eine Wahrscheinlichkeitsfamilie für $\langle J_r, H_r\rangle$. Sei weiter $\{V_1, ..., V_n\}$ eine additive Zerlegung von V bzgl. $\{J_1, ..., J_n\}$. Sei schließlich für alle $F_r \in \mathcal{Z}(H_r)$

$$U_r(F_r) = \sum_{A \in \mathcal{Z}(J_r)} V_r(A) \cdot P^r_{F_r}(A)$$

$(r = 1, ..., n)$. Dann ist für jedes $r = 1, ..., n$ $\langle J_r, H_r, P^r, V_r, U_r\rangle$ ein E1-Modell und, sofern $\{J_1, ..., J_n\}$ die feinste $\perp_{\mathrm{cw}}$-Zerlegung von I ist, sogar ein zusammenhängendes E1-Modell, und es gilt für alle $A_r \in \mathcal{A}(J_r)$, $F_r \in \mathcal{Z}(H_r)$ ($r = 1, ..., n$), $A = \bigcap_{r=1}^{n} A_r$ und $F = \bigcap_{r=1}^{n} F_r$:

(3.8) $\quad P_F(A) = \prod_{r=1}^{n} P^r_{F_r}(A_r)$

und

(3.9) $\quad U(F) = \sum_{r=1}^{n} U_r(F_r)$.

Satz 18: Seien umgekehrt für $r = 1, ..., n$ $\langle J_r, H_r, P^r, V_r, U_r\rangle$ E1-Modelle, wobei $J_1, ..., J_n$ paarweise disjunkt seien. Sei $I = \bigcup_{r=1}^{n} J_r$ und $I^h = \bigcup_{r=1}^{n} H_r$. Dann gibt es genau eine Wahrscheinlichkeitsfamilie P für $\langle I, I^h\rangle$, die (3.8) erfüllt. Sei weiter V auf $\mathcal{Z}(I)$ durch (3.7) (S. 133) und U auf $\mathrm{Z}(I^h)$ durch (3.9) definiert. Dann ist $\langle I, I^h, P, V, U\rangle$ ein E1-Modell und $\{J_1, ..., J_n\}$ eine $\perp_{\mathrm{cw}}$-Zerlegung von I und sogar die feinste, sofern alle $\langle J_r, H_r, P^r, V_r, U_r\rangle$ zusammenhängende E1-Modelle sind.

Zum Beweis der Sätze 17 und 18 ist folgendes zu bemerken:

(a) Daß in Satz 17 $P^r_{F_r}(A)$ wohldefiniert ist, liegt daran, daß $\{J_1, ..., J_n\}$ eine $\perp_{\mathrm{cw}}$-Zerlegung und daher auch eine $\perp$-Zerlegung von I ist, daß also $J_r \setminus I^h \perp I \setminus J_r$ und damit für alle $F, G \in \mathcal{Z}(I^h)$ mit $F, G \subseteq F_r$ und daher mit $F \approx_{H_r} G$ $P_F(A) = P_G(A)$ gilt. Die Gleichung (3.8) ergibt sich dann unmittelbar daraus, daß $\{J_1, ..., J_n\}$ eine $\perp$-Zerlegung von I ist.

(b) Die Behauptung in Satz 18, daß es zu den Wahrscheinlichkeitsfamilien P^r für $\langle J_r, H_r\rangle$ genau eine Wahrscheinlichkeitsfamilie P für $\langle I, I^h\rangle$, gibt, die (3.8) erfüllt, ist nicht trivial, aber in der Wahrscheinlichkeitstheorie unter der Überschrift „Existenz und Eindeutigkeit von Produktmaßen" hinlänglich bekannt.[28]

(c) So ist für die Sätze 17 und 18 eigentlich nur noch zu zeigen, daß (3.9) zutrifft: Seien also die Bezeichnungen wie in Satz 17. Dann gilt:

$$U(F) = \sum_{A \in \mathcal{Z}(I)} V(A) \cdot P_F(A) \qquad \text{(so ist } U \text{ festgelegt)}$$

$$= \sum_{A \in \mathcal{Z}(I)} \sum_{r=1}^{n} V_r(A_r) \cdot P_F(A) \qquad \text{(so waren } V_1,...,V_n \text{ angenommen)}$$

[28] S. etwa Bauer (1968), §§ 21-23+33.

$$
\begin{aligned}
&= \sum_{r=1}^{n} \sum_{A \in Z(I)} V_r(A_r) \cdot P_F(A) \\
&= \sum_{r=1}^{n} \sum_{A_r \in Z(J_r)} V_r(A_r) \cdot \left[\sum_{A' \in Z(I \setminus J_r)} P_F(A_r \cap A') \right] \\
&= \sum_{r=1}^{n} \sum_{A_r \in Z(J_r)} V_r(A_r) \cdot P_F(A_r) \\
&= \sum_{r=1}^{n} \sum_{A_r \in Z(J_r)} V_r(A_r) \cdot P_{F_r}^{r}(A_r) && \text{(gemäß der Festlegung der } F^r\text{)} \\
&= \sum_{r=1}^{n} U_r(H_r) && \text{(gemäß der Festlegung der } U_r\text{). Q.e.d.}
\end{aligned}
$$

Daß sich Entscheidungsmodelle in dieser Weise zerlegen und wieder zusammensetzen lassen, ist altbekannt und hier nur in unserem Rahmen reformuliert.[29] Dieser Sachverhalt ist natürlich insbesondere für die praktische Anwendung der Entscheidungstheorie bedeutsam; ein komplexes Entscheidungsproblem vereinfacht sich ja nicht wenig, wenn es sich in der in Satz 17 beschriebenen Weise zerlegen läßt.[30]

So weit die Antwort zur Anwendungsfrage. Wenn sie korrekt ist, so ist abermals klar, daß Savages „small worlds" nicht der Isolierung einzelner Entscheidungssituationen dienen; denn bei Savages Submodellen und den eben vorgestellten Zerlegungen handelt es sich offenkundig um zwei ganz verschiedene Dinge.

Wir stellten ja auch schon im Abschnitt 2.3 fest, daß Savage schließlich zeigt, wie sich Entscheidungsprobleme oder -situationen *verkleinern* lassen. Dies kann man auch mit den eben geschilderten Zerlegungen bewerkstelligen; doch bietet Savage eine darüber hinausgehende Verkleinerungsmethode, derer es nach wie vor bedarf. Denn auch die Spielraummengen, die in zusammenhängende Entscheidungsmodelle eingehen, können noch sehr umfangreich sein und insbesondere fast beliebig in die Zukunft ausufern. Jedenfalls dürfte es, kausale und wertmäßige Unabhängigkeit sehr strikt verstanden, ein Leichtes sein, Ketten von kausal oder wertmäßig zusammenhängenden Spielräumen zu konstruieren, die unseren üblichen zeitlichen Entscheidungshorizont bei weitem überschreiten.

---

[29] S. etwa Fishburn (l964), S. 406f.

[30] Einen ausführlichen informellen Überblick über dieses und weitere mehr oder weniger approximative Zerlegungsverfahren gibt Gäfgen (1974), S. 209-218.

Wenden wir uns also im nächsten Abschnitt dem Problem der Verkleinerung oder der – wie ich es nennen werde – Reduktion von Entscheidungssituationen zu. Abgesehen davon, daß dies in der entscheidungstheoretischen Praxis von Interesse ist, gelangen wir dadurch zu einem besseren Verständnis von Savages „small worlds“ und – dies sei jetzt schon angekündigt – auch zu weiteren Einblicken in die Beschaffenheit subjektiver Werte.

3.6 Reduktionen von Entscheidungsmodellen

Es geht uns jetzt also um die Frage, wie man selbst zusammenhängende Entscheidungsmodelle noch übersichtlicher gestalten kann. Anschauungsmaterial dazu lieferten wir schon im Abschnitt 2.3. Dennoch empfiehlt es sich, die abstrakten Erörterungen zur Verkleinerung von Entscheidungsmodellen mit einem weiteren Beispiel vorzubereiten.

Sei unser Entscheidender X diesmal ein Schüler, der mitten in der Reifeprüfung steht und sich darum bemüht, einen Studienplatz für Philosophie zu ergattern. Da dieses Fach außerordentlich überlaufen ist, herrscht in ihm ein strenger Numerus clausus. Nehmen wir an – damit probabilistische Überlegungen hereinspielen –, der Gesetzgeber habe die Zulassung zum Studium durch ein modifiziertes Losverfahren geregelt, das umso größere Chancen einräumt, je besser der Notendurchschnitt ist. X sei nun für Sonntagabend bei Freunden zu einer Party eingeladen, die ihm, wie die Erfahrung lehrt, für den nächsten Tag einen Kater beschert. Just an diesem Tag geht es für ihn aber in der mündlichen Prüfung in Erdkunde, seinem letzten Prüfungsteil, um die Note l oder 2, der, wie man weiß, für die Zulassung besonderes Gewicht zukommt. So steht er also vor der schweren Entscheidung, ob er sich einen schönen Abend machen oder besser noch einmal den Prüfungsstoff durcharbeiten soll.

Beschreiben wir diese Situation mit Hilfe eines Entscheidungsmodells. Da sind zunächst drei Spielräume auszumachen: der Handlungsspielraum i mit den Zuständen F_1 = „X geht zur Party“ und F_2 = „X büffelt Erdkunde“ und die Geschehensspielräume j, dessen Zustände A_r = „X erzielt die Erdkundenote r“ (r =1, 2) sind, und k, dessen Zustände B_1 = „X darf Philosophie studieren“ und B_2 = „X darf nicht Philosophie studieren“ sind. X glaube, daß sich der Spielraum k vor dem Spielraum i abschirmen lasse; d.h. für seine Wahrscheinlichkeitsfamilie P gelte $k \perp i \,/\, j$; ferner gelte $P_{F_1}(A_1) = 0{,}5$, $P_{F_2}(A_1) = 0{,}8$, $P_{F_r}(B_1 \mid A_1) = 0{,}9$ und $P_{F_r}(B_1 \mid A_2) = 0{,}7$ (r = 1, 2), wodurch Xs Wahrscheinlichkeitsfamilie P für $\langle\{i,j,k\}, \{i\}\rangle$ eindeutig bestimmt ist.

Seien außerdem für X alle drei Spielräume wertmäßig voneinander unabhängig. Xs subjektive Wertfunktion V läßt sich demnach in drei Funktionen V_i auf $Z(i)$, V_j auf $Z(j)$ und V_k auf $Z(k)$ (im Sinne von Definition 23, S. 133) additiv zerlegen; sei speziell $V_i(F_1) = 10$, $V_i(F_2) = -10$, $V_j(A_1) = V_j(A_2) = 0$, $V_k(B_1) = 200$ und $V_k(B_2) = -100$. Von diesen Zahlen dürfte allenfalls die Wahl von $V_j(A_1) = V_j(A_2) = 0$ unplausibel erscheinen. Darin soll jedoch nur zum Ausdruck kommen, daß *X* an der Erdkundenote als solcher gar nichts liegt; dies schließt ja nicht aus, daß sie für ihn im Hinblick auf die Zulassung sehr wichtig ist. Damit haben wir also ein E1-Modell konstruiert, wobei der Vollständigkeit halber angemerkt sei, daß $U(F_1) = \sum_{p,q=1}^{2} V(F_1 \cap A_p \cap B_q) \cdot P_{F_1}(A_p \cap B_q) = 150$ und entsprechend $U(F_2) = 148$ ist.

So dargestellt, steht *X* vor einem recht überschaubaren Entscheidungsproblem, das wir nun aber noch durchsichtiger gestalten wollen, indem wir es um den Spielraum *k* reduzieren. In das reduzierte Entscheidungsmodell gehen dann nur noch die Spielräume *i* und *j* ein. Die neue Wahrscheinlichkeitsfamilie P' ist einfach die Einschränkung der alten auf die Spielraummenge $\{i, j\}$; es gilt also $P'_{F_1}(A_1) = 0{,}5$ und $P'_{F_2}(A_1) = 0{,}8$, wodurch P' eindeutig bestimmt ist. Eine wesentliche Veränderung stellt sich jedoch bei den subjektiven Werten ein. Daß für *X* sehr wichtig ist, wie sich der Spielraum *k* realisiert, darf ja nicht unter den Tisch fallen; dies muß vielmehr in die neuen subjektiven Werte einfließen. In diesem Fall dürfte es das Natürlichste sein, die mit der Zulassung verknüpften subjektiven Werte allein auf die Erdkundenote zu übertragen; denn sie allein ist für die Chancen im Zulassungsverfahren direkt verantwortlich – jedenfalls gemäß der Abschirmbarkeitsrelation von *X*. Zum alten subjektiven Wert der Zustände A_1 und A_2 kommt also noch hinzu, daß sie die mehr oder weniger erwünschten Ergebnisse des Zulassungsverfahrens mehr oder weniger wahrscheinlich machen. Die neuen subjektiven Werte von A_1 und A_2 errechnen sich also gemäß der Formel

$$V'_j(A_p) = V_j(A_p) + \sum_{q=1}^{2} V_k(B_q) \cdot P_{F_r}(B_q \mid A_p) \quad (p = 1, 2)$$ [31]

d.h. es ist $V'_j(A_1) = 170$ und $V'_j(A_2) = 110$. Die (absoluten) subjektiven Werte der Handlungen werden dagegen in diesem Fall von der Reduktion nicht betroffen; daher gilt unverändert $V'_i(F_r) = V_i(F_r)$ $(r = 1,2)$. V'_i und V'_j summieren sich schließlich zur neuen subjektiven Wertfunktion V' für $\{i,j\}$. Damit haben wir ein

[31] Aufgrund unserer Annahmen über *P* ist es egal, ob wir hierin $r = 1$ oder $r = 2$ einsetzen.

zweites E1-Modell beschrieben, das als Reduktion des ersten um den Spielraum k noch um ein Stückchen einfacher ist als das erste.

In der Tat kann man die Situation von X durch das eine wie das andere Modell beschreiben. Daß der Spielraum k im zweiten Modell nicht ausdrücklich auftaucht, wird dadurch kompensiert, daß die subjektive Wertfunktion des reduzierten Modells nicht, wie das im ersten Modell geschah, den Wert der Erdkundenote als solcher, sondern den Wert der Erdkundenote unter Berücksichtigung aller ihrer Auswirkungen – hier auf die Zulassung – ausdrückt. Dies zeigt sich auch daran, daß der erwartete subjektive Wert der möglichen Handlungen in beiden Modellen gleich ist; wie leicht nachzuprüfen ist, gilt $U'(F_1) = 150$ und $U'(F_2) = 148$. Von daher dürfte es berechtigt sein, das reduzierte Modell als eine getreue Verkleinerung des ersten Modells zu bezeichnen.

Versuchen wir nun, den Mechanismus solcher Reduktionen, wie er sich an diesem Beispiel zeigt, allgemein zu beschreiben. Grob gesagt, dreht es sich bei einer Reduktion eines Entscheidungsmodells einfach darum, die Zahl der in es eingehenden Faktoren zu verringern und es dadurch kleiner und überschaubarer zu gestalten. Dies geschieht dadurch, daß zunächst gewisse Spielräume kurzerhand aus dem Entscheidungsmodell herausgestrichen werden. Damit daraus eine angemessene Verkleinerung resultiert, müssen dabei die mit den herausgestrichenen Spielräumen verknüpften subjektiven Werte auf einige, nicht unbedingt alle der übriggebliebenen Spielräume übertragen werden – und zwar in Form erwarteter subjektiver Werte. Das heißt genauer, daß die Zustände der fraglichen Spielräume neben ihrem alten subjektiven Wert zusätzlich mit ihrem erwarteten subjektiven Wert, der ihnen im Hinblick auf die herausgestrichenen Spielräume zukommt, belastet werden. So geschah das ja auch im obigen Beispiel. Natürlich ist weder willkürlich, welche Spielräume eliminiert werden, noch welche der verbliebenen Spielräume wertmäßig neu veranschlagt werden. In unserem Bemühen, solche Reduktionen formal möglichst allgemein zu charakterisieren, werden wir jedoch darauf zunächst keine Rücksicht nehmen. Später soll aber noch zur Sprache kommen, welche der formalen Reduktionsmöglichkeiten als natürlich bezeichnet werden können.

Zweierlei freilich muß jede Reduktion erfüllen: Erstens dürfen sich durch eine Reduktion die erwarteten subjektiven Werte der Handlungen oder, besser, der Handlungsverläufe nicht verändern. Und zweitens sollte die Reduktionsbeziehung transitiv sein; d.h. wenn ein Entscheidungsmodell um gewisse Spielräume reduziert wird und das Ergebnis davon um weitere Spielräume, so sollte dabei dasselbe herauskommen wie bei der Reduktion des ursprünglichen Modells um alle fraglichen Spielräume auf einmal. Schließlich dürfte auch klar sein, daß sich Handlungsspielräume auf diese Weise nicht wegreduzieren lassen. Denn da Handlungen keine Wahrscheinlichkeiten haben, gibt es keine erwarteten subjektiven Werte in Bezug auf Handlungsspielräume und damit keinen Spielraum, auf den sich die mit ihnen verknüpften subjektiven Werte in der gewünschten Weise übertragen ließen. Außerdem hätte es bei einer Reduktion um Handlungsspielräume keinen Sinn mehr zu verlangen, daß der erwartete subjektive Wert von Handlungsverläufen durch sie nicht angetastet werden dürfe.

In der folgenden Definition ist die Reduktion von E1-Modellen präzise und ganz allgemein beschrieben. Wir werden dabei zunächst nicht berücksichtigen, wie wir es bis jetzt taten, daß sich subjektive Wertfunktionen additiv zerlegen lassen und deswegen auch Teilzuständen subjektiver Wert zukommt; wir kümmern uns also vorläufig nur um die subjektiven Werte der Gesamtzustände.

Definition 26: Sei $\langle I, I^h, P, V, U\rangle$ ein E1-Modell und $K \subseteq I^g$ $(= I \setminus I^h)$. Dann heißt $\langle I', I^h, P', V', U'\rangle$ ein *K-Redukt* von $\langle I, I^h, P, V, U\rangle$ genau dann, wenn gilt:

(1) $I' = I \setminus K$,

(2) P' ist die Einschränkung von P auf $\langle I', I^h\rangle$,[32]

(3) V' ist eine subjektive Wertfunktion für I', für die gilt: für alle $F \in \mathcal{H}$ und $A \in \mathcal{Z}(I' \setminus I^h)$ mit $P_F(A) \neq 0$ ist

$$V'(A \cap F) = \sum_{B \in \mathcal{Z}(K)} V(A \cap B \cap P) \cdot P_F(B \mid A),$$

(4) U' ist die Funktion auf $\mathcal{H}$, für die für alle $F \in \mathcal{H}$

$$U'(F) = \sum_{A \in \mathcal{Z}(I')} V'(A) \cdot P'_F(A).$$

Es ist klar, daß ein K-Redukt eines E1-Modells wiederum ein E1-Modell ist. Jedoch ist darauf hinzuweisen, daß es in der Regel mehrere K-Redukte eines E1-Modells gibt. Dies ist aber nicht schlimm. Denn nennen wir zwei E1-Modelle $\langle I_1, I_1^h, P^1, V_1, U_1\rangle$ und $\langle I_2, I_2^h, P^2, V_2, U_2\rangle$ *äquivalent*, wenn $I_1 = I_2$, $I_1^h = I_2^h$, $P^1 = P^2$

---

[32] D.h. P' ist diejenige Wahrscheinlichkeitsfamilie für $\langle I', I^h\rangle$, für die für jedes $F \in \mathcal{H}$ P'_F die Einschränkung von P_F auf $\mathcal{A}(I')$ ist.

und für $A \in \mathcal{Z}(I_1^g)$ und $F \in \mathcal{Z}(I_1^h)$ $V_1(A \cap F) \neq V_2(A \cap F)$ allenfalls, wenn $P_F^1(A) = 0$. (Natürlich gilt dann auch $U_1 = U_2$.) Äquivalente E1-Modelle unterscheiden sich also allenfalls in ihren subjektiven Werten für Gesamtzustände, deren Wahrscheinlichkeit 0 ist und auf die es daher ohnehin nicht ankommt. Es ist dann klar, daß alle K-Redukte eines E1-Modells miteinander äquivalent sind, und in diesen Grenzen ist die Existenz mehrerer K-Redukte vollkommen tragbar. Wesentlicheres vermeldet der folgende

Satz 19:

(a) Ist $\langle I', I^h, P', V', U'\rangle$ ein K-Redukt von $\langle I, I^h, P, V, U\rangle$, so gilt für alle $F \in \mathcal{H}$ $U'(F) = U(F)$.

(b) Jedes E1-Modell ist ein $\emptyset$-Redukt von sich selbst.

(c) Ist Δ_2 ein K-Redukt von Δ_1 und Δ_3 ein L-Redukt von Δ_2, so ist Δ_3 ein $K \cup L$-Redukt von Δ_1.

(d) Ist P^* die Wahrscheinlichkeitsfamilie für $\langle I^h, I^h\rangle$[33], so ist $\langle I^h, I^h, P^*, U, U\rangle$ ein und sogar das einzige I^g-Redukt von $\langle I, I^h, P, V, U\rangle$.

Beweis: (b) und (d) sind trivial. (a) ergibt sich ebenso einfach aus (c) und (d). Bleibt also noch (c): Sei für $r = 1, 2, 3$ $\Delta_r = \langle I_r, I_r^h, P^r, V_r, U_r\rangle$, wobei $I_2 = I_1 \setminus K$ und $I_3 = I_2 \setminus L$. Es gilt dann für alle $F \in \mathcal{H}$, $A \in \mathcal{Z}(I_3 \setminus I^h)$ und $B \in \mathcal{Z}(L)$ mit $P_F^1(A) \neq 0$ und $P_F^1(B \mid A) = 0$

$$V_2(A \cap B \cap F) = \sum_{C \in \mathcal{Z}(K)} V_1(A \cap B \cap C \cap F) \cdot P_F^1(C \mid A \cap B) \quad \text{und}$$

$$V_3(A \cap F) = \sum_{B \in \mathcal{Z}(L)} V_2(A \cap B \cap C \cap F) \cdot P_F^1(B \mid A) \quad \text{und daher}$$

$$\begin{aligned} V_3(A \cap F) &= \sum_{B \in \mathcal{Z}(L)} \sum_{C \in \mathcal{Z}(K)} V_1(A \cap B \cap C \cap F) \cdot P_F^1(C \mid A \cap B) \cdot P_F^1(B \mid A) \\ &= \sum_{B \in \mathcal{Z}(L)} \sum_{C \in \mathcal{Z}(K)} V_1(A \cap B \cap C \cap F) \cdot P_F^1(B \cap C \mid A) \\ &= \sum_{D \in \mathcal{Z}(K \cup L)} V_1(A \cap B \cap D \cap F) \cdot P_F^1(D \mid A), \text{ q.e.d.} \end{aligned}$$

Satz 19, vor allem (a) und (c), garantiert also, daß der in Definition 26 definierte Reduktionsbegriff die gewünschten Eigenschaften hat. Natürlich drängt

[33] Da gibt es ja nur eine, für die lediglich die Trivialität gilt, daß für alle $F, G \in \mathcal{H}$ $P^*_F(G) = \begin{Bmatrix} 1, \text{ wenn } F = G \\ 0, \text{ wenn } F \neq G \end{Bmatrix}$.

sich hier ein Vergleich mit Savages „small worlds" unmittelbar auf: In Savage-Submodellen haben wir es mit gröberen Umständen und verkürzten Konsequenzen zu tun, bei K-Redukten mit kleineren Entscheidungsfeldern. Satz 19(b) korrespondiert mit Satz 1, Satz 19(c) mit Satz 2 und Satz 19(d) mit Satz 3. Was Savage zusätzlich voraussetzen mußte, ist hier freilich automatisch erfüllt: So gilt die Transitivität des Reduktionsbegriffs uneingeschränkt, und dementsprechend gilt hier, wie Satz 19(a) zeigt, unmittelbar, was Savage für echte gegenüber potentiellen Savage-Submodellen extra fordern mußte.

Überlegen wir uns nun, wie sich in Reduktionen allgemein einarbeiten läßt, was wir im einleitenden Abiturientenbeispiel bereits demonstrierten: nämlich daß sich aufgrund der additiven Zerlegbarkeit subjektiver Wertfunktionen die subjektiven Werte der Zustände der herausgestrichenen Spielräume möglicherweise sinnvoll auf bestimmte Teilzustände und nicht bloß pauschal auf die Gesamtzustände des reduzierten Modells übertragen lassen. Formal ausgedrückt, geht es also in Bezug auf die Reduktion eines E1-Modelles $\{I, I^h, P, V, U\}$ um die Spielraummenge $K \subseteq I^g$ um die Frage, bezüglich welcher $\perp_w$-Zerlegungen $\{J_1, ..., J_n\}$ von I sich in additiven Zerlegungen $\{V_1, ..., V_n\}$ von V jedes V_r *für sich* mittels Erwartungswertbildung bzgl. $\mathcal{Z}(J_r \cap K)$ auf eine subjektive Wertfunktion V_r' für $J_r \setminus K$ reduzieren läßt. Eine Antwort darauf gibt der

SATZ 20: Sei $\langle I, I^h, P, V, U\rangle$ ein E1-Modell, $K \subseteq I^g$, $I' = I \setminus K$ und P' die Einschränkung von P auf $\langle I', I^h\rangle$. Sei weiter $\{V_1, ..., V_n\}$ eine additive Zerlegung von V bzgl. der $\perp_w$-Zerlegung $\{J_1, ..., J_n\}$ von I, $K_r = J_r \cap K$ und $J_r' = J_r \setminus K$ ($r = 1, ..., n$). Wir können dabei ohne Beschränkung der Allgemeinheit annehmen, daß für ein $m \leq n$ $J_r' \neq \emptyset$ für $r = 1, ..., m$ und $J_r' = \emptyset$, d.h. $J_r = K_r$ für $r = m$+1, ..., n. Gelte nun für alle $r = 1, ..., n$:

$$K_r \perp I' \setminus J_r' \,/\, J_r' \setminus I^h. \tag{3.10}$$

Dann gibt es erstens für jedes $r = 1, ..., m$ eine subjektive Wertfunktion V_r' für J_r', für die gilt: für alle $A \in \mathcal{Z}(J_r' \setminus I^h)$, $F_r \in \mathcal{Z}(J_r' \cap I^h)$ und $F \in \mathcal{Z}(I^h)$ mit $F \subseteq F_r$ und $P_F(A) \neq 0$ ist

$$V_r'(A \cap F_r) = \sum_{B \in \mathcal{Z}(K_r)} V_r(A \cap B \cap F_r) \cdot P_F(B \mid A).$$

Und zweitens gibt es für jedes $r = m$+1, ..., n eine reelle Zahl z_r, so daß für alle $F \in \mathcal{Z}(I^h)$

$$z_r = \sum_{B \in Z(K_r)} V_r(B) \cdot P_F(B).$$

Sei zuletzt für alle $A_r \in Z(J'_r)$ $(r = 1, ..., m)$

$$V'(\bigcap_{r=1}^{m} A_r) = \sum_{r=1}^{m} V'_r(A_r) + \sum_{r=m+1}^{n} z_r$$

und U' wie in Definition 26, (4), (S. 144). Dann ist $\langle I', I^h, P', V', U' \rangle$ ein K-Redukt von $\langle I, I^h, P, V, U \rangle$.

Zum Beweis von Satz 20 ist folgendes zu bemerken: (a) Daß es solche Funktionen V'_r $(r = 1, ..., m)$ gibt, liegt daran, daß für alle $A \in Z(J'_r \setminus I^h)$, $B \in Z(K_r)$, $F_r \in Z(J'_r \cap I^h)$ und $F, G \in Z(I^h)$ mit $F, G \subseteq F_r$ und $P_F(A) \neq 0 \neq P_G(A)$ wegen (3.10) eben gilt, daß $P_F(B \mid A) = P_G(B \mid A)$. Und aus demselben Grunde gibt es auch die Zahlen z_r $(r = m+1, ..., n)$. In der Tat hätte für beides statt (3.10) auch die Annahme genügt, daß $K_r \perp I^h \setminus J'_r \,/\, J'_r \setminus I^h$ für alle $r = 1, ..., n$. Die volle Stärke von (3.10) brauchen wir jedoch im nächsten Beweisteil. (b) Es ist jetzt noch zu zeigen, daß V' Definition 26, (3), (S. 144), erfüllt: Sei $A_r \in Z(J'_r \setminus I^h)$, $F_r \in Z(J'_r \cap I^h)$ $(r = 1, ..., m)$, $A = \bigcap_{r=1}^{m} A_r$ und $F = \bigcap_{r=1}^{m} F_r$ und sei für $r = m+1, ..., n$ $A_r = F_r = W$; das folgende läßt sich dann einheitlicher schreiben. Dann gilt:

$$\begin{aligned}
V'(A \cap F) &= \sum_{r=1}^{m} V'_r(A_r \cap F_r) + \sum_{r=m+1}^{n} z_r \\
&= \sum_{r=1}^{m} \sum_{B_r \in Z(K_r)} V_r(A_r \cap B_r \cap F_r) \cdot P_F(B_r \mid A_r) + \sum_{r=m+1}^{n} \sum_{B_r \in Z(K_r)} V_r(B_r) \cdot P_F(B_r) \\
&= \sum_{r=1}^{n} \sum_{B_r \in Z(K_r)} V_r(A_r \cap B_r \cap F_r) \cdot P_F(B_r \mid A_r) \\
&= \sum_{r=1}^{n} \sum_{B_r \in Z(K_r)} V_r(A_r \cap B_r \cap F_r) \cdot P_F(B_r \mid A) && \text{(aufgrund (3.10))} \\
&= \sum_{r=1}^{n} \sum_{B_r \in Z(K_r)} V_r(A_r \cap B_r \cap F_r) \cdot \sum_{B' \in Z(K \setminus K_r)} P_F(B_r \cap B' \mid A) \\
&= \sum_{r=1}^{n} \sum_{B \in Z(K)} V_r(A_r \cap B_r \cap F_r) \cdot P_F(B \mid A) && \text{(wobei } B_r \text{ jetzt das Element} \\
& && \text{von } Z(K_r) \text{ mit } B \subseteq B_r \text{) sei)} \\
&= \sum_{B \in Z(K)} \sum_{r=1}^{n} V_r(A_r \cap B_r \cap F_r) \cdot P_F(B \mid A) \\
&= \sum_{B \in Z(K)} V(A \cap B \cap F) \cdot P_F(B \mid A) && \text{q.e.d.}
\end{aligned}$$

Es sind also jedenfalls die $\perp_w$-Zerlegungen, die (3.10) erfüllen, bezüglich derer sich die subjektiven Werte in reduzierten Modellen sozusagen komponentenweise neu veranschlagen lassen. Im übrigen läßt sich mit Satz 8 (S. 107) leicht zeigen, daß mit zwei Zerlegungen auch ihr Produkt (3.10) erfüllt; es gibt also eine feinste Zerlegung, die (3.10) genügt. Darüber, wie fein diese tatsächlich ist, läßt sich freilich allgemein nichts sagen.

Damit sind wir auch hinsichtlich der Frage, welche Sorten von Reduktionen man als besonders natürlich auszeichnen könnte, ein Stückchen weiter gekommen. Denn sicherlich entspricht eine Reduktion, die sich unter Ausnutzung wertmäßiger Unabhängigkeit komponentenweise vollzieht, eher dem intuitiven Vorgehen als eine pauschale Reduktion, wie sie in Definition 26 (S. 144) beschrieben wurde. Dabei funktioniert die in Satz 20 beschriebene Reduktion noch völlig allgemein, d.h. für beliebige Geschehensspielraummengen K; die Kehrseite davon ist allerdings, daß die Bedingung (3.10) noch recht unschön ist und außerdem unklar bleibt, wie feine $\perp_w$-Zerlegungen (3.10) erfüllen.

Dies ändert sich aber, wenn wir bedenken, daß die Natürlichkeit einer Reduktion vor allem davon abhängt, um welche Spielräume reduziert wird. Hätten wir z.B. das eingangs geschilderte Entscheidungsproblem des Abiturienten um den Spielraum der möglichen Erdkundenoten reduziert, so wäre dies wohl eher auf Befremden gestoßen. Der allgemeine Grund dafür dürfte sein, daß es besonders natürlich zu sein scheint, ein Entscheidungsproblem jeweils um seine am weitesten in der Zukunft gelegenen Geschehensspielräume zu reduzieren. Genauer: Wenn wir, Zeitliches berücksichtigend, von einem temporalen Entscheidungsfeld $\langle I, I^h, T, \leq, s\rangle$ ausgehen, so läuft die jetzt anvisierte Methode darauf hinaus, $\langle I, I^h, P, V, U\rangle$ zunächst um $I^g_{>t_1}$ für ein $t_1 \in T$ zu reduzieren. Das Ergebnis davon können wir, wenn wir wollen, in einem zweiten Schritt noch um $I^g_{>t_2}$ für ein $t_2 < t_1$ reduzieren etc., bis schließlich ganz I^g wegreduziert ist. Es dreht sich hier also, kurz gesagt, um eine sukzessive Reduktion der Zukunft. So hielten wir es ja auch in unserem Einführungsbeispiel.

Wieso liegt dieses Reduktionsverfahren intuitiv besonders nahe? Nicht deswegen, weil man so besonders feine $\perp_w$-Zerlegungen findet, die (3.10) erfüllen; in dieser Hinsicht könnten es allenfalls empirische Eigenheiten wertmäßiger Unabhängigkeit besser stellen als andere Reduktionen. Sondern vielmehr deswegen, weil bei ihm besonders leicht nachzuprüfen ist, ob die probabilistische Voraussetzung (3.10) für eine komponentenweise durchgeführte Reduktion erfüllt ist.

Das liegt daran, daß es *nur* bei diesem Vorgehen möglich ist, diese Voraussetzungen mit Hilfe kausaler Abschirmbarkeit auszudrücken. Der eine Teil dieser Behauptung, nämlich daß ohne dieses Vorgehen eine solche Formulierung nicht möglich ist, dürfte einleuchten. Denn wenn nicht jeweils die entfernteste Zukunft eines Entscheidungsmodells wegreduziert wird, so beinhaltet (3.10) unter anderem, daß gewisse eliminierte Spielräume von relativ zu ihnen zukünftigen Spielräumen probabilistisch unabhängig zu sein haben, und so etwas läßt sich offensichtlich nicht mit einer Abschirnbarkeitsrelation im Sinne von Definition 19 (S. 123) formulieren. Den anderen Teil der Behauptung liefert der folgende

Satz 21: Sei $\langle I, I^h, T, \leq, s, P, V, U \rangle$ ein temporales E1-Modell für irgendeinen Zeitindex und $\{J_1, ..., J_n\}$ eine Zerlegung von I. Sei $K = I^g_{>t_0}$ für ein $t_0 \in T$, $I' = I \setminus K$, $K_r = J_r \cap K$ und $J'_r = J_r \setminus K$ ($r = 1, ..., n$). Dann gilt für alle r = 1, ..., n:

$$\text{wenn } K_r \perp_a I \setminus J_r, \text{ dann } K_r \perp I' \setminus J'_r \setminus I^h.$$

Beweis: Sei $\{t \in T \mid t > t_0\} = \{t_1, ..., t_m\}$, wobei $t_p < t_q$ für $p < q$. Nach Definition 19 (S. 123) besagt unsere Prämisse, daß

$$(K_r)_{=t_p} \perp (I \setminus J_r)_{\leq t_p} / (J_r)_{<t_p} \setminus I^h \qquad \text{für } p = 1, ..., m.$$

Daraus folgt mit Satz 8(c) (S. 107)

$$(3.11) \qquad (K_r)_{=t_p} \perp I' \setminus J'_r / (J'_r)_{<t_p} \setminus I^h \qquad \text{für } p = 1, ..., m.$$

Und daraus ergibt sich

$$(3.12) \qquad (K_r)_{\leq t_{p'}} \perp I' \setminus J'_r / J'_r \setminus I^h \qquad \text{für alle } p' = 1, ..., m.$$

auf die folgende Weise: Für $p' = 1$ ist (3.12) identisch mit (3.11) für $p = 1$; hat man nun schon (3.12) für $p' = q$, so folgt daraus und aus (3.11) für $p = q + 1$ unter Anwendung von Satz 8(f) (S. 107) auch (3.12) für $p' = q + 1$. Für $p' = m$ besagt aber (3.12) gerade das, was wir beweisen wollen. q.e.d.

Die Voraussetzungen für eine komponentenweise durchgeführte Reduktion um die jeweils entfernteste Zukunft lassen sich also schon mit der einfachen kausalen Abschirmbarkeit und nicht erst mit der komplizierten bedingten probabilistischen Unabhängigkeit formulieren. Und bei einer solchen Reduktion werden die subjektiven Werte der Zustände der eliminierten Spielräume auf die Zustände der für diese Spielräume am direktesten kausal relevanten Spielräume übertragen.

Wir wollen uns nun nicht weiter in dieses Thema vertiefen, obwohl es sicherlich interessant wäre, sich noch einen systematischeren Überblick über die verschiedenen Reduktionsmöglichkeiten zu verschaffen, besonders elegant zu bewältigende Spezialfälle zu diskutieren etc. Doch ist eine Vertiefung auch gar nicht nötig. Denn selbstverständlich ist die Reduktion von Entscheidungsmodellen ein Hauptanliegen jeder praxisorientierten Entscheidungstheorie und daher Thema intensiver Forschung. Darin geht es natürlich nicht nur um exakte Reduktionen, wie sie hier beschrieben wurden, sondern vor allem um noch weiter tragende, möglichst praktikable approximative Reduktionsverfahren.[34] Insbesondere befinden wir uns in enger Nachbarschaft zu den sogenannten Markoffschen Entscheidungsprozessen, einem ebenfalls gründlich untersuchten Gegenstand der Entscheidungstheorie.[35] Diese sind in einer Hinsicht etwas spezieller als unsere Reduktionen von Entscheidungsmodellen, da sie durch speziellere Unabhängigkeitsannahmen charakterisiert sind. Und in einer anderen Hinsicht sind sie allgemeiner; sie arbeiten von vornherein nicht bloß mit Handlungen, sondern mit Strategien, die wir bisher notorisch außer acht gelassen haben.

Aus all dieser Reduziererei ist jedoch, wie ich glaube, noch eine wichtige allgemeine Lehre bezüglich der Natur quantitativer Wünschensdispositionen zu ziehen. Sie zeigt uns nämlich, daß man gar nicht sinnvoll von *dem* subjektiven Wert eines bestimmten Zustandes reden kann. Damit spreche ich nicht den trivialen Punkt an, daß für subjektive Werte im allgemeinen nur eine Intervallskala besteht und daß man daher subjektive Werte nur bis auf positiv lineare Transformationen eindeutig bestimmen kann. Denn selbst wenn Nullpunkt und Einheit dieser Skala festgelegt wären, dürfte man nicht sagen, ein Zustand habe für den Entscheidenden den und den subjektiven Wert. Das liegt daran, daß man den Entscheidenden mit mehr oder weniger stark reduzierten Entscheidungsmodellen beschreiben kann, und in der Regel ist der subjektive Wert eines Zustandes, sofern er überhaupt einen subjektiven Wert erhält, in den verschiedenen Modellen

[34] Vgl. dazu die informelle Schilderung von Gäfgen (1974), S. 205-209, S. 351-363 und S. 469-475, der statt Reduktion auch von Kondensation von Entscheidungsfeldern spricht.

[35] S. etwa Fishburn (1964), S. 412ff., oder Mine, Osaki (1970).

verschieden. Der subjektive Wert eines Zustandes läßt sich also nur relativ zu dem jeweiligen Entscheidungsmodell, mit dem der Entscheidende gerade beschrieben wird, d.h. nur relativ zu der in das jeweilige Modell eingehenden Gesamtspielraummenge, angeben.

So befremdlich dies auf den ersten Blick sein mag, so gut stimmt dieser Sachverhalt mit unseren alltäglichen Erfahrungen und Redewendungen überein. „Eigentlich würde ich viel lieber zu Hause bleiben; aber man muß ja den Gästen wohl oder übel mal die Oper vorführen." „So ein saftiges Steak wäre schon etwas Feines; wenn's nur nicht gar so teuer wäre." „Eine interessante Paarung, das Fußballspiel heute nachmittag – aber bei dem Wetter ins Stadion gehen, nein – na ja, ist vielleicht immer noch besser als mit der Schwiegermutter am Kaffeetisch – und mit dem Ede zusammen erst recht, ich muß ihn doch gleich mal anrufen." Weitere Beispiele: die verhaßten Pflichten, die dann doch erledigt werden; der Verzicht auf durchsetzbare Wünsche um des lieben Friedens willen; die Befolgung von Konventionen, die einem ganz gegen den Strich gehen etc. In all diesen Fällen – und die Liste ließe sich ohne Mühe beliebig verlängern – werden intuitiv Reduktionen vollzogen, bei denen der subjektive Wert eines Zustandes immer wieder anders beurteilt wird.

Auch die philosophische Terminologie ist von diesen Dingen nicht unberührt geblieben. So waren ja die Begriffspaare „gut als Zweck – gut als Mittel" und „Wert an sich – Wert als Mittel" dazu gedacht, solche Situationen, allerdings nicht auf rein subjektiver, sondern auf verbindlicherer ethischer oder moralischer Ebene, in den Griff zu bekommen, wenngleich auch auf denkbar grobschlächtige Weise. Mit Entscheidungsmodellen, die mit komplexeren Kausal- und Wertbegriffen ausgestattet sind, und ihren Reduktionen lassen sich solche Situationen entschieden differenzierter und präziser erfassen.

In Anbetracht dieser alltäglichen und philosophischen Tatbestände hätten wir also gar nichts anderes erwarten dürfen, als daß der subjektive Wertbegriff stärker zu relativieren ist als nur bezüglich Person, Zeitpunkt und Nullpunkt und Einheit der Meßskala.

Man könnte denken, daß es hier einfach um die Unterscheidung zwischen absolutem und erwartetem subjektiven Wert geht. Doch wäre das eine Fehleinschätzung der Lage. Denn zum einen können die Zustände einer Spielraummenge J viele erwartete subjektive Werte haben: nämlich im Hinblick auf jeweils andere Spielraummengen, für deren Realisierung J verantwortlich ist. Und zum anderen – und dies ist entscheidend – lassen sich diese erwarteten subjektiven Werte mit vollem Recht auch als absolute subjektive Werte verstehen; sie fungieren ja im jeweiligen Modell als absolute, nicht-erwartete subjektive Werte. Es ist also schon der Begriff des absoluten subjektiven Wertes in der genannten Art und Weise zu relativieren – sofern man bei dieser Sachlage die Unterscheidung zwischen erwarteten und absoluten subjektiven Werten überhaupt aufrecht erhalten will. Besser wäre es freilich, einfach von subjektiven Werten relativ zu der und der Spielraummenge zu reden.

Damit klärt sich auch, inwiefern schon in die quantitativen Wünschensdispositionen selbst – und damit in Wünschensdispositionen allgemein, so weit sie sich als Vergröberungen der quantitativen verstehen lassen – Glaubensdispositionen eingehen – ein Punkt, den wir schon im ersten Kapitel (S. 24) als klärungsbedürftig empfanden. Denn (absolute) subjektive Werte von Zuständen in reduzierten Entscheidungsmodellen sind in der Regel erwartete subjektive Werte dieser Zustände im Hinblick auf die wegreduzierten Spielräume. Und so ist nichts Überraschendes daran, daß subjektive Werte, d.h. quantitative Wünschensdispositionen mit Glaubensdispositionen verquickt sind, und zwar nicht erst in ihren Manifestationen.

KAPITEL 4

Dynamisches und Strategisches

4.1 Der Grundgedanke

Bisher haben wir eine rein statische Entscheidungstheorie betrieben. Teils haben wir auf Zeitpunkte gar nicht explizit Bezug genommen, und wenn wir es taten, so haben wir nur Entscheidungsmodelle zu einem bestimmten Zeitpunkt betrachtet und sind auf zeitliche Änderungen der Situation und insbesondere der subjektiven Wahrscheinlichkeiten und Werte nicht eingegangen. Dabei ist selbst die statische Theorie noch unvollständig; damit meine ich nicht, daß es noch viele unerwähnte und unerforschte Dinge gibt – die gibt es immer –, sondern vielmehr, daß die mit E1-Modellen verknüpfte empirische Behauptung noch sehr vorläufig war. Sie sagte, daß eine Person X, die zu t durch das E1-Modell $\langle I, I^h, \leq, s, P, V, U\rangle$ für t charakterisiert wird, beginnt, einen Handlungsverlauf $F \in \mathcal{H}$ zu realisieren, für den $U(F)$ maximal ist. Und ihre Schwäche war, daß die zeitlich erste Handlung eines optimalen Handlungsverlaufs mitnichten die zeitlich erste Handlung einer optimalen Strategie zu sein braucht. Wir müßten uns also schon im Hinblick auf eine bloß statische Entscheidungstheorie mit Strategien beschäftigen. Dies taten wir aber bisher nicht mit der konstanten Entschuldigung, Strategien erforderten eine dynamische Theorie. Darin liegt nichts Paradoxes; es ist einfach so, daß wir erst die noch unfertige statische Theorie dynamisieren müssen, bevor wir die statische Theorie vervollständigen können. Dies hat den folgenden Grund:

Unsere statische Theorie ist zwar noch unfertig, aber für gewisse Individuen ist sie doch schon angemessen – soweit die hier diskutierte Standard-Entscheidungstheorie überhaupt angemessen ist –, nämlich gerade für solche Individuen, die nicht in Rechnung stellen, daß sich ihre Entscheidungssituation und insbesondere ihre subjektiven Wahrscheinlichkeiten und Werte ändern können und werden. Ein solches Individuum braucht nicht strategisch zu denken; denn Strategien sagen ja gerade, wie man sich auf veränderte Gegebenheiten einstellen soll, und daher kann ein Individuum, das die Möglichkeit leugnet oder gar nicht erst sieht, daß sich an seiner Situation etwas ändert, Strategien getrost außer acht lassen und direkt auf optimale Handlungsverläufe ausgehen. Für solche beschränkte Individuen haben wir also eine statische Theorie; im Abschnitt 4.2 wollen wir einiges Dynamische über sie sagen.

Erst danach, das dürfte nun klar sein, können wir uns den umsichtigeren Individuen zuwenden, die ihre eigene Dynamik berücksichtigen, die also die verschiedenen Entscheidungssituationen, in die sie geraten können, voraussehen und sich Strategien zurechtlegen, was in welcher Situation zu tun sei. Das soll in den Abschnitten 4.3 und 4.4 geschehen; sie sollen die bisherige statische Theorie so ergänzen, daß sie auch für solche umsichtigen Individuen angemessen ist. Dynamische Aussagen für diese Individuen, die über die für beschränkte Individuen hinausgehen, werden sich nicht ergeben.

Doch machen wir uns, bevor wir in die Details einsteigen, erst ganz klar, was es mit dieser Art von Umsichtigkeit auf sich hat: Ein damit ausgestattetes Individuum zieht, so sagten wir, in Betracht (genauer: hält es unter den und den Handlungen für so und so wahrscheinlich), daß es in eine bestimmte zukünftige Entscheidungssituation gerät. Es überlegt sich dann bereits zum früheren Zeitpunkt, was es in dieser Situation am ehesten tun sollte und also auch tatsächlich täte; und dies macht es für alle Entscheidungssituationen, in die es glaubt geraten zu können (und die irgendetwas mit seiner jetzigen Situation zu tun haben). Erst aufgrund dessen wird es entscheiden, welches weitere Vorgehen am besten ist. Ein solches Individuum fängt also, mit anderen Worten, an, über seine eigenen zukünftigen Handlungen zu theoretisieren, und die Theorie, die es dabei hat, ist just die Entscheidungstheorie, d.h. es nimmt an, daß es mit seinen zukünftigen Handlungen den erwarteten subjektiven Wert maximieren wird.

Dies hört sich so an, als ob ein solches Individuum sich über all das im klaren sei und seine Situation ausdrücklich in dieser Weise analysiere. Es läßt sich auch schlecht anders formulieren; aber natürlich muß es keineswegs so sein, solche Umsichtigkeit verlangt keinerlei entscheidungstheoretische Ausbildung. Zumindest in einfachen Fällen wird ein Individuum, ohne jemals etwas von Entscheidungstheorie gehört zu haben, allein aufgrund eines groben Überblicks eine Handlung ausführen, die sich auch nach einer expliziten, umsichtigen Entscheidungsfindung als optimal erwiese. Häufig wird ein optimales Verhalten schon so gut gelernt oder gewohnheitsmäßig verfestigt sein, daß bei ihm überhaupt nichts überlegt wird. Dies zeigt sich z.B. an der allgemein geübten Praxis, kostenlose, relevante Informationen abzuwarten, sofern man es sich leisten kann, noch zu warten. Erst in sehr komplexen Situationen wird es nötig sein, das erwähnte Theoretisieren ausdrücklich zu vollziehen. Aber so ist es ja überall; für komplizierte Rechnungen bzw. logische Schlüsse sollte man sich auch etwas in der Arithmetik bzw. in der Logik auskennen.

Doch ist diese Frage jetzt nicht wichtig. Wichtig ist für uns vielmehr, wie wir solcherart umsichtige Individuen entscheidungstheoretisch erfassen sollen, und dazu müssen wir sie eben als Individuen beschreiben, die selbst ihre eigenen zukünftigen Handlungen stillschweigend oder ausdrücklich entscheidungstheoretisch bestimmt sehen. Genau diese Selbstanwendung der Entscheidungstheorie steckt hinter strategischem Denken. In den Abschnitten 4.3 und 4.4 wird dies noch ganz deutlich werden.

Dieser Gedankengang fordert drei Fragen heraus, zu denen ich aber lediglich einige spekulative Bemerkungen machen will. Erstens drängt sich natürlich die Frage auf, ob das so skizzierte Vorgehen auch Personen gerecht werden kann, die ausdrücklich eine völlig andere und womöglich mit der Entscheidungstheorie unverträgliche Theorie von sich und ihren Handlungen haben. Es ist mir nicht klar, was da zu tun ist; deshalb will ich diese gewiß interessante Frage einfach im Raum stehen lassen. Immerhin ist zu vermuten, daß dann auch unsere Beschreibung solcher Personen entsprechend zu ändern ist.

Zweitens kann, wer will, auch von hier aus in anthropologische Tiefen vordringen. Jedenfalls liegt die ebenfalls interessante Frage nahe, welche Individuen zu dieser Art von Umsichtigkeit fähig sind. Der Mensch ist es sicherlich (in unterschiedlichem Maße). Ob auch Primaten oder niederere Tierarten wenigstens in primitiven Fällen dazu fähig sind, entzieht sich dagegen meiner Kenntnis. Allerdings könnte man dafür, daß wir einem Lebewesen überhaupt diese Fähigkeit zubilligen, für wesentlich halten, daß es sich auch sprachlich zu erläutern vermag.

Schließlich sind hier noch einige Anmerkungen zum Verhältnis zwischen Entscheidungs- und Spieltheorie am Platze. Dieses wird in der Regel so beschrieben, daß die Entscheidungstheorie schlicht der personale Spezialfall der allgemeineren interpersonalen Spieltheorie ist. Dies ist erstens richtig, zweitens häufig irreführend so formuliert, daß die Entscheidungstheorie Spiele gegen die Natur erfasse – eine Wendung, die die Experten selbst nicht mehr verwirrt, aber dauernd zu Vorkehrungen gegen falsche Assoziationen anderer nötigt –, und drittens vor allem nicht sonderlich lehrreich. Wesentlich erkenntnisträchtiger wäre es, wenn sich umgekehrt die Spieltheorie irgendwie auf die Entscheidungstheorie zurückführen ließe. In der Tat reicht auch die gewöhnliche Entscheidungstheorie zur Beschreibung der interpersonalen Situation völlig hin, wenn man es mit lauter Individuen zu tun hat, die in Bezug auf die jeweils anderen im obigen Sinne beschränkt sind; denn dann können wir jedes Individuum z.B. mit einem E1-Modell beschreiben, in dem die Handlungen der anderen einfach gewisse Geschehensspielräume bilden. Kompliziert wird es erst, wenn wir es mit Personen zu tun haben, die über die anderen Personen und ihre Handlungen Theorien entwickeln, womöglich Annahmen darüber bilden, welche Theorien die anderen entwickelt haben könnten, etc.; wenn wir also gerade die Sorte von Umsichtigkeit, die wir hier im rein personalen Bereich diskutieren wollen, auch auf der interpersonalen Ebene in den Griff bekommen müssen. Daß sich Menschen in dieser Weise gegenseitig auszurechnen versuchen, bildet ja auch den eigentlichen Gegenstand der Spieltheorie.[1] Daß dem so ist, zeigt sich merkwürdigerweise jedoch vor allem in den von Spieltheoretikern verwandten einführenden Beispielen[2]; in ihrer Theorie selbst scheint es sich nicht mehr explizit niederzuschlagen. Noch kurioser wird die Sachlage dadurch, daß der einzige Ort, wo die gegenseitigen Durchschauungsversuche der Menschen ausdrücklich, wenngleich in rein qualitativer Form und ohne exakt ausgeführte Theorie, bedacht werden, die Sprachphilosophie zu sein scheint.[3] Jedenfalls glaube ich, daß es der Spieltheorie wie der Sprachphilosophie gut tun könnte, wenn man versucht, der interpersonalen Situation entscheidungstheoretisch dadurch Herr zu werden, daß man die

---

[1] Und macht das Leben so elend (und) kompliziert.

[2] Vgl. etwa Luce, Raiffa (1957), Abschnitte 4.3 und 4.7.

[3] Das fing dort mit Grice (1957) an, der damit, daß ein Sprecher gewisse Wünsche hat und aufgrund gewisser Annahmen über seine Hörer durch eine bestimmte Äußerung. verwirklichen zu können glaubt, zu explizieren versuchte, daß der Sprecher mit seiner Äußerung etwas meint; setzte sich unter anderem mit Lewis (1969) und seinem Begriff des gemeinsamen Wissens fort; und mittlerweile sind verschachtelte Formulierungen der Form „*x* glaubt, daß *Y* glaubt, daß *X* glaubt, daß ...“ unter Sprachphilosophen gang und gäbe und nachgerade Gegenstand eines Denksports.

Theoretisiererei der einzelnen Personen über die anderen ausdrücklich und exakt (d.h. eigentlich: erst einmal übertrieben ausdrücklich und exakt) in den sie beschreibenden Entscheidungsmodellen nachvollzieht.

4.2 Zur Dynamik von Entscheidungsmodellen

Wenden wir uns also zunächst der Dynamik von E1-Modellen zu. Die aus T, $\leq$ und s bestehende zeitliche Struktur wollen wir dabei in diesem Abschnitt fest vorgeben; wir wollen daher auch darauf verzichten, sie in temporalen E1-Modellen explizit anzuführen. Die Frage, um die es geht, lautet dann: Werde der Entscheidende X zum Zeitpunkt t durch das temporale E1-Modell $\langle I, I^h, P, V, U\rangle$ für t charakterisiert. Was läßt sich dann über ein E1-Llodell $\langle I', I'^h, P', V', U'\rangle$ für $t' > t$ sagen, das X zum späteren Zeitpunkt t' charakterisiert?

Als erstes ist durchaus zu gewärtigen, daß I' andere Spielräume enthält als I. X kann ja plötzlich ganz neue Spielräume berücksichtigen und alte vergessen haben. Doch dürfte man darüber nur im Einzelfall Gescheites sagen können. Und selbst wenn hier eine allgemeine Aussage möglich wäre, so könnte man aus $\langle I, I^h, P, V, U\rangle$ nichts über Xs probabilistische und wertmäßige Beurteilung der zu t' neu hinzugekommenen Spielräume gewinnen. Wir wollen daher annehmen, daß $I' = I$, daß also Xs dynamischer Entwicklung ein fester Spielraumbestand unterliegt. Nur so können wir auf substantielle dynamische Aussagen hoffen. Klar ist, daß man dann auch annehmen muß, daß $I'^h = I^h_{>t'}$.

Über die Dynamik der subjektiven Wahrscheinlichkeiten, d.h. über das Verhältnis zwischen P und P', läßt sich dagegen entschieden mehr sagen. Der Zusammenhang, der da praktisch immer gesehen wird, kommt in der einen oder anderen Weise über das Verfahren der Konditionalisierung zustande, das wir schon auf S. 27 angedeutet haben und das sich dort für dynamische Theorien ganz allgemein als bedeutsam erwies.

Die älteste und bekannteste dynamische Aussage liefert die sogenannte einfache Konditionalisierung: Sie geht davon aus, daß X zwischen t und t' gewisse Erfahrungen sammelt, die sich durch eine Proposition aus $\mathcal{A}(I)$ ausdrücken lassen. Dabei kann man speziell annehmen, daß diese Proposition von der Form $B_0 \cap F_0$ ist, wobei $B_0 \in \mathcal{A}(I^g_{>t} \cap I^g_{\leq t'})$ und $F_0 \in \mathcal{Z}(I^h_{\leq t'})$ sei. Darin kommt nur zum Ausdruck, daß zu Xs Erfahrung zwischen t und t' zum einen seine eigenen Handlungen in diesem Zeitraum gehören und zum anderen ein weiteres Ereignis, in das Handlungen nicht eingehen und das natürlich auch aus diesem Zeitraum stammen muß.

Der durch die einfache Konditionalisierung hergestellte Zusammenhang zwischen P und P' lautet dann, sofern $P_{F_0 \cap F}(B_0) \neq \varnothing$ für ein $F \in \mathcal{Z}(I^h_{>t'})$[4]: Für alle $A \in \mathcal{A}(I)$ und $F \in \mathcal{Z}(I^h_{>t'})$ ist

(4.1) $$P'_F(A) = P_{F_0 \cap F}(A \mid B_0).$$

Dies ist aber noch eine recht spezielle Form der Wahrscheinlichkeitsdynamik. Sie geht davon aus, daß X so scharfe Erfahrungen macht, daß er zu t' von gewissen Propositionen mit Gewißheit überzeugt ist. Bezüglich seiner eigenen Handlungen mag dies noch angehen, aber allgemein liegt darin, vorsichtig ausgedrückt, eine unnötige Beschränkung. Dieser Punkt bekümmerte auch Jeffrey; deshalb schlug Jeffrey in (1965), Kap. 11, eine Verallgemeinerung der einfachen Konditionalisierung vor, die dem Umstand Rechnung tragen soll, daß Beobachtungen vielfach, insbesondere bei ungünstigen Beobachtungsbedingungen, zu keinen definitiven Resultaten führen. Machen wir uns diese Verallgemeinerung erst an einem kleinen Beispiel klar: Die abstiegsbedrohten Stuttgarter Kickers haben ein vorentscheidendes Heimspiel. Meine subjektiven Wahrscheinlichkeiten mögen folgendermaßen aussehen:[5]

P(Sieg) = 0,5	P(Klassenerhalt | Sieg) = 0,7,
P(Unentschieden) = 0,3	P(Klassenerhalt | Unentschieden) = 0,4,
P(Niederlage) = 0,2	P(Klassenerhalt | Niederlage) = 0,2;

entsprechend die bedingten Wahrscheinlichkeiten für den Abstieg; und schließlich als Konsequenz der vorigen Werte

P(Klassenerhalt) = 0,7 · 0,5 + 0,4 · 0,3 + 0,2 · 0,2 = 0,51,
P(Abstieg) = 0,49.

Höre ich nun im Radio, die Kickers hätten Unentschieden gespielt, so wird meine neue Wahrscheinlichkeitsverteilung P' gerade das durch dieses Ereignis bedingte P sein. Dies besagte ja gerade die einfache Konditionalisierung. Nehmen wir nun aber an, ich hätte das Radio etwas zu spät eingeschaltet und erführe nur

[4] Wenn dies für ein solches F gilt, dann auch für *alle*, da P ja eine Wahrscheinlichkeitsfamilie für ein *temporales* Entscheidungsfeld gemäß Definition 17 (S. 115) ist.

[5] Da es sich im Beispiel um lauter handlungsunabhängige Ereignisse dreht, können wir die Handlungsindizes an P und P' weglassen.

noch, daß die Kickers zu drei gespielt hätten. Über Sieg und Niederlage ist mit dieser Information nicht entschieden, aber es ist anzunehmen, daß sich meine Wahrscheinlichkeiten dafür drastisch ändern. Folgende Werte dürften nicht unplausibel sein:

P'(Sieg) = 0,1,
P'(Unentschieden) = 0,4,
P'(Niederlage) = 0,5.

Welche Hoffnungen sollte ich nun noch auf den Klassenerhalt der Kickers setzen? Jeffrey (1965), Abschnitte 11.3 und 11.6, macht hier nun die vernünftige Annahme, daß die *bedingten* Wahrscheinlichkeiten P(Klassenerhalt | Sieg) etc. durch diese Information nicht im geringsten tangiert werden, daß also wie für P gilt:

P'(Klassenerhalt | Sieg) = 0,7,
P'(Klassenerhalt | Unentschieden) = 0,4,
P'(Klassenerhalt | Niederlage) = 0,2.

P'(Klassenerhalt) errechnet sich dann zu 0,33.

Allgemein funktioniert diese *verallgemeinerte Konditionalisierung* folgendermaßen: Wie gesagt, können wir bei ihr nicht mehr davon ausgehen, daß X zwischen t und t' gleich mit Sicherheit gewisse Ereignisse erfährt; vielmehr dürfen wir nur annehmen, daß X irgendwelche neue Wahrscheinlichkeiten für die Ereignisse aus diesem Zeitraum hat, die wir durch ein Wahrscheinlichkeitsmaß Q auf $\mathcal{A}(J)$ mit $J = I_{>t} \cap I_{\leq t'}$ repräsentieren wollen; dabei könnte man J den Erfahrungsbereich und Q ein Erfahrungsmaß nennen.[6] Plausibel erscheint auch die Annahme, daß es ein $F_0 \in \mathcal{Z}(I^h_{\leq t'})$ mit $Q(F_0) = 1$ gibt, d.h. daß die Beobachtung der eigenen Handlungen nicht unscharf ist. Unerläßlich ist dann allerdings die Bedingung, daß für alle $B \in \mathcal{A}(J)$ mit $P_{F_0 \cap F}(B) = 0$ für irgendein $F \in \mathcal{Z}(I^h_{>t'})$[4] auch $Q(B) = 0$ gilt; andernfalls wäre die nachfolgende Formel (4.2) mathematisch nicht sinnvoll. Gemäß der verallgemeinerten Konditionalisierung bestimmt sich dann P' so: Für alle $A \in \mathcal{A}(I)$ und $F \in \mathcal{Z}(I^h_{>t'})$ ist

---

[6] Man beachte, daß Q ein echtes Wahrscheinlichkeitsmaß ist. Das liegt daran, daß J von $I^h_{>t'}$ probabilistisch unabhängig ist.

$$(4.2) \qquad P'_F(A) \;=\; \sum_{B \in \mathcal{Z}(J \setminus I^h)} P_{F_0 \cap F}(A \mid B) \cdot Q(B) \;.$$

Es ist klar, daß das so festgelegte P' eine Wahrscheinlichkeitsfamilie: für $\langle I,\ I^h_{>t'} \rangle$ ist und daß insbesondere, wie gewünscht, für alle $A \in \mathcal{A}(J)$ und $F \in \mathcal{Z}(I^h_{>t'})$ $P'_F(A) = Q(A)$ gilt.

Man kann die Gleichung (4.2) auf zwei verschiedene Weisen mit Worten erläutern: Einmal kann man sagen, daß man die neue Wahrscheinlichkeit von X für A unter F mit Rückgriff darauf berechnen kann, daß Xs durch die möglichen Beobachtungsergebnisse bedingten Wahrscheinlichkeiten für A gleich geblieben sind. Und zum andern kann man die neue Wahrscheinlichkeitsfamilie P' von X als Mischung aus den Wahrscheinlichkeitsfamilien charakterisieren, die X gemäß der einfachen Konditionalisierung hätte, wenn er nicht nur F_0, sondern auch eines der $B \in \mathcal{Z}(J \setminus I^h)$ definitiv beobachtet hätte; die Mischungsfaktoren sind dabei gerade die neuen Wahrscheinlichkeiten für die $B \in \mathcal{Z}(J \setminus I^h)$.

Levi (1967) hält dieses verallgemeinerte Konditionalisierungsverfahren für überflüssig, ja sogar für unbegründet. Entweder, so läuft seine Argumentation, läßt sich etwas anführen, was den Übergang von P zu P' rechtfertigt; dann kann man aber diesen Übergang durch einfache Konditionalisierung beschreiben. Oder man kann keinen Grund für den Übergang angeben; dann jedoch ist die Änderung der Wahrscheinlichkeiten nicht vernünftig. Levi akzeptiert die verallgemeinerte Konditionalisierung also gerade dann, wenn sie sich auf die einfache Konditionalisierung zurückführen läßt.

Eine solche Zurückführung ist nun sicherlich häufig möglich. Im obigen Beispiel etwa könnte man mir nicht nur Wahrscheinlichkeiten der Form P(Sieg), P(Niederlage) etc. zuschreiben, sondern auch solche von der Form P(Sieg | der Gegner erzielt 0 Treffer), P(Unentschieden | der Gegner erzielt 2 Treffer) etc., und dann in den neuen Wahrscheinlichkeiten für Sieg und Niederlage gerade das Ergebnis der einfachen Konditionalisierung bezüglich des Ereignisses, daß der Gegner 3 Tore geschossen hat, sehen. Wie Jeffrey selbst bemerkt[7], bereitet jedoch die sprachliche Formulierung des Beobachteten häufig große Schwierigkeiten. Über Wendungen wie „mir scheint, als wäre das und das geschehen“ kommt man oft nicht hinaus. Und ob das überhaupt Propositionen sind, bezüglich derer man bedingte Wahrscheinlichkeiten ausrechnen kann, ist äußerst frag-

---

[7] Jeffrey (1965), S. 154f.

lich. Solche Äußerungen dürften vielmehr ausdrücken, daß man gerade eine Wahrscheinlichkeit kleiner als 1 für das fragliche Geschehnis hat, und nicht, daß man im Grade 1 daran glaubt, daß einem etwas scheint.

Levis Standpunkt läuft also auf eine Art Vollständigkeitsforderung hinaus: Für jede vernünftige Änderung subjektiver Wahrscheinlichkeiten läßt sich eine Proposition formulieren, so daß sich diese Änderung als eine Änderung gemäß der einfachen Konditionalisierung bezüglich dieser Proposition ansehen läßt. Es ist schon zweifelhaft, ob selbst so etwas Reichhaltiges wie unsere Umgangssprache dieser Forderung genügt. Umso eher sollte man sie nicht an eine begrenzte Propositionenmenge stellen, mit der man eine Entscheidungssituation zu erfassen versucht. Insofern stellt Jeffreys allgemeine Konditionalisierung ohne Frage eine wesentliche Bereicherung für die Behandlung der Dynamik von Glaubensdispositionen dar.

Im übrigen hat Teller (1976), S. 241ff., nachgewiesen, daß das Verfahren der verallgemeinerten Konditionalisierung mit der folgenden Bedingung an P und P' äquivalent ist: Für alle $A, A' \in \mathcal{A}(I)$ mit $A, A' \subseteq B$ für ein $B \in \mathcal{Z}(J)$ gelte:

$$\text{wenn } P(A) = P(A'), \text{ dann } P'(A) = P'(A').$$

Wichtig ist dabei, daß man diese Bedingung rein qualitativ, soll heißen, mit Hilfe einer komparativen Glaubensrelation formulieren kann. Die durch (4.2) gegebene verallgemeinerte Konditionalisierung läßt sich also auf eine sehr harmlose und auch entschieden leichter begründbare und nachprüfbare Bedingung zurückführen.

Die eben geschilderte verallgemeinerte Konditionalisierung bedarf im Grunde noch einer weiteren Verallgemeinerung: Mit ihr ist zwar das Problem der unscharfen Beobachtung gelöst; doch das bisher nicht erwähnte, eigentlich aber viel drängendere Problem der unscharfen Erinnerung bleibt unbehandelt. In der Tat wurde es der einfachen Konditionalisierung oft zum Vorwurf gemacht, daß sie ein perfektes Gedächtnis voraussetze, das alle einmal registrierten Einzelheiten fest speichere.[8] Und insoweit Jeffrey, mit seiner verallgemeinerten Konditionalsierung nur das Problem der unscharfen Beobachtung angeht, ist ihm derselbe Vorwurf zu machen. Glücklicherweise läßt sie sich leicht zur Behandlung unscharfer Erinnerungen einsetzen: Es ist jetzt davon auszugehen, daß *alle* vergan-

[8] S. etwa Suppes (1966), S. 28f., oder Stegmüller (1973a), S. 402.

genen Ereignisse neue Wahrscheinlichkeiten erhalten können, wie es dem Gedächtnis eben beliebt. Vergangene Handlungen sind davon natürlich nicht ausgeschlossen, wie etwa die bei Urlaubsantritt besonders gerne auftretenden Zweifel dokumentieren, ob man den Gashahn auch wirklich zugedreht habe. Diese neuen Wahrscheinlichkeiten werden formal wieder durch ein Wahrscheinlichkeitsmaß Q auf $A(J)$ ausgedrückt, wobei diesmal aber $J = I_{\leq t'}$ sei, also die gesamte Vergangenheit umfasse; Q könnte man jetzt Erinnerungs- oder Gedächtnismaß nennen. Aus mathematischen Gründen muß dabei für alle $B \in \mathcal{A}(J)$, $F_0 \in \mathcal{Z}(I^h_{\leq t'})$ und $F \in \mathcal{Z}(I^h_{>t'})$ gelten: Wenn $P_{F_0 \cap F}(B) = 0$, dann $Q(F_0 \cap B) = 0$. Für P' gilt dann eine Verallgemeinerung von (4.2): Für alle $A \in \mathcal{A}(I)$ und $F \in \mathcal{Z}(I^h_{>t'})$ ist

$$(4.3) \qquad P'_F(A) \;=\; \sum_{F_0 \in \mathcal{Z}(I^h_{\leq t'})} \; \sum_{B \in \mathcal{Z}(J \setminus I^h)} P_{F_0 \cap F}(A \mid B) \cdot Q(F_0 \cap B) \;.$$

Ich will nun nicht diskutieren, inwieweit man damit dem menschlichen Gedächtnis gerecht wird. Mir kam es nur darauf an zu zeigen, wie sich möglicherweise ein erster Schritt tun ließe, um die Entscheidungstheorie in einem wichtigen Aspekt realistischer zu gestalten. Freilich wäre es nur ein sehr kleiner erster Schritt; eigentlich interessant wird es ja erst, wenn es gilt, über das Gedächtnismaß detailliertere Aussagen zu treffen.

So weit die verschiedenen Konditionalisierungsformen der Wahrscheinlichkeitsdynamik. Natürlich kommen auch andere Änderungen subjektiver Wahrscheinlichkeiten vor. Häufig werden neue Informationen nicht nach den Konditionalisierungsverfahren, sondern auf andere mehr oder weniger bizarre Weise verarbeitet. Die Änderung muß auch gar nicht durch neue Informationen oder durch einen veränderten Erinnerungsstand ausgelöst sein. Z.B. kann auch der Alkohol die verblüffendsten Wirkungen zeitigen. Doch läßt sich zu all diesen Fällen nur durch empirische Einzelforschung Genaueres sagen. Auch ist festzuhalten, daß einer Änderung der subjektiven Wahrscheinlichkeiten allein nicht anzumerken ist, ob sie auf neue Erfahrungen oder andere Faktoren zurückzuführen ist. Darüber läßt sich erst etwas sagen, wenn man weiß, worauf die Sinnesorgane des fraglichen Individuums ausgerichtet waren und welchen sonstigen Einflüssen es ausgesetzt war.

Zum Verhältnis zwischen V und V', also zur Dynamik der subjektiven Werte, läßt sich allgemein wiederum sehr wenig sagen. Man könnte allenfalls eine gewisse zeitliche Konstanz der subjektiven Werte annehmen. Doch läßt sich auch hier nur in konkreten Fällen mehr sagen. Ein Punkt ist hier jedoch noch zu erwähnen: Die E1-Modelle, die X zu t und t' charakterisieren, können ja bereits Reduktionen von umfassenderen E1-Modellen sein, so daß V und V' in Bezug auf diese umfassenderen Modelle erwartete subjektive Werte ausdrücken. Daher kann es geschehen, daß sich eine Wahrscheinlichkeitsänderung von t zu t' nicht nur in den Wahrscheinlichkeitsfamilien P und P' der reduzierten Modelle (wenn überhaupt), sondern auch direkt in einer korrespondierenden Änderung von V' gegenüber V niederschlägt. Dies gilt insbesondere für die erwarteten subjektiven Werte von Handlungsverläufen; U' ändert sich gegenüber U nach Maßgabe der Änderungen von P' und V' gegenüber P und V.

4.3 Entscheidungsmodelle, die zukünftige Erfahrungen berücksichtigen

Nachdem wir im letzten Abschnitt auf die Dynamik von E1-Modellen eingegangen sind und dabei immerhin in einem Punkt einigermaßen Ergiebiges sagen konnten, wollen wir nun auf dem im Abschnitt 4.1 vorgezeichneten Weg fortschreiten und in möglichst großer Allgemeinheit auch für diejenigen Individuen eine statische Theorie formulieren, die ihre eigene Dynamik mitbedenken, die sich also darauf einrichten, daß sie selbst solchen Änderungen, wie wir sie eben diskutierten, unterworfen sind. Aus Gründen, die noch klar werden, ist dieses Ziel nicht im direkten Angriff, sondern nur durch eine zweistufige Operation zu nehmen. Deren erster Teil besteht in der Bewältigung des Spezialfalls, in dem das zu beschreibende Individuum – sei es mal wieder unser treuer Begleiter X – lediglich in Betracht zieht, daß es diese oder jene Erfahrungen sammeln wird und daß sich seine subjektiven Wahrscheinlichkeiten gemäß der einfachen oder verallgemeinerten Konditionalisierung ändern werden.

Wir können dabei davon ausgehen, daß X seiner eigenen dynamischen Entwicklung einen festen Spielraumbestand unterlegt. Denn wenn X jetzt mutmaßt, daß er später einen bestimmten Spielraum in Betracht ziehen wird, so stellt er diesen Spielraum ja auch jetzt schon in Rechnung.

Der allerelementarste Fall sieht dann so aus: X bedenke zu t_0 seine Situation, in der er sich zu einer der Handlungen $F_1, ..., F_m$ zu t_2 entscheiden muß; er erwarte dabei, vorher zu t_1 einen Spielraum mit den Realisierungsmöglichkeiten $A_1, ..., A_n$ zu beobachten. Dann wird sich X natürlich nicht schon zu t_0 auf eine Handlung F_p festlegen, die zu t_0 maximalen erwarteten subjektiven Wert hat. Vielmehr wird er abwarten, welches A_q zu t_1 geschieht, und erst dann mit Hilfe seiner durch das beobachtete A_q bedingten subjektiven Wahrscheinlichkeiten eine Handlung F_p mit maximalem erwartetem subjektivem Wert bestimmen. Allenfalls wird er sich schon zu t_0 überlegen, welche Handlung er bei welcher Beobachtung wählen, d.h. welche Strategie er einschlagen soll.

Im allgemeinen wird X natürlich mehrere Planungsstufen zu bewältigen haben. Es werden mehrere Handlungsspielräume zu verschiedenen Zeitpunkten zur Debatte stehen, und die X offenstehenden Strategien werden auf mehrere, zu verschiedenen Zeitpunkten zu beobachtende Spielräume abzustimmen zu sein. Wichtiger ist freilich, daß *X* nicht fixe Erwartungen darüber haben wird, welche Spielräume er beobachten wird; dies werden in der Regel vielmehr handlungsabhängige Erwartungen sein.

Dies ist ein ganz trivialer Punkt; schon daß *X* überhaupt seine Augen offen hält, ist eine Handlung mit ganz entscheidendem Einfluß auf seine Erfahrungen; und erst recht ist es so, wenn *X* mehr oder weniger raffinierte Experimente anstellen muß, um zu seinen Beobachtungen zu gelangen. Doch tut sich hier die interessante Nebenfrage auf, welche Beobachtungen Handlungen sind und welche nicht. Auch ohne genauere Erörterung dieser Frage scheint mir jedenfalls klar zu sein, daß die meisten, wenngleich nicht alle Beobachtungen Handlungen sind. So kann man sich gewisser Sinneseindrücke nicht erwehren, z.B. sehr lauter Geräusche oder taktiler Eindrücke. Insbesondere aber darf man nicht verlangen, daß Beobachtungen der eigenen Handlungen wiederum Handlungen sind; dies führte sofort in einen unendlichen Regreß. Es erscheint umgekehrt die Annahme plausibel, daß man von den eigenen Handlungen, so weit sie Gegenstand der Entscheidungstheorie sind, nicht erst durch gesonderte Beobachtung, sondern direkt aufgrund ihrer Durchführung Kenntnis hat. Doch habe ich zu diesem Punkt – leider – nicht mehr zu sagen.

Wenden wir uns wieder unserem *X* und seinen sich komplizierenden Verhältnissen zu. Wir waren bei dem Fall angelangt, der durch die spezielle Annahme gekennzeichnet ist, daß *X sichere*, aber *handlungsabhängige* Erwartungen darüber hat, welche Spielräume er beobachtet. Dieser Fall dürfte auch der wichtigste und am häufigsten vorliegende sein. In der Tat ist er derjenige, den Lehrbücher der Entscheidungstheorie in der Regel behandeln.[9] Und in ihm tun sich schon sehr komplizierte Probleme auf, wie z.B. ein Blick in die statistische Entscheidungstheorie lehrt.[10] Freilich bleibt man in der Regel auch bei diesem Fall stehen. Das liegt an einer zu engen Sichtweise strategischen Denkens, die meint, mit Strategien stelle man sich lediglich auf verschiedene äußere Gegebenheiten ein. Nach unserer Auffassung stellt man sich damit aber vielmehr auf verschiedene neue subjektive Beurteilungen der Situation ein und natürlich mittelbar dann auch auf neue äußere Gegebenheiten, insoweit sie beobachtet werden und

---

[9] Vgl. etwa Fishburn (1964), Abschnitt 2.3, Raiffa (1973), Kap .2,3,7, oder Bühlmann, Loeffel, Nievergelt (1975), Kap.3.

[10] Vgl. etwa Wald (1950) oder Bühlmann, Loeffel, Nievergelt (1975), 3. Teil.

dadurch eine veränderte subjektive Beurteilung erzeugen. Es gilt also, auch die genannte spezielle Annahme noch aufzulockern.

Zunächst kann man zulassen, daß X nur noch mit einer gewissen Wahrscheinlichkeit erwartet, einen bestimmten Spielraum zu beobachten. Dies wird z.B. immer dann der Fall sein, wenn sich X zur Beobachtung irgendeiner Apparatur bedient, in deren Funktionstüchtigkeit er kein volles Vertrauen hat. Vor allem aber kann man die auch jetzt noch gemachte Annahme aufgeben, daß X in der einfachen Konditionalisierung das Verfahren sieht, nach dem sich seine subjektiven Wahrscheinlichkeiten ändern. X kann ja nicht nur berücksichtigen, daß er einen Spielraum entweder ganz oder gar nicht beobachtet, sondern auch, daß er ihn in unterschiedlichem Maße unscharf beobachtet und daß dann Jeffreys verallgemeinerte Konditionalisierung heranzuziehen ist. X hat dann also handlungsabhängige Wahrscheinlichkeiten für verschiedene mögliche Wahrscheinlichkeitsverteilungen über bestimmte Ereignisse: Verteilungen, die sich bei ihm aufgrund verschiedener unscharfer Beobachtungen einstellen können. So kompliziert sich das anhört, so häufig werden solche Überlegungen oder zumindest Konsequenzen daraus virulent. Da sich nämlich eine Entscheidungssituation im allgemeinen umso günstiger darstellt, je schärfer die gemachten Beobachtungen sind, versucht man in der Regel, sofern es nicht zu aufwendig ist, zu möglichst scharfen Beobachtungen zu gelangen – indem man z.B. eine Brille aufsetzt, zusätzlich zuverlässige Linienrichter postiert, die Scheibenwischer oder die Scheinwerfer anstellt etc.

Die verallgemeinerte Konditionalisierung in der Form, in der sie auch unvollkommenen Erinnerungen gerecht werden könnte, kommt jetzt freilich noch nicht zum Zuge. Denn wie gesagt, geht es uns im Augenblick ja nur um den Fall, in dem X verschiedene, mögliche Informations*verbesserungen* in Betracht zieht. In der Tat dürften wir diesen Fall mittlerweile bis zur größtmöglichen Allgemeinheit ausgereizt haben. Wie lassen sich nun all diese Überlegungen in einem präzisen Entscheidungsmodell wiedergeben? Wie ist dafür ein Strategiebegriff zu formulieren? Und wie lautet vor allem die zugehörige Entscheidungsregel bzw. die zugehörige empirische Behauptung?

Die Beantwortung dieser Fragen fällt wesentlich leichter, wenn wir uns eines in der Entscheidungstheorie üblichen gedanklichen Tricks bedienen: wenn wir den rekursiven Charakter dieser Überlegungen ausnutzen. Dieser besteht in folgendem: Umfasse der Planungszeitraum von X gerade die Zeitpunkte t_0, ..., t_n

(wobei natürlich $t_p < t_q$ für $p < q$). D.h. *X*s gegenwärtiger Zeitpunkt sei t_0; der erste Zeitpunkt, zu dem er seines Erachtens aufgrund neuer Erfahrungen eine veränderte Einschätzung seiner Situation haben kann, sei t_1; mit der nächsten Änderung rechne er für t_2; etc; und Änderungen, die sich nach t_n möglicherweise noch einstellen, berücksichtige er nicht. Dann müssen alle Entscheidungsmodelle, die die von ihm für möglich gehaltenen Situationen zu t_n beschreiben, E1-Modelle sein; sein strategisches Denken trägt ja nicht über t_n hinaus. Faßt er nun eine mögliche Situation zu $t_{n\text{-}1}$ ins Auge, so muß er neben dem aus E1-Modellen Bekannten auch noch berücksichtigen, daß ihn zu t_n dieses oder jenes E1-Modell charakterisieren wird. Eine Strategie für diese Situation zu $t_{n\text{-}1}$ wird dann erstens sagen, was er zwischen $t_{n\text{-}1}$ und t_n tun soll, und zweitens für jede mögliche Situation zu t_n, was er nach t_n tun solle, wenn er in sie geriete. Bedenkt er im nächsten Schritt eine mögliche Situation zu $t_{n\text{-}2}$, so hat er in Betracht zu ziehen, daß er sich zu $t_{n\text{-}1}$ in der einen oder anderen Situation der eben beschriebenen Art befinden wird; und eine Strategie für diese Situation zu $t_{n\text{-}2}$ wird erstens sein Verhalten zwischen $t_{n\text{-}2}$ und $t_{n\text{-}1}$ und zweitens für jede dieser Situationen zu $t_{n\text{-}1}$ eine Strategie festlegen, die das weitere Verhalten nach $t_{n\text{-}1}$ regelt. Und so weiter, bis wir bei t_0 angelangt sind.

Der Witz ist nun, daß wir die drei obigen Fragen nur in Bezug auf den Rekursionsschritt zu beantworten haben. D.h. wir müssen lediglich, davon ausgehend, daß wir sie bereits für die etwa zu t_1 gegebenen Situationen erledigt haben, sie dann auch für Situationen zu t_0 lösen. Denn den Rekursionsanfang haben wir schon; die spätesten, in Betracht gezogenen Situationen werden ja durch E1-Modelle adäquat beschrieben. Mit dem Rekursionsschritt haben wir dann also alle in diesem Abschnitt diskutierten Entscheidungssituationen im Griff. Zu betonen ist freilich, daß dieses rekursive Verfahren nur funktioniert, wenn *X*s Planungszeitraum endlich ist; denn nur dann ist gewährleistet, daß es einen Rekursionsanfang gibt.[11]

Die formale Wiedergabe dieses Rekursionsschritts dürfte aufgrund des bisher Gesagten ziemlich klar sein:

---

[11] Da hier alles, insbesondere die Menge der Zeitindizes, als endlich angenommen ist, ist diese Voraussetzung hier automatisch erfüllt. In der Entscheidungstheorie gibt es Methoden, auch mit aus unendlich vielen Zeitpunkten bestehenden Planungsperioden zurechtzukommen; vgl. etwa Mine, Osaki (1970). Ob sich diese Methoden auf die hier betrachteten Situationen übertragen lassen, habe ich nicht nachgeprüft.

Definition 27: $\Delta_0 = \langle I, I^h, T, i, \leq, s, P, V, U\rangle$ ist ein *E2-Modell für t* genau dann, wenn gilt:

(1) P ist eine Wahrscheinlichkeitsfamilie für $\langle I \cup \{i\}, I^h, T, \leq, s\rangle$, falls $i \neq \varnothing$, sonst für $\langle I, I^h, T, \leq, s\rangle$,

(2) wenn P' die Einschränkung von P auf $\langle I, I^h\rangle$ ist, so ist $\langle I, I^h, T, \leq, s, P', V, U\rangle$ ein E1-Modell für t,

(3) wenn $i \neq \varnothing$, dann ist i ein Spielraum mit $i \notin I$ und $s_i > t$ derart, dass i eine Menge von E2-Modellen für s_i der Form $\Delta = \langle I, I^h_{>s_i}, i_\Delta, T, \leq, s, P^\Delta, V_\Delta, U_\Delta\rangle$ ist.

Wir bezeichnen i dann auch als *die Rekursionsquelle von* Δ_0.

Da in Bedingung (3) wieder von E2-Modellen die Rede ist, ist Definition 27 eine *rekursive* Definition und überdies eine korrekte. Denn E2-Modelle mit leerer Rekursionsquelle sind nichts anderes als El-Modelle, und da T endlich und $s_i > t$ sein muß, müssen wir nach endlich vielen Schritten bei diesem elementaren Fall angelangt sein. E2-Modelle mit nicht-leerer Rekursionsquelle gehen dann einfach darin über E1-Modelle hinaus, daß in ihnen noch ein besonderer Spielraum auftaucht, der aus all den E2-Modellen besteht, die die zu einem späteren Zeitpunkt möglichen Situationen beschreiben. Dieser Spielraum muß natürlich in die subjektiven Wahrscheinlichkeiten einbezogen sein; das Modell muß ja sagen, mit welcher Wahrscheinlichkeit man in welche Situationen dieses Spielraums zu kommen meint. Dagegen sah ich keinen Sinn darin, diesen Spielraum auch in die subjektive Wertfunktion des Modells einzubeziehen.

Wie der Strategiebegriff für E2-Modelle auszusehen hat, ist ebenfalls klar:

Definition 28: Sei Δ_0 ein E2-Modell für t und i seine Rekursionsquelle. Dann heißt S eine *Strategie für* Δ_0 genau dann, wenn S eine Funktion auf $\{\Delta_0\} \cup i$ ist derart, daß $S(\Delta_0) \in \mathcal{Z}(I^h_{\leq s_i})$[12] und für jedes $\Delta \in i$ $S(\Delta)$ eine Strategie für Δ ist.

Dies ist also wieder eine rekursive Definition, deren Korrektheit auf denselben Gründen wie die von Definition 27 beruht. Eine Strategie für Δ_0 sagt also, wie zwischen t und s_i zu verfahren ist und wie das Vorgehen nach s_i, je nachdem, in welches $\Delta \in i$ man gerät, auszusehen hat.

---

[12] Wobei $s_i = \max_{t \in T} t$; falls $i = \varnothing$; diese Regelung gelte gleich für den Rest des Kapitels.

Doch werden E2-Modelle noch nicht dem Umstand gerecht, daß es in diesem Abschnitt Entscheidungssituationen zu repräsentieren gilt, in denen ausschließlich zukünftige Erfahrungen berücksichtigt werden. Dazu ist zumindest die folgende Verschärfung nötig:

Definition 29: $\Delta_0 = \langle I, I^h, i, T, \leq, s, P, V, U\rangle$ ist ein E3-Modell für t genau dann, wenn gilt:

(1) Δ_0 ist ein E2-Modell für t,

(2) wenn $i \neq \varnothing$, dann ist i eine Menge von E3-Modellen für s_i, für die gilt: für alle $A \in \mathcal{A}(I)$, $B \in \mathcal{Z}(I^g)$, $F \in \mathcal{Z}(I^h_{\leq s_i})$ und $G \in \mathcal{Z}(I^h_{>s_i})$ ist

$$P_{F\cap G}(A) = \sum_{\Delta\in i} P_G^\Delta(A)\cdot P_{F\cap G}(\Delta) \tag{4.4}$$

[13]

und

$$V(B\cap F\cap G) = \sum_{\Delta\in i} V_\Delta(B\cap F\cap G)\cdot P_{F\cap G}(\Delta) \ . \tag{4.5}$$

Die Gleichungen (4.4) und (4.5) sind eben notwendige Bedingungen dafür, dass die in einem E3-Modell für t anvisierten Änderungen der Situation durch zusätzliche Erfahrungen ausgelöste Änderungen sein sollen. Besonders deutlich ist dies bei (4.4), das gewissermaßen das Analogon zur Formel von der totalen Wahrscheinlichkeit ist, welche für ein Wahrscheinlichkeitsmaß Q, ein Ereignis A und eine Menge $\{B_1, ..., B_n\}$ paarweise disjunkter und insgesamt erschöpfender Ereignisse besagt, daß

$$Q(A) = \sum_{r=1}^{n} Q(A \mid B_r)\cdot Q(B_r).$$

(4.4) verlangt also, daß die *alte* probabilistische Einschätzung gleich einem gewichteten Mittel der für möglich gehaltenen neuen Einschätzungen ist, wobei die Gewichte gerade die Wahrscheinlichkeiten für die neuen Einschätzungen sind. Man beachte, daß aus (4.4) folgt – wenn man $A = F$ setzt –, daß es ein $\Delta \in i$ mit $P_G^\Delta(F) = 1$ gibt, ja dass sogar $\sum_\Delta P_{F\cap G}(A) = 1$, summiert über diejenigen Δ mit $P_G^\Delta(F) = 1$ (wobei das G ganz unerheblich ist). Das Wissen zu s_i um die zwischen t und s_i ausgeführten Handlungen ist also Bestandteil von Definition 29.

[13] Die Schreibweise „$P_{F\cap G}(\Delta)$“ erzeugt bestimmt kein Mißverständnis obwohl sie nicht ganz korrekt ist. Genaugenommen müßte darin für Δ der entsprechende Zustand von i stehen.

Neben (4.4) ist aber auch (4.5) notwendig. Das ist die Folge des am Ende von Abschnitt 4.2 Bemerkten. Die neuen Erfahrungen könnten ja Spielräume betreffen, die im E3-Modell selbst nicht enthalten sind und sich daher im Übergang von P zu einem der P^{Δ} gar nicht oder nur unzulänglich, dafür aber im Übergang von V zu einem der V_{Δ} niederschlagen; deswegen nämlich, weil V und die V_{Δ} *erwartete* subjektive Werte in Bezug auf diese Spielräume darstellen können. Daher ist auch die zu (4.4) analoge Bedingung für subjektive Werte zu verlangen.

Es ist freilich darauf hinzuweisen, daß (4.4) und (4.5) lediglich notwendige Bedingungen dafür darstellen, daß die für möglich gehaltenen Änderungen vom ursprünglichen Modell für t zu einem der Modelle aus i allein auf Informationsgewinn beruhen. Ob und wie sich mit den bisher verwandten begrifflichen Mitteln auch hinreichende Bedingungen dafür formulieren lassen, ist mir nicht klar.

Wenden wir uns nun der entscheidenden Frage nach einem für E3-Modelle angemessenen Optimalitätskriterium für Strategien zu. Die übliche Antwort darauf besteht darin, auch für Strategien einen erwarteten subjektiven Wert zu definieren und diesen dann zu maximieren. In dem normalerweise behandelten, spezielleren Rahmen, den die Annahme kennzeichnet, daß der Entscheidende sichere, aber handlungsabhängige Erwartungen über seine Beobachtungen hat, bieten sich zwei äquivalente Verfahren zur Definition eines erwarteten subjektiven Wertes für Strategien an. Das erste beruht auf der Tatsache, daß in diesem Rahmen bei jedem Zustand $A \in \mathcal{Z}(I^g)$ der Geschehensspielräume aus jeder Strategie S genau ein Handlungsverlauf $S_A \in \mathcal{Z}(I^h)$ resultiert. Man kann dann nämlich die Wahrscheinlichkeit $P_S(A)$ von A unter der Strategie S mit $P_{S_A}(A)$ gleichsetzen, erhält daraus allgemein durch Strategien bedingte Wahrscheinlichkeiten für Ereignisse und kann schließlich ebenso wie für Handlungsverläufe auch für Strategien einen erwarteten subjektiven Wert bestimmen[14]; konkret ergibt sich

$$U^*(S) = \sum_{A \in \mathcal{Z}(I^g)} V(S_A \cap A) \cdot P_{S_A}(A) \; .$$

In unserem allgemeinerem Rahmen gilt aber die genannte Tatsache nicht mehr, so daß dieses Verfahren hier nicht anwendbar ist; insbesondere lassen sich also in unserem Rahmen durch Strategien bedingte Wahrscheinlichkeiten nicht sinnvoll erklären.

[14] Vgl. etwa Fishburn (1964), Abschnitte 2.3-3.1.

Das zweite Verfahren[15] hingegen ist verallgemeinerungsfähig; es läßt sich an der früher geschilderten Entscheidungssituation von X mit den Planungsstufen von t_0 bis t_n veranschaulichen: Die Strategien für die möglichen Situationen zu t diktieren einfach Handlungsverläufe ab t_n, deren erwarteten subjektiven Wert wir schon definiert haben. Eine Strategie für eine Situation zu $t_{n\text{-}1}$ legt dann erstens fest, mit welcher Wahrscheinlichkeit X in welche mögliche Situationen zu t_n kommt (da sie einen Handlungsverlauf zwischen $t_{n\text{-}1}$ und t_n bestimmt), und zweitens, welchen erwarteten subjektiven Wert X in welcher möglichen Situation zu t_n hat (da sie für jede solche Situation einen Handlungsverlauf nach t_n bestimmt). Eine erneute Mittelung ergibt dann einen erwarteten subjektiven Wert für diese Strategie. Und so weiter.

Präzise formuliert heißt das:

Definition 30: Sei $\Delta_0 = \langle I, I^h, i, T, \leq, s, P, V, U\rangle$ ein E3-Modell für t. Dann heißt U^* *die erwartete subjektive Wertfunktion für Strategien für* Δ_0 genau dann, wenn U^* diejenige Funktion ist, für die für alle Strategien S für Δ_0 gilt: wenn $i = \varnothing$, dann $U^*(S) = U(S(\Delta_0))$, und wenn $i \neq \varnothing$, dann

$$U^*(S) = \sum_{\Delta \in i} U^*_\Delta(S(\Delta)) \cdot P_{S(\Delta_0)}(\Delta)$$ [16]

wobei U^*_Δ die erwartete subjektive Wertfunktion für Strategien für Δ sei.

Die mit E3-Modellen verknüpfte empirische Behauptung lautet dann: Werde X zu t durch das E3-Modell Δ_0 für t charakterisiert, und sei U^* die erwartete subjektive Wertfunktion für Strategien für Δ_0. Dann fängt X an, eine Strategie S_0 für Δ_0 durchzuführen, für die $U^*(S_0)$ maximal ist (für die also $U^*(S_0) \geq U^*(S)$ für alle Strategien S für Δ_0).

Grob, aber einigermaßen bündig könnte man also sagen, daß der *gegenwärtige* Wert einer Strategie immer vom Standpunkt der *zukünftigen*, veränderten Situationseinschätzungen aus beurteilt wird. Wie wir sehen werden, kennzeichnet das den in diesem Abschnitt beschriebenen Fall, in dem der Entscheidende le-

---

[15] Vgl. etwa Savage (1954), Abschnitt 6.2.

[16] Der Ausdruck „$P_{S(\Delta_0)}(\Delta)$" enthält neben der in Anm. 13 vermerkten Unkorrektheit noch eine weitere; korrekterweise müßte man „$P_{S(\Delta_0)\cap G}(\Delta)$" für ein $G \in Z(I^h_{>s_i})$ schreiben. Da aber P eine Wahrscheinlichkeitsfamilie für ein *temporales* Entscheidungsfeld ist (vgl. Definition 17, S. 115), kommt es auf das G nicht an.

diglich Änderungen seiner Entscheidungssituation in Betracht zieht, die auf Informationszuwachs beruhen.

4.4 Entscheidungsmodelle, die beliebige Entwicklungen berücksichtigen

Wir können nun zum zweiten Teil unserer Operation übergehen, die nach wie vor das Ziel ansteuert, adäquate Entscheidungsmodelle für Individuen von größtmöglicher Umsichtigkeit zu formulieren; die Beschreibung von Individuen mit geringerer Umsichtigkeit erfordert dann ja nichts weiter als eine geeignete Spezialisierung. Die bisher diskutierte Umsichtigkeit stellte nur zukünftige Erfahrungen in Rechnung. Doch ist aufgrund von Abschnitt 4.2 klar, was ein kluges Individuum, etwa unser X, noch alles berücksichtigen könnte.

Zunächst könnte X bedenken, daß nicht nur eine neue Erfahrung, sondern auch sein mangelhaftes Gedächtnis eine Quelle für die Änderung seiner subjektiven Wahrscheinlichkeiten sein kann; er wird dann eine solche Änderung gemäß der verallgemeinerten Konditionalisierung in unserer etwas erweiterten Lesart ins Auge fassen. Auch wenn ein solcher Gedankengang noch komplizierter zu sein scheint als alles, was wir im letzten Abschnitt diskutierten, er hat doch seine alltäglichen Anwendungen, z.B. immer dann, wenn man sich etwas aufschreibt, bevor man es vergißt. In der Tat müßte so unser gesamter Aufwand an mnemotechnischen Hilfsmitteln seine entscheidungstheoretische Begründung finden können; und da dieser beträchtlich ist, wundere ich mich, daß er, meines Wissens jedenfalls, in der Entscheidungstheorie bisher nicht diskutiert wurde.

Darüber hinaus könnte X sogar völlig wilde Änderungen seiner subjektiven Wahrscheinlichkeiten in Betracht ziehen, und selbst hiervon dürfte es noch Anwendungen geben. So könnte X, solange er sich noch in zurechnungsfähigem Zustand befindet, Vorkehrungen für den Fall treffen, daß er in dem Rauschzustand, in den er sich versetzen will oder der ihn erfahrungsgemäß demnächst überkommt, aus ganz unsinnigen Überzeugungen heraus Dummheiten macht. Z.B. wird er zu dem geplanten Saufgelage mit seinen Freunden nicht mit dem eigenen Auto anreisen.

In einem letzten Schritt könnte X auch noch gewärtigen, daß sich seine subjektiven Werte mit der Zeit ändern, und zwar selbständig und nicht als Folge von Wahrscheinlichkeitsänderungen; letzteres war uns ja schon begegnet. Auf eine

solche Änderung subjektiver Werte werden wir gleich noch etwas ausführlicher zu sprechen kommen.
Wie lassen sich nun diese zusätzlichen Überlegungen formal wiedergeben? Dazu scheinen sich justament die im letzten Abschnitt definierten E2-Modelle zu eignen. Denn es ist klar, daß hier dieselbe rekursive Prozedur zur Anwendung kommt wie dort; wiederum gilt es nur den Rekursionsschritt zu formulieren, nur die zeitlich nächste Änderung der Situation ins Auge zu fassen. Denn die möglichen geänderten Situationen sind, so können wir voraussetzen, bereits so beschrieben, daß sie die endlich vielen weiteren Änderungen berücksichtigen. Dabei sind E2-Modelle in voller Allgemeinheit heranzuziehen, da die einschränkenden Bedingungen (4.4) und (4.5), die E3-Modelle kennzeichneten, hier nicht mehr angebracht sind; mögliche Wahrscheinlichkeitsänderungen aufgrund Erinnerungslücken oder beliebige Änderungen subjektiver Werte werden (4.4) und (4.5) im allgemeinen nicht erfüllen.

Aber ganz so fix ist die Angelegenheit doch nicht abgehandelt, alldieweil die Frage nach einem Optimalitätskriterium für Strategien nun völlig anders zu beantworten ist als zuvor. Wie ist sie denn zu beantworten? Dies gleich in voller Allgemeinheit zu überschauen, wollen wir uns nicht zumuten. Wenden wir uns daher erst Spezialfällen zu, am besten Vorbildern aus der Literatur. Die gibt es nämlich; um genau zu sein, ich kenne gerade eines. Ich meine damit die in der Wirtschaftstheorie seit einiger Zeit schwelende Debatte um sogenannte endogene Präferenz- oder Nutzenänderungen.

Laut v. Weizsäcker (1971), S. 345f., haben die Ökonomen immer anerkannt, daß sich die Nutzenfunktionen der Verbraucher ändern, es aber zumeist nicht als ihr Geschäft angesehen, dies auch theoretisch zu verarbeiten. So gingen die meisten von konstanten Wünschen aus. Aber es scheint die Meinung zunehmend ernst genommen zu werden, daß eine angemessene Wirtschaftstheorie sich zumindest der endogenen Nutzenänderungen anzunehmen habe, d.h. derjenigen Nutzenänderungen, die durch Faktoren aus dem Produktions- und Konsumtionsprozeß selbst ausgelöst werden. Solche Faktoren sind z.B. die Werbung und vor allem der vorausgegangene Verbrauch, wie die Entwicklung von Vorlieben und Süchten zeigt. So soll es Leute geben, die sich immer überkandideltere HiFi-Anlagen erstehen, weil ihr Gehör durch ihre bisherigen Anschaffungen immer besser trainiert wurde. Ein vertrautes Beispiel liefert der Zigarettenkonsum, und besonders drastisch ist es beim Heroin, wonach der Normalbürger kein großes

Verlangen hat, wo man aber schon nach dem ersten Schuß sonst was für den zweiten gäbe. Diese Abhängigkeit vom vergangenen Verbrauch ist auch die am meisten diskutierte, weil anscheinend am einfachsten zu behandelnde, und in der Initialarbeit von Strotz (1955/56) und in den Arbeiten von Pollak (1968), Peleg und Yaari (1973) und Yaari (1977) ist sie in einer zu unserer Problemstellung besonders affinen Weise diskutiert. Sie ist für uns deswegen ein unkomplizierter Spezialfall, weil bei ihr nicht nur Wahrscheinlichkeitsänderungen, sondern überhaupt Wahrscheinlichkeiten nicht in Erscheinung treten. Schauen wir uns also den Fall, wo die Nutzenfunktion einer Person in einer ihr bekannten Weise mit ihrem Verbrauch variiert und sie nun einen optimalen Verbrauchsplan zu entwerfen hat, an, wie er sich nach den genannten Arbeiten darstellt und soweit er für uns relevant ist.

Sei unser *X* diesmal auf dem Wege zum Alkoholiker. *X* möge sich im Besitz von 100 Vierfünftel-Quartflaschen Jack Daniels' Old No.7 befinden, deren Verbrauch er über 10 Intervalle t_1, ..., t_{10} (von je etwa einem Monat) hinweg plant; vorzeitiger Nachschub ist dabei aus irgendwelchen obskuren Gründen, die hier nicht interessieren, ausgeschlossen. Schauen wir uns zunächst an, wie sich *X*s Konsum entwickelt, wenn er in dem Sinne beschränkt ist, daß er den Einfluß seines Konsums auf seine subjektiven Werte für Bourbon nicht berücksichtigt. Anfänglich wird er einen gleichmäßigen Verbrauch von 10 Flaschen pro Intervall für optimal halten. Also wird er während t_1 10 Flaschen verbrauchen. Aufgrund dessen wird er aber eine leichte Ungeduld entwickeln, so daß Konsumtionspläne, die die nähere Zukunft begünstigen, nun bei ihm höher im Kurs stehen. Z.B. mag er nach t_1 einen Verbrauch von zunächst 15, dann 13, 12, 11, 10, 9, 8, 7 und schließlich 5 Flaschen als optimal ansehen. Also wird er sich während t_2 15 Flaschen zu Gemüte führen. Dies verstärkt seinen Drang zum schnellen Verbrauch, so daß er während t_3 20 und während t_4 25 Flaschen leert. Dann bleiben ihm noch 30 Flaschen, und mittlerweile wird ihm der Gedanke an jedwede Verzögerung derart unerträglich sein, daß es (nach t_4) optimal für ihn sein wird, den Rest gleich wegzuputzen. Daran wird er sich in t_5 auch halten, und in den restlichen fünf Intervallen wird er dann wohl oder übel auf dem Trockenen sitzen. Insgesamt ergibt sich also eine ganz unglückliche Entwicklung, die von keinem Standpunkt aus optimal ist, obwohl er zu jedem Zeitpunkt das tut, was von seiner beschränkten Sicht her jeweils optimal ist. Und das Unglück rührt eben genau daher, daß sich seine subjektiven Werte für Whisky in einer von ihm nicht vorhergesehenen Weise ändern.

Welchen Weg wird X einschlagen, wenn er gleich von Anfang an hinreichende Umsichtigkeit aufbringt? Er wird dann rechtzeitig, also vor t_1, in der folgenden Weise rekursiv räsonieren: „Angenommen, ich hätte nach t_9 insgesamt x_9 Flaschen verbraucht. Dann wird mir in Abhängigkeit von x_9 ein Verbrauch von $f_{10}(x_9)$ Flaschen während x_{10} optimal erscheinen.[17] Gehen wir also einen Schritt weiter zurück; nehmen wir an, ich hätte nach t_8 insgesamt x_8 Flaschen verbraucht. Wieviel soll ich nun für t_9 einplanen? Wenn ich während t_9 y_9 Flaschen trinke, dann werde ich, so viel ist schon klar, während t_{10} $f_{10}(x_8 + y_9)$ Flaschen trinken. Ich werde also y_9 so wählen, daß der Verbrauch von erst y_9 und dann $f_{10}(x_8 + y_9)$ Flaschen optimal ist, und zwar von dem Standpunkt aus, den ich nach t_8 nach dem Konsum von x_8 Flaschen innehaben werde. Liefere die Funktion f_9 dieses optimale y_9 in Abhängigkeit von x_8. Nehmen wir für den nächsten Schritt an, ich hätte nach t_7 insgesamt x_7 Flaschen getrunken. Wenn ich dann während t_8 y_8 Flaschen leere, dann werde ich, weil mir das optimal erscheinen wird, während $t_9 f_9(x_7 + y_8)$ und während $t_{10} f_{10}(x_7 + y_8 + f_9(x_7 + y_8))$ Flaschen verbrauchen. Also werde ich y_8 so ansetzen, daß y_8 zusammen mit dieser weiteren Entwicklung gemäß der subjektiven Wertfunktion, die ich nach t_7 habe, optimal ist. $f_8(x_7)$ sei dieses optimale y_8 in Abhängigkeit von x_7. Und so weiter. Im letzten Schritt muß ich mir überlegen, wieviel ich denn nun gleich während t_1 konsumieren soll. Ich habe mir schon ausgerechnet: Wenn ich während t_1 y_1 Flaschen verbrauche, dann wird mir während t_2 ein Verbrauch von $f_2(x_1)$ Flaschen optimal erscheinen; also beträgt mein Konsum während t_2 $f_2(x_1)$ Flaschen. Sodann werde ich während t_3 $f_3(y_1 + f_2(y_1))$ Flaschen verbrauchen, während t_4 $f_4(y_1 + f_2(y_1) + f_3(y_1 + f_2(y_1)))$ Flaschen etc. Ich muß also ein Anfangsquantum wählen, bei dem diese ganze Entwicklung (das Anfangsquantum eingeschlossen) mindestens ebenso gut ist wie die Entwicklungen, die sich bei anderen Anfangsquanten einstellen."

Daran wird sich X dann halten, und wenn er die Änderungen seiner subjektiven Werte richtig vorausgesehen hat, wird er genau auf dem vorausberechneten Weg voranschreiten und dabei in jedem Augenblick das tun, was ihm zu diesem

[17] Es wird dabei die vereinfachende Annahme gemacht, daß die subjektive Wertfunktion von X nach t_9 nur vom bisherigen Gesamtverbrauch x_9 und nicht noch davon abhängt, wie sich x_9 über $t_1, ..., t_9$ verteilt.

Augenblick in seiner Umsichtigkeit optimal erscheint.[18] Dieses Verfahren zur Ermittlung eines optimalen Konsumtionsplans hat erstmals Strotz (1955/56) vorgeschlagen, Pollak (1968) hat es dann ganz präzise formuliert, und Yaari (1977), S. 167-175, schildert es noch einmal ausführlich in sehr durchsichtiger Form.

Peleg und Yaari (1973) weisen noch auf den folgenden Sachverhalt hin: Wenn wir die in der obigen Schilderung enthaltenen Endlichkeitsannahmen aufgeben und zum kontinuierlichen Fall übergehen, um so z.B. auch beliebig fein aufteilbare Güter zu behandeln, so braucht ein optimaler Verbrauchsplan, wie wir ihn oben beschrieben haben, nicht zu existieren. In der Tat konstruieren sie ein Beispiel, wo er nicht existiert, und schlagen für diesen Fall ein andersartiges Optimalitätskriterium vor. Doch wollen wir uns jetzt nicht darum, sondern nach wie vor nur um den endlichen Fall kümmern. Wie man das im endlichen Fall als vernünftig Erkannte möglichst ungebrochen auf den unendlichen Fall übertragen kann, steht auf einem anderen Blatt.

Welches Prinzip steht hinter Xs obigen Überlegungen zu einem optimalen Verbrauchsplan? Ein ganz einfaches: X mögen jetzt die Handlungen $F_1, ..., F_n$ offenstehen. Er sieht voraus, daß er mit der Handlung F_r in die Entscheidungssituation Δ_r gerät. Ferner weiß er schon, daß in Δ_r das weitere Verhalten G_r optimal ist und daher von ihm gewählt werden wird. Also muß er sich nur noch überlegen, welche der Vorgehensweisen $F_r \cap G_r$ er *jetzt* für am besten hält. Dieses Prinzip scheint nicht nur im Fall sich ändernder subjektiver Werte, sondern auch bei Wahrscheinlichkeitsänderungen aufgrund mangelhafter Erinnerungen oder anderer *nicht* auf neuen Erfahrungen beruhender Einflüsse vernünftig zu sein. Z.B. steht jemand, der sich überlegt, wie er zum Oktoberfest gelangen soll, vor der folgenden Alternative: Entweder benutzt er sein eigenes Auto. Auch mit zwei Promille intus wird er sich dann noch für fahrtüchtig halten und demnach mit seinem Auto wieder nach Hause fahren. Oder er läßt sich hinfahren (mit öffentlichen Verkehrsmitteln oder sonst wie) und daher auch wieder zurückfahren. Er wird dann diejenige Möglichkeit wählen, die ihm *vorher* angenehmer erscheint; und dies wird unter anderem davon abhängen, welche Meinung er *jetzt* im nüchternen Zustand über seine Fahrtüchtigkeit bei einem Alkoholpegel von zwei Promille hat.

---

[18] Freilich: Überlegte er sich das alles tatsächlich, so müßten ihm gewisse tiefgründigere Lebensweisheiten noch verborgen geblieben sein.

Es empfiehlt sich also, dieses Prinzip zur Grundlage eines Optimalitätskriteriums für den in diesem Abschnitt diskutierten allgemeinen Rahmen zu machen. Zu mehr als einer Grundlage taugt es freilich nicht; es tun sich da nämlich zwei bisher nicht bedachte Schwierigkeiten auf.

Erstens dürfen wir nicht, wie wir es in unseren hochprozentigen Beispielen taten, annehmen, daß *X* schon genau weiß, in welche Entscheidungssituation er bei welcher Handlung gerät. Vielmehr müssen wir im allgemeinen davon ausgehen, daß *X* lediglich handlungsbedingte Wahrscheinlichkeiten für all die verschiedenen Situationen hat, in die er als nächste kommen zu können glaubt. Welche Strategien für das weitere Vorgehen in diesen Situationen jeweils optimal sind, weiß er schon; das bemißt sich nach den subjektiven Werten und Wahrscheinlichkeiten, die er in diesen Situationen jeweils haben wird. Nun muß er aber gemäß unserem Prinzip all diese Strategien, die ja Teilstrategien der von ihm gesuchten Gesamtstrategie sind, außerdem noch von seinem jetzigen Standpunkt aus bewerten, und zwar in einer quantitativen Weise; denn für die Beurteilung einer Gesamtstrategie, die dann ebenfalls quantitativ ausfällt, ist wesentlich, mit welcher Wahrscheinlichkeit sie in welche wie nützliche Teilstrategie mündet. Wie das präzise zu formulieren ist, lehrt der Whisky noch nicht.

Die zweite Schwierigkeit besteht darin, daß wir Individuen beschreiben wollen, die beliebige Änderungen ihrer Entscheidungssituation ins Auge fassen, d.h. sowohl Änderungen aufgrund zukünftiger Erfahrungen, wie wir sie im letzten Abschnitt diskutierten, als auch Änderungen aus sonstigen Gründen, wie wir sie aus diesem Abschnitt kennen. Ein Vergleich zeigt jedoch, daß bei diesen zwei Sorten von Änderungen völlig verschiedene Überlegungen zum Zuge kommen. Bei Änderungen aufgrund neuer Informationen wird das weitere Vorgehen bereits jetzt schon nach den jeweiligen veränderten subjektiven Werten und Wahrscheinlichkeiten beurteilt. Bei sonstigen Änderungen hingegen muß das weitere Vorgehen auch nach den alten subjektiven Werten und Wahrscheinlichkeiten beurteilt werden. Der Entscheidende muß also je nachdem, welcher Sorte von Änderung er entgegensieht, die eine oder andere Überlegung anstellen. Doch schlimmer noch. Es können ja in einem Schritt Änderungen beider Sorten gleichzeitig enthalten sein; der Entscheidende kann z.B. in einem Zuge eine Beobachtung machen sowie etwas vergessen, und er kann auch mit beidem rechnen. Welche der beiden Überlegungen sollte er dann anwenden?

Der folgende Vorschlag scheint mir geeignet zu sein, all dieser Schwierigkeiten Herr zu werden. Sein Grundgedanke ist, daß in ein Entscheidungsmodell, das all diese Komplikationen repräsentieren soll, dreierlei hineinzustecken ist: erstens die verschiedenen tatsächlichen zukünftigen Entscheidungssituationen, mit denen der Entscheidende aufgrund wer weiß was für Änderungen rechnet; zweitens, wie diese Entscheidungssituationen aussähen, wenn allein neue Erfahrungen für Änderungen sorgten; und drittens sozusagen eine Übersetzung der vorausgesehenen tatsächlichen Situationen in diese hypothetischen, von Erfahrungsfremdem bereinigten Situationen, die insbesondere sagt, welcher Strategie für eine tatsächliche Situation welche Strategie für die zugehörige hypothetische Situation entspricht.

Dem letztgenannten Punkt dient die folgende vorbereitende

Definition 31: Seien $\Delta_0 = \langle I, I^h, i, T, \leq, s, P, V, U\rangle$ und $\Delta'_0 = \langle I, I^h, i', T, \leq, s, P', V', U'\rangle$ E2-Modelle für t mit $s_i = s_{i'}$.

(a) Dann heißt $\langle f_1, f_2\rangle$ ein *Isomorphismus von* Δ_0 *nach* Δ'_0 genau dann, wenn f_1 eine umkehrbar eindeutige Abbildung von i auf i' und f_2 eine Funktion auf i ist derart, daß für jedes $\Delta \in i$ $f_2(\Delta)$ ein Isomorphismus von Δ nach $f_1(\Delta)$ ist.

(b) Sei nun $\langle f_1, f_2\rangle$ ein Isomorphismus von Δ_0 nach Δ'_0 und S eine Strategie für Δ_0. Dann heißt S' *die* S *gemäß* $\langle f_1, f_2\rangle$ *korrespondierende Strategie für* Δ'_0 genau dann, wenn $S'(\Delta'_0) = S(\Delta_0)$ und wenn für jedes $\Delta \in i$ $S'(f_1(\Delta))$ die $S(\Delta)$ gemäß $f_2(\Delta)$ korrespondierende Strategie für $f_1(\Delta)$ ist.

Mit dieser aus den bekannten Gründen korrekten rekursiven Definition ist unser „Übersetzungsproblem" gelöst. E2-Modelle, zwischen denen ein Isomorphismus besteht, sind völlig gleichartig aufgebaut, und Strategien, die einander über diesen Isomorphismus entsprechen, beschreiben innerhalb dieses Aufbaus genau die gleichen Wege. Der geschilderte Grundgedanke läßt sich dann folgendermaßen realisieren:

Definition 32: $\Delta_0 = \langle I, I^h, i, T, \leq, s, P, V, U, f_1, f_2\rangle$ ist ein *E4-Modell für t* genau dann, wenn gilt:

(1) Δ_0 ist ein E2-Modell für t,[19]

---

[19] Ich bitte, über die kleine Schlamperei hinwegzusehen, daß in der ursprünglichen Definition von E2-Modellen f_1 und f_2 keinen Platz hatten; ein Mißverständnis kann dadurch nicht entstehen.

(2) wenn $i \neq \varnothing$, dann ist i eine Menge von E4-Modellen für s_i,

(3) wenn $i' = \{f_1(\Delta) \mid \Delta \in i\}$ und P' die P entsprechende Wahrscheinlichkeitsfamilie für $\langle I \cup \{i'\rangle, I^h\rangle$ bzw. $\langle I, I^h\rangle$ ist[20], dann ist $\Delta_0' = \langle I, I^h, i', T, \leq, s, P', V, U\rangle$ ein E3-Modell für t und $\langle f_1, f_2\rangle$ ein Isomorphismus von Δ_0 nach Δ_0'.

Δ_0' heißt dann auch das *Erfahrungsdestillat* von Δ_0 und $\langle f_1, f_2\rangle$ sein *Erzeuger*.

Wie üblich ist dies eine korrekte rekursive Definition. Ein E4-Modell Δ_0 ist also, wie gewünscht, erstens ein E2-Modell, in dem beliebige zukünftige Entwicklungen beschrieben werden können, und enthält zweitens eine Vorrichtung, die ein zu Δ_0 völlig strukturgleiches E3-Modell Δ_0' entwirft, das diejenige Situation beschreiben soll, die vorläge, wenn von all den in Δ_0 ins Auge gefaßten Änderungen nur diejenigen zur Geltung kämen, die von neuen Erfahrungen herrühren.

Dies führt uns zur letzten

Definition 33: Sei Δ_0 ein E4-Modell für t, i seine Rekursionsquelle, Δ_0' sein Erfahrungsdestillat und $\langle f_1, f_2\rangle$ dessen Erzeuger. Dann heißt eine Strategie S_0 für Δ_0 *optimal* genau dann, wenn für jedes $\Delta \in i$ $S_0(\Delta)$ eine optimale Strategie für Δ ist und wenn für alle Strategien S für Δ_0 gilt: wenn für jedes $\Delta \in i$ $S(\Delta)$ eine optimale Strategie für Δ ist, wenn S_0' bzw. S' die S_0 bzw. S gemäß $\langle f_1, f_2\rangle$ korrespondierende Strategie für Δ_0' ist und wenn U^* die erwartete subjektive Wertfunktion für Strategien für Δ_0' ist, dann gilt $U^*(S_0') \geq U^*(S')$.

Wiederum ist dies eine korrekte rekursive Definition. Und sie faßt die Überlegungen dieses Abschnitts zusammen: Zunächst können nur solche Strategien in Betracht gezogen werden, deren Teilstrategien vom jeweiligen späteren Standpunkt aus optimal sind, da nur solche Teilstrategien weiterverfolgt werden; und aus all diesen Strategien mit optimalen Teilstrategien ist dann eine vom ursprünglichen Standpunkt aus beste herauszusuchen.

Haben wir nun aber doch nicht bloß die eine Überlegung berücksichtigt, die bei nicht erfahrungsbedingten Änderungen der Entscheidungssituation angemessen ist? Nein; die andere Überlegung, die bei erfahrungsbedingten Änderungen am Platze ist, geht ja in die erwartete subjektive Wertfunktion für Strategien für

[20] D.h. genauer: wenn $i = \varnothing$, so $P' = P$, und wenn $i \neq \varnothing$, so soll für alle $A \in \mathcal{A}(I)$. $F \in \mathcal{Z}(I^h)$ und $\Delta \in i$ $P_F'(A) = P_F(A)$ und $P_F'(f_1(\Delta)) = P_F(\Delta)$ gelten.

das zugehörige Erfahrungsdestillat und damit in die Beurteilung der Strategien vom anfänglichen Standpunkt aus ein. Dies zeigt sich auch, wenn wir den Fall, in dem der Entscheidende nur zukünftige Erfahrungen bedenkt, statt mit einem E3-Modell Δ'_0 etwas umständlich mit einem E4-Modell Δ_0 beschreiben, das mit seinem Erfahrungsdestillat Δ'_0 identisch ist. Dann fällt nämlich das Optimalitätskriterium für E4-Modelle mit dem für E3-Modelle zusammen, d.h. eine Strategie für Δ_0 ist in diesem Fall genau dann optimal, wenn die korrespondierende Strategie für Δ'_0 (unter *allen* Strategien für Δ'_0) maximalen erwarteten subjektiven Wert hat.

Formulieren wir der Vollständigkeit halber noch die mit E4-Modellen verknüpfte empirische Behauptung, die besagt: Werde X zu t durch das E4-Modell Δ_0 für t charakterisiert. Dann fängt X an, eine optimale Strategie für Δ_0 durchzuführen. Natürlich können wir nur sagen, daß er damit anfängt; die tatsächliche Entwicklung seiner Situation kann ja ganz anders sein, als er sie sich vorgestellt hat.

Nun, da sich dieses Kapitel seinem Ende zuneigt, ist einzugestehen, daß im Grunde zweierlei schiefgelaufen ist. Erstens haben wir uns mit E4-Modellen zu einer Allgemeinheit aufgeschwungen, die irgendwo schon nicht mehr sinnvoll ist. Denn in einem E4-Modell und in der Bestimmung von optimalen Strategien dafür sind so viele Rekursionen versteckt, daß schon bei vier Planungsstufen kein Mensch mehr die in einem E4-Modell gebotenen Möglichkeiten voll ausschöpfen dürfte.[21] Der Vorteil der Verallgemeinerung ist so eher ein technischer. Doch nutzen wir diesen nicht mehr.

Denn zweitens wollen wir es bei der gelieferten Explikation und den daraus resultierenden Definitionen belassen, obwohl diese doch eigentlich nur die Präliminarien für eine Theorie mit vielen Sätzen und Theoremen abgeben sollten. Aber mir kam es in diesem Kapitel vor allem darauf an, deutlich zu machen: daß hinter strategischem Denken nichts weiter als eine Selbstanwendung der Entscheidungstheorie steckt; wie sich damit recht Unterschiedliches unter einen begrifflichen Hut bringen läßt; daß zwei verschiedene Optimierungsüberlegungen

---

[21] Allgemein gilt – dies als Schmankerl zum Nachrechnen: Wenn mit einem E4-Modell n Planungsstufen bewältigt werden sollen, wobei jede Rekursionsquelle gleich viele, nämlich $m \geq 2$ Elemente enthalte, so ist es insgesamt aus $q = \dfrac{n \cdot m^n - 1}{m-1} - \dfrac{m^{n-1} - 1}{(m-1)^2}$ Entscheidungsmodellen aufgebaut. Für $n = 1$ ergibt sich also $q = 1$, wie nicht anders zu erwarten; und für $n = 4$ und $m = 2$, also bei vier Planungsstufen und jeweils minimaler Rekursionsquelle, ergibt sich bereits $q = 56$.

anzustellen sind, je nachdem, ob die erwogenen Änderungen der Entscheidungssituation erfahrungsbedingt sind oder nicht; und wie diese beiden Überlegungen zusammenspielen.

Über beides mag vielleicht die Tatsache hinwegzutrösten, daß es viele Spezialisierungen von E4-Modellen gibt, die alle sinnvolle Anwendungen finden und die teilweise schon sehr gründlich untersucht worden sind. Letzteres bezieht sich natürlich vor allem auf E3-Modelle, in denen es als sicher angenommen wird, daß dieser oder jener Spielraum beobachtet wird, und in denen diese Erfahrungen dann gemäß einfacher Konditionalisierung verarbeitet werden. Diese bilden ja auch nach wie vor den wichtigsten Spezialfall.

Statt Theoremen will ich noch zwei Beobachtungen nachtragen. Erstens ist es aufgrund von Definition 33 recht trivial, daß es gut ist, sich korrekt zu erinnern. (Das kann man auch als Bestätigung für Definition 33 auffassen.) Denn sei Δ_0 ein E4-Modell und bestehe seine Rekursionsquelle aus Δ_1 und Δ_2. Der einzige Unterschied zwischen Δ_1 und Δ_2 sei dabei, daß sich in Δ_1 gegenüber Δ_0 nichts geändert hat, daß sich also der Entscheidende insbesondere in Δ_1 aller Dinge aus Δ_0 korrekt erinnert, während dies in Δ_2 nicht der Fall sei. Dann ist eine optimale Strategie für Δ_1 auch schon vom Standpunkt von Δ_0 aus optimal, während eine optimale Strategie für Δ_2 von Δ_0 aus betrachtet in der Regel nicht optimal sein wird. Deshalb wird eine Strategie für Δ_0 umso besser sein, je eher sie zu Δ_1 statt zu Δ_2 führt. (Die Sachlage ändert sich natürlich, wenn die Erinnerungsbemühungen erhebliche Kosten verursachen.)

Zweitens möchte ich noch einmal kurz auf die Zerlegung und die Reduktion von Entscheidungsmodellen eingehen, nachdem diese ja einen wesentlichen Gegenstand von Kapitel 3 bildeten. Die verschiedenen Dinge, zu denen sie gut waren, ließen sich bereits hinreichend an Hand von E1-Modellen demonstrieren. Dennoch liegt natürlich die Frage nahe, wie es damit in Bezug auf E4-Modelle aussieht. Nun, es sieht gut aus, auch wenn ich das jetzt nicht mehr im technischen Detail darlegen möchte.

Was zunächst die Zerlegung betrifft: Sei Δ_0 ein E4-Modell. Mit Δ_0 ist eine Vielzahl weiterer Entscheidungsmodelle verknüpft: sein Erfahrungsdestillat, die Modelle in seiner Rekursionsquelle, deren Erfahrungsdestillate, die Modelle in den Rekursionsquellen der jeweiligen Erfahrungsdestillate etc. Allen diesen Modellen liegt dieselbe Spielraummenge I zugrunde, und jedes dieser Modelle induziert $\perp_{cw}$-Zerlegungen von I. Sei nun $\{J_1, ..., J_n\}$ eine Zerlegung von I, die bezüglich jedes dieser Modelle eine $\perp_{cw}$-Zerlegung von I ist. Dann läßt sich, das will ich hier lediglich behaupten, Δ_0 in eindeutiger Weise in die E4-Modelle Δ_1, ..., Δ_n zerlegen derart, daß jedes Δ_r über der Spielraummenge J_r aufgebaut ist (r = 1, ..., n), Δ_0 sich aus Δ_1, ..., Δ_n wieder zurückgewinnen läßt und die optimalen Strategien für Δ_0 sich gerade aus den optimalen Strategien für Δ_1, ..., Δ_n zusammensetzen lassen; d.h. es gelten die Analoga zu den Sätzen 17 und 18 (S. 137f.).

Die Reduktion von E4-Modellen ist auch nicht problematischer: Sei $\Delta_0 = \langle I, I^h, i, T, \leq, s, P, V, U, f_1, f_2 \rangle$ ein E4-Modell für t, das wir nun um $K \subseteq I^g$ reduzieren wollen. Das ist nicht schwer; wir brauchen nur der üblichen rekursiven Prozedur zu folgen. D.h. wir können voraussetzen, daß wir die Modelle in i bereits um K reduziert haben – die reduzierten Modelle mögen die Menge $\tilde{i}$ bilden – und ebenso die Modelle $f_1(\Delta)$ ($\Delta \in i$). Dann sind natürlich die Funktionen f_1 und f_2 durch entsprechende Funktionen $\tilde{f}_1$ und $\tilde{f}_2$ auf $\tilde{i}$ zu ersetzen. Und schließlich ist Δ_0 selbst wie in Definition 26 (S. 144) zu reduzieren, wobei P, V und U in $\tilde{P}$, $\tilde{V}$ und $\tilde{U}$ übergehen möge. Dann läßt sich $\tilde{\Delta}_0 = \langle I, I^h, \tilde{i}, T, \leq, s, \tilde{P}, \tilde{V}, \tilde{U}, \tilde{f}_1, \tilde{f}_2 \rangle$ als ein K-Redukt von Δ_0 definieren, und es gilt das Analogon zu Satz 19 (S. 145).

Es ist dabei hervorzuheben, daß ohne Jeffreys verallgemeinerte Konditionalisierung die Reduktion speziell von E3-Modellen, in denen neue Informationen nur gemäß der einfachen Konditionalisierung verarbeitet werden, nicht möglich wäre. Dies zeigt das folgende Beispiel: In einem E3-Modell Δ_0 tauche ein Spielraum j auf, der aus den verschiedenen Zeigerstellungen eines Meßinstrumentes besteht, das nicht absolut zuverlässige, aber doch beherzigenswerte Aussagen über einen Spielraum k, z.B. die Temperatur im Erdinnern, trifft. Außerdem sei in Δ_0 die Möglichkeit erwogen, j zu beobachten, d.h. zum rechten Moment ein genaues Auge für das Meßinstrument zu haben, um daraus mittels einfacher Konditionalisierung Aufschluß über k zu gewinnen. Wenn nun Δ_0 um $\{j\}$ zum Modell Δ_0^* reduziert wird, so kann man über Δ_0^* nicht sagen, daß darin die Beobachtungen von j in Betracht gezogen wird. In Δ_0^* stellt sich der Sachverhalt vielmehr so dar, daß direkt der Spielraum k beobachtet wird – und zwar in unscharfer Weise, da das Meßinstrument nicht ganz zuverlässig ist – und daß sich daraus direkt Wahrscheinlichkeitsänderungen gemäß der verallgemeinerten Konditionalisierung ergeben. Insofern ist die verallgemeinerte Konditionalisierung eine notwendige theoretische Ergänzung zur einfachen Konditionalisierung. Dies als Nachtrag zur Diskussion auf S. 161f.

KAPITEL 5

Nachträge

5.1 Newcombs Problem

Newcombs Problem, der Öffentlichkeit erstmals von Nozick (1969) vorgestellt, hat unter Entscheidungstheoretikern und anderen Leuten eine veritable Verwirrung gestiftet. Es ist nicht nur ein entscheidungstheoretisches, sondern ebenso ein den Kausalbegriff betreffendes Problem, das zwei miteinander unverträgliche, aber anscheinend gleichermaßen überzeugend begründbare Lösungen hat (dies ist das eigentliche Problem). Da sich aus unseren bisherigen Ausführungen unmittelbar ergibt, daß eine der beiden Lösungen der Grundlage entbehrt, will ich hier kurz darauf eingehen:

Sie, der geschätzte Leser, stehen vor zwei geschlossenen Urnen a_1 und a_2. Die erste, a_1, enthält, das wissen Sie, 1 TDM (in Worten: eintausend Deutsche Mark); die zweite, a_2, enthält entweder 1000 TDM oder 0 TDM, und Sie wissen nicht, was. Sie haben die Wahl zwischen zwei Handlungen: der Handlung F_1, *nur* die zweite Urne samt Inhalt zu nehmen, und der Handlung F_2, *beide* Urnen samt deren Inhalt zu nehmen.

So geschildert, scheint völlig klar, was zu tun ist. Bei F_2 fahren Sie in jedem Fall besser, haben Sie in jedem Fall 1 TDM mehr als bei F_1 – egal ob a_2 nun 1000 TDM oder 0 TDM enthält; Ihre Wahrscheinlichkeiten dafür, daß a_2 1000 TDM bzw. 0 TDM enthält, scheinen dabei überhaupt keine Rolle zu spielen. Technisch ausgedrückt, F_2 dominiert F_1, und deshalb sollten Sie F_2 wählen.

Die Geschichte geht aber noch weiter. Es gibt da nämlich ein Wesen, das Handlungen von Menschen allgemein und speziell von Ihnen mit großer Zuverlässigkeit voraussagen kann. Sie selbst haben schon sehr viele dieser Voraussagen, insbesondere in Bezug auf das Verhalten von Leuten in der geschilderten Situation, kontrolliert, und sie trafen praktisch immer zu. (Wie dieses Wesen zu dieser wunderlichen Fähigkeit kommt, tut dabei nichts zur Sache.) Nun wissen Sie ferner, daß das Wesen bereits eine Voraussage darüber gemacht hat, ob Sie F_1 oder F_2 wählen, und Sie wissen Folgendes: Wenn es vorausgesagt hat, daß Sie beide Urnen nehmen, legt es nichts in die zweite Urne, so daß sie 0 TDM enthält; und wenn es vorausgesagt hat, daß Sie nur die zweite Urne nehmen, legt es 1000 TDM in die zweite Urne. Wollen Sie nun mit all diesem Wissen im Hintergrund an der obigen Überlegung festhalten und immer noch F_2 wählen?

Oder sollten Sie sich nicht vielmehr sagen: Wenn ich beide Urnen nehme, so wird das Wesen dies aller Wahrscheinlichkeit nach vorhergesagt und nichts in a_2 gelegt haben; in diesem Fall erhalte ich also fast sicher nur die 1 TDM aus a_1. Wenn ich hingegen nur die zweite Urne nehme, so wird das Wesen auch dies mit großer Sicherheit vorhergesagt und dementsprechend a_2 mit 1000 TDM gefüllt haben; in diesem Fall bekomme ich also fast sicher diese 1000 TDM. Der erwartete subjektive Wert von F_1 ist somit entschieden größer als der von F_2, und daher sollte ich natürlich F_1 wählen.

Doch wenn dies richtig ist, was war dann an der ersten Überlegung falsch? Malen wir, um das Dilemma noch bedrängender zu gestalten, beide Überlegungen mit Nozicks eigenen Worten aus.

Zur zweiten Überlegung: „Sie wissen, daß viele Personen dieses Experiment durchlaufen haben. All jene, die nur die zweite Urne nahmen, auch jene, die die erste Überlegung kannten, ihr aber nicht folgten, kamen mit 1000 TDM heraus. Sie wissen auch, daß all die Schlauköpfe, die der ersten Überlegung folgend beide Urnen nahmen, es nur zu 1 TDM brachten. Sie haben keinen Grund anzunehmen, daß Sie sich, was Ihre Vorhersagbarkeit anbetrifft, von all den anderen unterscheiden. Da Sie weiterhin wissen, daß ich all das bisher Gesagte weiß, wissen Sie auch, daß ich mit einem hohen Satz darauf wetten würde – und mich dabei rational verhalten würde –, daß Sie nur 1 TDM erhielten, wenn Sie beide Urnen nähmen. Wenn Sie nun doch unwiderruflich beide Urnen nähmen und das Ergebnis aber noch ein Weilchen geheim gehalten würde, wäre es dann nicht für Sie rational, inzwischen mit einer dritten Seite zu hohen Sätzen darauf zu wetten, daß Sie nur 1 TDM bekommen? Wenn Sie hingegen nur die zweite Urne nähmen, wäre es dann nicht rational für Sie, mit einer dritten Seite eine Nebenwette einzugehen, daß Sie 1000 TDM bekommen? Das alles ist Ihnen klar (auch wenn tatsächlich niemand zur Hand ist, mit dem Sie wetten könnten). Wollen Sie dann wirklich beide Urnen nehmen und damit dem zuwiderhandeln, worauf Sie rationalerweise gerne wetten würden?“ (Nozick (1969), S. 115f.)

Zur ersten Überlegung: „Das Wesen hat seine Vorhersage bereits getroffen, je nachdem die 1000 TDM in die zweite Urne gelegt oder nicht, und ist dann weggegangen. Dies mag vor einer Woche oder vor einem Jahr geschehen sein. Die Urne a_1 ist durchsichtig. Sie können die 1 TDM darin sehen. Die 1000 TDM sind bereits entweder in der Urne a_2 oder auch nicht (auch wenn Sie nicht sehen können, was der Fall ist). Wollen Sie wirklich nur a_2 nehmen? Um die Sache noch deutlicher zu machen: Sie können von Ihrer Seite aus nicht in a_2 hineinsehen, aber von der anderen Seite aus möge a_2 durchsichtig sein. Ich sitze auf der anderen Seite von a_2, schaue in sie hinein und sehe, was da ist. Ich sitze da schon die ganze Woche. Entweder schaue ich seit einer Woche auf 1000 TDM oder ich sehe seit einer Woche eine leere Urne. Wenn das Geld schon da ist, so bleibt es auch da, was immer Sie tun. Es wird nicht verschwinden. Wenn es noch nicht da ist, wenn ich auf eine leere Urne schaue, so erscheint es auch nicht plötzlich, wenn Sie nur die zweite Urne wählen. Wollen Sie also nur die zweite Urne nehmen und sich die zusätzliche 1 TDM entgehen lassen, die Sie deutlich sehen können? Ich sitze da, meinen Blick auf den Urnen, und hoffe, daß Sie eine bestimmte Handlung durchführen. Innerlich gebe ich Ihnen einen Rat. Und natürlich wissen Sie bereits, welchen Rat ich Ihnen innerlich gebe. In jedem Fall, ob ich 1000 TDM in der zweiten Urne sehe oder nicht, hoffe ich, daß Sie beide Urnen nehmen werden. Werden Sie da nur die zweite Urne nehmen, sich die deutlich sichtbare, zusätzliche 1 TDM entgehen lassen und meine stumme Hoffnung ignorieren, daß Sie beide Urnen nehmen mögen? Natürlich kommt es auf meine Gegenwart nicht an. Sie sitzen da allein, aber Sie wissen: Wenn da ein Freund, dem Ihr Wohl am Herzen liegt, säße und von der anderen Seite aus in die beiden Urnen schaute, so hoffte er, daß Sie beide Urnen nehmen. Wollen Sie wirklich nur die zweite Urne nehmen und sich die deutlich sichtbare, zusätzliche 1 TDM entgehen lassen?“ (Nozick (1969), S. 116f.) Und, wie ich hinzufügen möchte, würden Sie sich, falls Sie nun doch nur die zweite Urne genommen haben sollten, nicht ärgern, die 1 TDM aus der ersten Urne liegengelassen zu haben – was immer Sie in der zweiten Urne finden? Sie hätten dort ja genau dasselbe gefunden, wenn Sie zusätzlich noch die erste Urne genommen hätten.

Das Problem ist weniger, für sich eine Entscheidung in dieser hypothetischen Situation zu treffen; da sind die meisten Leute ohne großes Zögern einer recht dezidierten Meinung. Das Problem ist vielmehr, daß die Leute recht dezidiert entgegengesetzter Meinung sind, daß man sich also nicht nur von der einen Überlegung überzeugen lassen darf, sondern auch noch überzeugende Gründe dafür finden sollte, wieso die andere Überlegung doch nicht so stichhaltig ist, wie sie scheint.

Es ist wichtig, sich klar zu machen, daß nur bei ungefähr den genannten Zahlen bzw. Zahlenverhältnissen beide Entscheidungen prima facie ähnlich plausibel sind und daß man die Befürworter sowohl der einen wie der anderen Entscheidung in Verlegenheit bringen kann, wenn man die Zahlen anders wählt. So war es, wie Nozick (1969), S. 140f., zu Recht betont, erstens wesentlich, daß wir davon sprachen, daß jenes Wesen *fast sicher* die richtige Voraussage trifft, daß also die bedingte Wahrscheinlichkeit dafür, daß das Wesen F_n vorhergesagt hat, sofern F_n ausgeführt wird (n = 1, 2), irgendwo zwischen, sagen wir, 0,95 und 0,99 liegt.[1] Der Befürworter von F_1 kommt in Schwierigkeiten, wenn jenes Wesen nur noch etwas häufiger richtige als falsche Voraussagen macht, wenn man also diese Wahrscheinlichkeit z.B. auf 0,6 oder gar auf 0,51 senkt; nach der zweiten Überlegung müßte er auch dann nur die zweite Urne nehmen.[2] Dies erscheint absurd – kann es doch jeder halbwegs intelligente Mensch bei etwas Training zu Voraussagen dieser geringen Oualität bringen. Umgekehrt führt man den Befürworter von F_2, wie es Nozick ist, aufs Glatteis, wenn man diese Wahrscheinlichkeit noch erhöht und insbesondere wenn man diesem Wesen Unfehlbarkeit zuschreibt. Dann immer noch beide Urnen zu nehmen, ist zumindest sehr mutig. Und zweitens war es wesentlich, den festen Betrag in a_1 gegenüber dem möglichen in a_2 klein, aber (für Leute normalen Einkommens) nicht vernachlässigbar zu wählen. Erhöbt man ihn etwa von 1 TDM auf 200 TDM, so müßte der Befürworter von F_1 immer noch F_1 wählen,[2] was wiederum bedenklich anmutet. Und erniedrigt man ihn auf 5 DM, so gehört beim Befürworter von F_2 schon eine Portion Dickköpfigkeit dazu, dann noch an seiner Wahl festzuhalten.

---

[1] Levi (1975) weist zu Recht darauf hin, daß *diese* Wahrscheinlichkeiten spezifiziert sein müssen und daß es für die Entscheidung nicht viel hülfe, wenn stattdessen die bedingten (statistischen) Wahrscheinlichkeiten dafür bekannt wären, daß F_n ausgeführt wird, sofern das Wesen F_n vorhergesagt hat (n = 1, 2). Wieso er Nozick vorwirft, dies übersehen zu haben, ist mir allerdings nicht klar.

[2] Sofern seine subjektive Wertfunktion für Geld nicht allzu stark abflacht.

Ich für meinen Teil bin kompromißloser Anhänger davon, beide Urnen zu nehmen. Im Prinzip würde ich selbst dann beide Urnen nehmen, wenn in der ersten Urne nur ein Groschen wäre und wenn ich mit der mir größtmöglichen Gewißheit von der Richtigkeit der Voraussage jenes Wesens überzeugt wäre.[3] Freilich – wenn ich tatsächlich in die glückliche Lage käme, diese Wahl zu haben, würde ich vielleicht doch nur die zweite Urne nehmen. Dies ist kein Widerspruch. Darin kommt lediglich zum Ausdruck, daß ich mir, wenn ich nun tatsächlich vor den Urnen stünde, der Voraussetzungen der Geschichte nicht so sicher sein könnte, wie es hier angenommen wurde. Mein Argwohn, daß da ein ganz fauler Trick dahintersteckt, wäre vermutlich nicht zu besänftigen. Vielleicht würde ich auch, um die Rätsel, die mir jenes Wesen aufgibt, in mein Weltbild einzupassen, einfach meinen Sinnen nicht mehr trauen oder zu phantasieren anfangen, daß die Zeit zuweilen rückwärts liefe, daß jeder Körper seine eigene Zeit habe oder was da sonst noch alles in Frage käme. Aber, wie gesagt, die Voraussetzungen der Geschichte sind so, daß all dies auszuschließen ist, und unter diesen Voraussetzungen würde ich in jedem Fall beide Urnen nehmen.

Im folgenden geht es mir weniger darum, andere zu meiner Meinung zu bekehren oder den Unbelehrbaren irgendwelche Denkfehler nachzuweisen. Vielmehr interessiert mich die eher theoretische, nicht-normative Frage, wie man die jeweilige Neigung mit einer kohärenten Entscheidungstheorie in Einklang bringen kann. Der Befürworter von F_1 scheint da keine Schwierigkeiten zu haben; wie sonst auch maximiert er den erwarteten Nutzen. Davon scheint der Befürworter von F_2 abzuweichen, was eine Erklärung von ihm fordert. Tatsächlich ist es, glaube ich, gerade umgekehrt. Aber hören wir erst einige andere Meinungen zu diesem Thema.

Zunächst die Meinung Nozicks. Wie sich in den Zitaten schon andeutete, sieht er hier zwei verschiedene entscheidungstheoretische Prinzipien im Widerstreit. Er geht dabei von der Repräsentation einer Entscheidungssituation durch das Grundmodell, also mit Hilfe von Handlungen, Umständen, Konsequenzen, subjektiven Wahrscheinlichkeiten für Umstände und subjektiven Werten für Konsequenzen aus und läßt aber zu, daß diese Wahrscheinlichkeiten handlungsabhän-

[3] Mit Absicht wähle ich diese Formulierung. Es scheint mir nämlich sehr fraglich, ob ich jemals von der Unfehlbarkeit jenes Wesens überzeugt sein könnte; vermutlich würde ich eher an meinem Verstande zweifeln.

gig sind.[4] Eine Handlung heißt dann *dominant*, wenn sie unter jedem Umstand eine mindestens ebenso gute Konsequenz zeitigt wie die anderen möglichen Handlungen und unter mindestens einem Umstand zu einer besseren Konsequenz führt als alle anderen Handlungen. Das *Dominanzprinzip* besagt dann, daß man eine dominante Handlung wählen soll, sofern sie existiert. Nach Nozick hält sich dann derjenige, der beide Urnen nimmt, an das Dominanzprinzip, während derjenige, der nur die zweite Urne nimmt, das Bayessche Prinzip von der Maximierung des erwarteten subjektiven Wertes befolgt.

Konsequenterweise versucht er dann zu ermitteln, welches Prinzip in welchem Fall angemessen ist. Er unterscheidet dabei zwei mal drei Fälle: Zunächst die Fälle, (a) daß eine dominante Handlung existiert, und (b) daß keine dominante Handlung existiert. Und dann die drei Fälle, (1) daß die Handlungen kausalen Einfluß auf die Umstände haben und die handlungsbedingten Wahrscheinlichkeiten für die Umstände je nach Handlung verschieden ausfallen, (2) daß die Handlungen keinen kausalen Einfluß auf die Umstände haben und die handlungsbedingten Wahrscheinlichkeiten für die Umstände dennoch verschieden sind, und (3) daß die Handlungen keinen kausalen Einfluß auf die Umstände haben und die handlungsbedingten Wahrscheinlichkeiten für die Umstände gleich sind.[5] Dabei haben die Handlungen insbesondere dann keinen Einfluß auf die Umstände, wenn die Umstände sich zeitlich vor den Handlungen realisieren, wie es bei Newcombs Problem der Fall ist.

Welches Prinzip gilt nun in welchem Fall? Im Fall (3a) fallen das Dominanzprinzip und das Bayessche Prinzip zusammen; beide sind angemessen. Im Fall (3b) gilt das Bayessche Prinzip. Ferner ist klar, daß das Dominanzprinzip im Fall (1a) inadäquat ist.[6] So ist im Fall (1a) und erst recht im Fall (1b) das Bayessche Prinzip zu empfehlen. Kritisch ist der Fall (2a), der bei Newcombs Problem gegeben scheint. Hier kommt Nozick nach sorgfältiger Diskussion zu der Überzeugung, daß nur das Dominanzprinzip und nicht das Bayessche Prinzip angemessen ist. Im Fall (2b) ist natürlich nicht einmal das Dominanzprinzip anwendbar;

---

[4] Daß dann unserer Argumentation aus Abschnitt 2.2 (S. 50) zufolge die Unterscheidung zwischen Umständen und Konsequenzen nicht mehr sonderlich sinnvoll ist, braucht uns jetzt nicht zu stören.

[5] Den vierten Fall, daß die Handlungen (versteckten) Einfluß auf die Umstände ausüben, die handlungsbedingten Wahrscheinlichkeiten für die Umstände aber dennoch gleich sind, erwähnt Nozick nicht; er scheint ihn nicht für möglich zu halten.

[6] Vgl. etwa Jeffrey (1965), Abschnitt 1.5.

Nozick schlägt stattdessen ein etwas komplizierteres Prinzip vor, das nun nicht interessiert.

Dementsprechend zieht es Nozick in der von Newcomb entworfenen Situation vor, beide Urnen zu nehmen. Freilich bleibt Nozick, wie erwähnt, nicht ganz konsequent bei dieser Lösung; in dem Fall, in dem jenem merkwürdigen Wesen Unfehlbarkeit zuzuschreiben ist, weicht er von ihr wieder ab. Mir mißfällt, daß die Entscheidungstheorie mit dieser Lösung ihre Einheitlichkeit einbüßt und in verschiedene, für verschiedene Fälle zuständige Entscheidungsregeln zersplittert. Dies ist nicht nur unelegant, sondern verlangt vor allem noch nach weiterer Erklärung; es ist nicht damit getan, daß man in jedem Fall die jeweilige Regel für überzeugend hält.

Levi (1975) mißfällt etwas anderes. Er hält die Gegenüberstellung des Bayesschen und des Dominanzprinzips für irreführend. Davon ausgehend, dass in jedem Fall sowohl handlungsbedingte wie nicht-bedingte, absolute Wahrscheinlichkeiten für die Umstände vorliegen, sieht er das Problem vielmehr darin, ob das Bayessche Prinzip in Bezug auf die handlungsbedingten Wahrscheinlichkeiten oder in Bezug auf die absoluten Wahrscheinlichkeiten anzuwenden ist. Diese Sicht mag vielleicht hilfreich sein; aber es ist festzustellen, wenn schon nicht einzuwenden, daß die theoretische Einheitlichkeit damit nicht wiederhergestellt wird.

Der entschieden am anregendste Beitrag zu Newcombs Problem ist neben dem von Nozick (1969) freilich der von Gibbard und Harper (1976). Ihr Grundgedanke ist es, das subjunktive Konditional „$\Box\!\!\rightarrow$“ zur Repräsentation von Entscheidungssituationen einzusetzen. Für eine Handlung f und eine Konsequenz c heißt „$f \Box\!\!\rightarrow c$“ dabei soviel wie „wenn f ausgeführt würde, so würde c eintreten“. Den erwarteten subjektiven Wert einer Handlung f hatten wir bisher – mit C als Menge möglicher Konsequenzen – definiert als

$$U(f) = \sum_{c \in C} V(c) \cdot P(c \mid f).$$

Gibbard und Harper weisen nun auf die Möglichkeit hin, daß man ihn auch so definieren könnte:

$$U'(f) = \sum_{c \in C} V(c) \cdot P(f \Box\!\!\rightarrow c).$$

Begeben wir uns in den Rahmen des leicht liberalisierten Grundmodells mit W als Menge möglicherweise *handlungabhängiger* Umstände (und mit Handlungen als Funktionen von W in C), so ist

$$U(f) = \sum_{w \in W} V(f(w)) \cdot P(w \mid f)$$

und

$$U'(f) = \sum_{w \in W} V(f(w)) \cdot P(f \Box\!\!\rightarrow w).$$

Gilt denn nicht immer $U'(f) = U(f)$? Das hängt davon ab, ob immer $P(f \Box\!\!\rightarrow w) = P(w \mid f)$ gilt. Und das gilt laut Gibbard und Harper zwar in normalen Fällen[7] – deshalb wurde hier bisher kein Unterschied gemacht –, aber eben nicht immer. Gerade die von Newcomb entworfene Situation illustriert dies: Wenn A_n der Umstand ist, daß jenes Wesen vorausgesagt hat, daß F_n ausgeführt wird ($n = 1, 2$), so gilt $P(A_1 \mid F_1) > P(A_1 \mid F_2)$ und entsprechend $P(A_2 \mid F_2) > P(A_2 \mid F_1)$ (wobei sich die Differenz danach bemißt, für wie zuverlässig der Entscheidende jenes Wesen hält) und dennoch $P(F_1 \Box\!\!\rightarrow A_1) = P(F_2 \Box\!\!\rightarrow A_1)$ und $P(F_1 \Box\!\!\rightarrow A_2) = P(F_2 \Box\!\!\rightarrow A_2)$; die Situation war ja so beschrieben, daß $F_1 \Box\!\!\rightarrow A_n$ genau dann, wenn $F_2 \Box\!\!\rightarrow A_n$ ($n = 1, 2$), und daß der Entscheidende dies auch weiß. Wer U, den erwarteten subjektiven Wert bisher bekannter Art, maximiert, wird also nur die zweite Urne nehmen, sofern er jenes Wesen für hinreichend zuverlässig hält. Wer dagegen U', Gibbards und Harpers erwarteten subjektiven Wert, maximiert, wird in jedem Fall beide Urnen nehmen.

Nach Gibbard und Harper ist das Problem also, welche Sorte von erwartetem subjektivem Wert zu maximieren ist. Ihre Ansicht ist, daß U' zu maximieren sei. Demgemäß würden sie beide Urnen nehmen, und zwar – zu meiner Freude – auch in dem Fall, in dem jenem Wesen Unfehlbarkeit zugeschrieben wird. Jedenfalls eröffnen sie beiden Positionen die Möglichkeit zu einer kohärenten Entscheidungstheorie; sowohl die Befürworter von F_1 wie die von F_2 können jeweils eine einzige Entscheidungsregel angeben, die all ihr Verhalten regelt. Die Befürworter von F_2, also auch Gibbard und Harper, haben dafür freilich einen beträchtlichen Preis zu zahlen: Sie müßten erstens eine akzeptable Logik des subjunktiven Konditionals formulieren. Der ausgereifteste Vorschlag dafür dürfte Lewis (1973) sein. Aber auch dieser ist zumindest als umstritten zu bezeich-

---

[7] Genauer gesagt: Gibbard und Harper zeigen, daß dies jedenfalls dann gilt, wenn $P(f \Box\!\!\rightarrow w \mid f) = P(f \Box\!\!\rightarrow w)$.

nen, auch wenn ich nun in diesen Streit nicht eintreten will. Zweitens hätten sie eine Nonstandard-Wahrscheinlichkeitstheorie zu schreiben, in der Wahrscheinlichkeitsmaße auch für subjunktive Konditionalsätze, genauer, für alle Sätze einer solchen Logik erklärt sein müßten. Und drittens wäre die gesamte Entscheidungstheorie entsprechend umzuformulieren, was allerdings vermutlich nicht mehr besonders schwierig wäre, wenn man die ersten beiden Dinge schon hätte. Diese Dinge waren uns aber bereits im Abschnitt 2.4 zu problematisch, um von einer echten Reduzierbarkeit von Fishburn-Modellen auf Savage-Modelle zu reden; und so ist, wie ich meine, auch hier der Preis so hoch, daß es große Anstrengungen lohnt zu vermeiden, ihn zu bezahlen.

In der Tat scheinen mir hier – und der Leser wird sich dies aufgrund der Abschnitte 2.5 und 3.3 bereits gedacht haben – keine großen Anstrengungen nötig zu sein. Denn da die Vorhersage jenes Wesens der Urnenwahl des Entscheidenden zeitlich vorausgeht, ist es unseren Ausführungen auf S. 113ff. zufolge falsch, handlungsabhängige Wahrscheinlichkeiten des Entscheidenden für die Ergebnisse der Vorhersage anzusetzen. Vielmehr ist davon auszugehen, daß diese subjektiven Wahrscheinlichkeiten des Entscheidenden einen festen Wert haben, unabhängig von seiner Entscheidung. Demgemäß ist für den Maximierer des erwarteten subjektiven Wertes nichts anderes optimal, als beide Urnen zu nehmen – wobei die jeweiligen Wahrscheinlichkeitswerte dabei gar nicht wesentlich eingehen, eben weil F_2 F_1 dominiert. Nach unserer Auffassung befindet sich also gerade der Befürworter von F_2 in Übereinstimmung mit der Standard-Entscheidungstheorie, während das Argument, das den Anhänger von F_1 als Befolger der Bayesschen Regel hinstellte, letztlich unser Verbot von subjektiven Wahrscheinlichkeiten für die eigenen Handlungen aus Abschnitt 2.5 verletzt. Es bedarf also gar keiner kunstvollen Manöver, um F_2 zu vertreten.

Insbesondere ist es recht aufschlußreich, die Auswirkungen dieses Standpunktes auf Nozicks Fallunterscheidung zu betrachten. Nach unserer Explikation kausaler Unabhängigkeit im Abschnitt 3.3 kann nämlich Nozicks Fall (2), in dem für einen Geschehensspielraum i und einen Handlungsspielraum j zwar $i \perp_c j$, aber nicht $i \perp j$ gilt, überhaupt nicht auftreten, jedenfalls dann nicht, wenn keine zwei Spielräume denselben Zeitindex erhalten: Falls i sich zeitlich vor j realisiert, muß gemäß unseren Überlegungen, die zu Definition 17 (S. 115) führten, $i \perp j$ gelten. Und falls i sich zeitlich nach j realisiert, so folgt aus $i \perp_c j$ gemäß Definition 18 (S. 121f.) $i \perp j \,/\, I^g_{<s_j}$; und da gemäß Definition 17 $I^g_{<s_j} \perp j$ gilt, folgt

daraus mit Satz 8(f) (S. 107) $\{i\} \cup I^g_{<s_j} \perp j$ und damit $i \perp j$. Nozicks Aufsplitterung der Entscheidungstheorie in verschiedene Entscheidungsregeln mit separaten Anwendungsbreichen erweist sich damit als überflüssig.[8]

Welche Möglichkeiten hat der Befürworter von F_1, sich hier aus der Affäre zu ziehen? Nun, erstens könnte er behaupten, daß in Newcombs Problem nicht-handlungsbedingte Wahrscheinlichkeiten des Entscheidenden für die möglichen Voraussagen jenes Wesens keinen Sinn haben. Und als solcher stellt sich der Befürworter von F_1 von unserem Standpunkt aus auch dar. D.h., da wir Personen von unserem Standpunkt aus nach wie vor keine Wahrscheinlichkeiten für ihre eigenen Handlungen unterstellen können, müßten wir jemandem, der in Newcombs Situation in Abhängigkeit von der beobachteten Zuverlässigkeit jenes Wesens und von den involvierten Geldbeträgen F_1 oder F_2 tut, unterstellen, daß er nur handlungsbedingte und keine absoluten Wahrscheinlichkeiten für die Voraussagen jenes Wesens hat. Gemäß unserer Feststellung auf S. 114 müßten wir ihm damit auch merkwürdige Kausalvorstellungen, nämlich die Annahme rückwirkender Einflußnahme zuschreiben.

Diese Annahme sollte aber nach der Schilderung von Newcombs Situation gerade nicht gemacht werden. Insofern empfiehlt sich diese Verteidigung nicht. In der Tat scheint es mir ganz offensichtlich zu sein, daß hier absolute Wahrscheinlichkeiten für jene Voraussagen sinnvoll sind. Der Entscheidende könnte sie möglicherweise sogar mit mehr oder weniger fundierten Vermutungen darüber stützen, welche Daten jenes Wesen über ihn gesammelt hat und mit was für einer Theorie es operiert. Hier scheint mir übrigens der Irrtum in dem Zitat von Nozick (1969) zur Unterstützung von F_1 (s. S. 186f.) zu liegen; für den Entscheidenden wäre es in jedem Fall rational, bei seiner Wette darauf, daß die zweite Urne gefüllt bzw. leer ist, von diesen absoluten Wahrscheinlichkeiten auszugehen.

Der Befürworter von F_1 hat aber noch die andere Möglichkeit, unser Prinzip aus Abschnitt 2.5 abzulehnen, daß dem Entscheidenden keine subjektiven Wahrscheinlichkeiten für seine eigenen Handlungen zu unterstellen sind. Er hätte dann freilich unsere dortigen Argumente zu entkräften. Andererseits könnte er mit

[8] Der in Anm. 5 aufgeführte vierte Fall kann dagegen sehr wohl auftreten. Das ist nicht schlimm, da in ihm wiederum die Bayessche Regel angemessen ist. Daß Nozicks Fall (2) dann nicht unbedingt ausgeschlossen ist, wenn mehrere Spielräume den gleichen Zeitindex haben, ist vielleicht ein Zeichen dafür, daß unsere Behandlung gleichzeitiger Spielräume im Abschnitt 3.3 noch nicht ideal ist.

Recht darauf hinweisen, daß auch wir noch zu erläutern haben, wie sich unser Prinzip mit der Tatsache verträgt, daß wir häufig doch probabilistische Einschätzungen unserer zukünftigen Handlungen und auch handlungsabhängige Wahrscheinlichkeiten für Ereignisse, die vor der fraglichen Handlung liegen, haben; und er könnte hinzusetzen, daß gerade die von Newcomb entworfene Situation ein glänzendes Beispiel dafür sei. Vielleicht stellt ihn die folgende Erläuterung zufrieden.

5.2 Nochmals: Wahrscheinlichkeiten für Handlungen

Natürlich soll unser Prinzip aus Abschnitt 2.5 einem Entscheidenden – sei es mal wieder unser *X* – nicht verbieten, subjektive Wahrscheinlichkeiten für seine eigenen zukünftigen Handlungen zu haben. Das tut es auch nicht, und das zu tun, wäre in der Tat verquer. Denn *X* kann ja und wird auch häufig Annahmen über die Handlungen anderer Personen haben; und was sollte an ihm so Besonderes sein, daß dies für ihn in bezug auf sich selbst nicht möglich sein sollte? Auch wird *X* vielleicht einige Überzeugungen hegen oder gar bestimmte Theorien entwickelt haben, die alle Menschen und ihr Verhalten betreffen und nicht bloß alle Menschen außer ihm.

Dagegen, daß *X* auch über seine eigenen Handlungen zu theoretisieren beginnt, ist also gar nichts einzuwenden. Im Gegenteil. In den Abschnitten 4.3 und 4.4 sind wir ja selbst davon ausgegangen, daß der strategisch Denkende Annahmen über die Entscheidungssituationen hat, in die er geraten wird, und außerdem glaubt, daß er in diesen Situationen den erwarteten subjektiven Wert maximiert. Damit hat er natürlich auch Annahmen über seine zukünftigen Handlungen. Ebenso gut können Vermutungen über die eigenen Handlungen auch von anderen Theorien herrühren; es muß ja nicht jeder der Entscheidungstheorie anhängen. Es ist nicht einmal nötig, daß da irgendeine formulierbare Theorie im Hintergrund steht. Ich kann es für sehr ungewöhnlich und unwahrscheinlich halten, daß ich im nächsten Winter außer Hauses kurze Hosen tragen werde, auch ohne dies über ein „Aber das habe ich doch noch nie getan!“ hinaus begründen zu können. Schließlich braucht das Theoretisieren nicht allein fern in der Zukunft liegende Handlungsmöglichkeiten zu betreffen, bei denen eine Entscheidung noch nicht aktuell ist; es kann auch die unmittelbar zur Entscheidung anstehenden Handlungen betreffen. So wird jemand, der sich zu einer bestimmten Handlung entschlossen hat, dann auch davon überzeugt sein, daß er sie ausführen wird.

Der springende Punkt ist eben nur, daß sich all dieses Theoretisieren über die eigenen Handlungen nicht in diesen Handlungen selbst manifestieren kann. Es manifestiert sich anderweitig, und wie mir vorkommt, weitgehend im Reden über diese Handlungen. Oder für den, der es hochtrabend liebt; Sprache ist eine Voraussetzung für Reflexion. Und deshalb hat all das, so argumentierten wir in

Abschnitt 2.5, in Entscheidungsmodellen, die der theoretischen Erfassung lediglich dieser Handlungen dienen, nichts zu suchen.

Damit rückt freilich das eigentliche Problem erst ins rechte Licht. Denn nehmen wir an, unser X sei einer, der über die Zeitläufte und damit auch über seine eigenen Handlungen theoretisiert, und nehmen wir weiter an, wir kennten irgendwoher seine theoretische Einschätzung. Gleichzeitig wollen wir ihn aber in bezug auf diese Handlungen mit einem Entscheidungsmodell beschreiben; und da wir ihn insgesamt in einer kohärenten Weise beschreiben wollen, erhebt sich natürlich die Frage: Welches und ein wie reibungsloses Verhältnis besteht zwischen Xs subjektiven Wahrscheinlichkeiten aus dem Entscheidungsmodell und seiner allgemeinen theoretischen Einschätzung?

Fassen wir die Angelegenheit etwas präziser, und zwar einfachheitshalber zunächst innerhalb eines atemporalen Rahmens: Sei also ein atemporales Entscheidungsfeld $\langle I, I^h \rangle$ gegeben, in dem I^h sich aus Handlungsspielräumen von X zusammensetzt. Die entscheidungsrelevanten subjektiven Wahrscheinlichkeiten werden dann durch eine Wahrscheinlichkeitsfamilie P für $\langle I, I^h \rangle$ gegeben. Nun theoretisiert X aber noch über dieses Entscheidungsfeld, und in unserem Kontext ist es angemessen, das Ergebnis dieser Bemühungen durch ein Wahrscheinlichkeitsmaß Q zu repräsentieren, das auf *ganz* $\mathcal{A}(I)$, also insbesondere auch für Handlungspropositionen definiert ist. Was ist dann der Zusammenhang zwischen P und Q? Die Antwort ist klar: P soll ja gerade die Wahrscheinlichkeiten aller Ereignisse in Abhängigkeit der Handlungsverläufe aus $\mathcal{Z}(I^h)$ liefern und sonst nichts. Also muß, wenn unsere Zuschreibung von P und Q konsistent sein soll, für alle $A \in \mathcal{A}(I)$ und $F \in \mathcal{Z}(I^h)$ mit $Q(F) \neq 0$ gelten: $P_F(A) = Q(A \mid F)$.

Es ist dabei ganz interessant sich anzusehen, was man von Q sozusagen wegschneiden muß, damit P übrigbleibt, d.h. welche Teile von Q es sind, aus denen sich zusammen mit P Q eindeutig zurückgewinnen läßt. Diese Frage gestattet natürlich mehr als eine Antwort. Z.B. ist, dies ist trivial, Q durch P und die Einschränkung von Q auf $\mathcal{A}(I^h)$ eindeutig bestimmt. Eine andere mögliche Antwort ist die folgende: Sei P' die dadurch festgelegte Wahrscheinlichkeitsfamilie für $\langle I, I^g \rangle$[9], daß für alle $A \in \mathcal{A}(I)$ und $G \in \mathcal{Z}(I^g)$ $P'_G(A) = Q(A \mid G)$. Dann gilt, daß Q durch P und P' eindeutig bestimmt ist, d.h. daß jedes Wahrscheinlichkeitsmaß Q' auf $\mathcal{A}(I)$, das zu P und P' in derselben Beziehung steht wie Q, mit Q identisch

[9] Es ist, wie erinnerlich, $I^g = I \setminus I^h$.

ist.[10] P' ist dabei sogar die *minimale* derartige Wahrscheinlichkeitsfamilie in dem Sinne, daß Q von P und der Wahrscheinlichkeitsfamilie P^J für $\langle I, J\rangle$, für die für alle $A \in \mathcal{A}(I)$ und $G \in \mathcal{Z}(J)$ $P^J_G(A) = Q(A \mid G)$, nur dann eindeutig bestimmt ist, wenn $J \subseteq I^g$.[11] Insofern können wir also auch sagen, daß P gerade der Rest ist, der übrigbleibt, wenn man P' von Q wegnimmt.

Wie stellt sich nun die Sache innerhalb eines temporalen Rahmens dar? Nicht ganz so einfach, dafür umso interessanter: Gegeben sei ein temporales Entscheidungsfeld $\langle I, I^h, T, \leq, s\rangle$, eine Wahrscheinlichkeitsfamilie P dafür, das Xs entscheidungsrelevante subjektive Wahrscheinlichkeiten ausdrücke, und ein Wahrscheinlichkeitsmaß Q auf $\mathcal{A}(I)$, das Xs Theoretisiererei über die Spielräume aus I zusammenfasse. Der Anschaulichkeit halber kann man sich vorstellen, daß es etwa um Xs Besuch eines Freibades am 17.7.1990 geht. X wird da jetzt eine Menge Vermutungen haben. Er wird es z.B. für relativ unwahrscheinlich halten, daß er an jenem Tage um 14h im Freibad ist, da er ohnehin selten ins Freibad geht und sich auch nicht ganz sicher ist, daß er dann überhaupt noch lebt. Unter der Bedingung, daß an jenem Tage um 12h oder auch um 15h warmes, sonniges Wetter ist, ist diese Wahrscheinlichkeit schon etwas höher. Und umgekehrt wird er es unter der Bedingung, daß er um 14h im Freibad ist, als fast sicher annehmen, daß um 12h schönes Wetter ist, und als fast ebenso sicher, daß um 15h schönes Wetter ist. Beides wäre ohne diese Bedingung nicht der Fall. Und so weiter. All das ist in Q niedergelegt. Ferner können wir uns vorstellen, daß wir unseren X jetzt einfrieren – eine freilich zweifelhafte Therapie gegen seinen früheren Liebeskummer und die daraus resultierenden Alkoholprobleme – und ihn am 17.7. 1990 um 13h wieder auftauen, so daß seine Einschätzung Q unverändert erhalten geblieben und weder durch neue Informationen verbessert, noch durch Gedächtnislücken getrübt ist. Nun soll er sich entscheiden, ob er ins Freibad gehen will oder nicht. Die dafür relevanten Wahrscheinlichkeiten seien durch die Wahrscheinlichkeitsfamilie P gegeben. Wie hängt P mit Q zusammen?

Sicherlich gilt nicht $P_F(A) = Q(A \mid F)$ für alle $A \in \mathcal{A}(I)$ und $F \in \mathcal{Z}(I^h)$.Denn dann wäre P in der Regel gar keine Wahrscheinlichkeitsfamilie für ein *tempo-*

---

[10] Dieser Sachverhalt ist mitnichten trivial. Sein Beweis läuft ebenso wie der von Satz 8(g), S. 107. Vgl. dazu Spohn (1980).

[11] Dies ergibt sich ebenfalls aus dem Beweis von Satz 8(g). „Minimal“ ist insofern der angemessene Ausdruck, als P' durch jedes P^J mit $J \subseteq I^g$ festgelegt ist, P' seinerseits aber nur die P^J mit $J \subseteq I^g$ festlegt. Ferner ist zu betonen, daß P' nicht in dem Sinne minimal ist, daß gälte: Jeder Teil von P', der zusammen mit P Q eindeutig bestimmt, bestimmt für sich allein P' eindeutig. Dies gilt, wie gesagt, nicht.

rales Entscheidungsfeld, da nicht gesichert ist, daß in bezug auf das so festgelegte P für alle $t \in T$ $I^g_{\leq t} \perp I^h_{>t}$ gilt. Vielmehr darf P die Wahrscheinlichkeit eines Ereignisses zu t von Q nur übernehmen, soweit sie von Handlungsverläufen bis t abhängt. Es muß also für alle $t \in T$, $A \in \mathcal{Z}(I_{<t})$, $B \in \mathcal{Z}(I_{=t})$, $F \in \mathcal{Z}(I^h_{\leq t})$ und $G \in \mathcal{Z}(I^h_{>t})$ mit $Q(A \cap F) \neq 0$ gelten:

$$(5.1) \qquad P_{F \cap G}(B \mid A) = Q(B \mid A \cap F).$$

In der Tat ist durch (5.1) P als Wahrscheinlichkeitsfamilie für $\langle I, I^h, T, \leq, s \rangle$ eindeutig bestimmt: Sei nämlich $T = \{t_1, ..., t_n\}$, wobei $t_p < t_q$ für $p < q$. Für $t = t_1$ sichert (5.1), daß für jedes $F \in \mathcal{Z}(I^h)$ P_F auf $\mathcal{A}(I_{\leq t_1})$ festgelegt ist; und wenn P_F bereits für $\mathcal{A}(I_{\leq t_r})$ bestimmt ist, so garantiert (5.1) mit $t = t_{r+1}$, daß P_F dann auch auf $\mathrm{A}(I_{\leq t_{r+1}})$ festgelegt ist. Ebenso ist klar, daß für das so festgelegte P tatsächlich $I^g_{\leq t} \perp I^h_{>t}$ alle $t \in T$ gilt.

Die Frage, was von Q wegzuschneiden ist, um zu diesem P zu gelangen, läßt sich analog zum atemporalen Fall beantworten: Sei nämlich P' die dadurch festgelegte Wahrscheinlichkeitsfamilie für $\langle I, I^h, T, \leq, s \rangle$, daß für alle $t \in T$, $A \in \mathcal{Z}(I_{<t})$, $B \in \mathcal{Z}(I_{=t})$, $F \in \mathcal{Z}(I^g_{\leq t})$ und $G \in \mathcal{Z}(I^g_{>t})$ $P'_{F \cap G}(B \mid A) = Q(B \mid A \cap F)$. Dann ist Q wiederum durch P und P' eindeutig bestimmt[12], und außerdem ist P' wieder die *minimale* derartige Wahrscheinlichkeitsfamilie insofern, als Q von P und der entsprechend festgelegten Wahrscheinlichkeitsfamilie P^J für $\langle I, J, T, \leq, s \rangle$ nur dann eindeutig bestimmt ist, wenn $J \subseteq I^g$. Abermals läßt sich P als Differenz von Q und P' auffassen.

Wesentlich ist nun, zu betrachten, was den temporalen Fall auszeichnet: die durch Q und P (und P') gegebenen Kausalverhältnisse (von denen im atemporalen Fall nicht sinnvoll geredet werden konnte). Q legt für alle $J, J' \subseteq I$ fest, ob J' von J kausal abhängt oder nicht; P regelt die kausalen Abhängigkeiten aller $K \subseteq I^g$ von allen $J \subseteq I$ und P' entsprechend die aller $H \subseteq I^h$ von allen $J \subseteq I$. Dabei gilt: *$K \subseteq I^g$ ist von $J \subseteq I$ gemäß P genau dann kausal unabhängig, wenn K von J ge-*

[12] Dies läßt sich folgendermaßen einsehen: Sei $T = \{t_1, ..., t_n\}$, wobei $t_p < t_q$ für $p < q$. Dann ist zunächst durch die Beschränkungen aller P_F ($F \in \mathcal{Z}(I^h)$) und aller P'_G ($G \in \mathcal{Z}(I^g)$) auf $\mathcal{A}(I_{=t_1})$ Q auf $\mathcal{A}(I_{=t_1})$ eindeutig bestimmt – dies lehrt der atemporale Fall. Sei nun Q bereits für $\mathcal{A}(I_{\leq t_r})$ bestimmt. Dann ist für jedes $C \in \mathcal{Z}(I_{\leq t_r})$ durch die Beschränkungen aller $P_F(. \mid C)$ und $P'_G(. \mid C)$ auf $\mathcal{A}(I_{=t_{r+1}})$ auch $Q(. \mid C)$ auf $\mathcal{A}(I_{=t_{r+1}})$ eindeutig bestimmt – dies lehrt wiederum direkt der atemporale Fall; und damit ist Q auf ganz $\mathcal{A}(I_{\leq t_{r+1}})$ eindeutig festgelegt.

mäß Q kausal unabhängig ist,[13] und ebenso ist $H \subseteq I^h$ von $J \subseteq I$ gemäß P' genau dann kausal unabhängig, wenn H von J gemäß Q kausal unabhängig ist. Dasselbe gilt für die Abschirmbarkeit. Umgekehrt brauchen wir von den Wahrscheinlichkeiten $Q(B \mid A)$, in denen das Ereignis B dem Ereignis A folgt, für die es also ein $t \in T$ mit $B \in \mathcal{A}(I_{\geq t})$ und $A \in \mathcal{A}(I_{\leq t})$ gibt, gerade diejenigen, die P gemäß (5.1) festlegen, um für alle $K \subseteq I^g$ und $J \subseteq I$ festzustellen, ob K von J gemäß Q kausal abhängt bzw. abschirmbar ist oder nicht. Dies wird insbesondere an der Aussage „$I^g_{=t} \perp_{\mathrm{a}} I^h_{=t'}$ gemäß Q" (wobei $t' < t$) deutlich, die ja nach Definition 19 von $\perp_{\mathrm{a}}$ (S. 123) mit der Aussage äquivalent ist, daß $I^g_{=t} \perp I^h_{=t'} \,/\, I^g_{<t} \cup (I^h_{\leq t} \setminus I^h_{=t'})$ gemäß Q. Und zu deren Überprüfung benötigen wir von den genannten Wahrscheinlichkeiten gerade diejenigen, die in (5.1) auftauchen.

Zusammenfassend läßt sich also sagen, daß P gerade derjenige Teil von Q ist, in dem die kausalen Abhängigkeiten der Geschehenspielräume voneinander und von den Handlungsspielräumen, wie sie gemäß Q bestehen, festgehalten sind, während der weggeschnittene Teil P' gerade die durch Q gegebenen kausalen Abhängigkeiten der Handlungsspielräume voneinander und von den Geschehensspielräumen ausdrückt. Und P ist gerade derjenige Teil von Q, der sich in Xs Handlungen aus diesen Handlungsspielräumen niederschlägt; allenfalls er läßt sich wieder aus ihnen entnehmen. Eine Person nimmt also *als Handelnde* ihre Handlungen aus dem kausalen Ereignisfluß heraus und sieht sie nur als Bewirkende und nicht als Bewirkte. So ist z.B. für unseren frisch aufgetauten und über den Freibadbesuch sinnierenden X nicht relevant, daß er unter der Bedingung, daß er ins Freibad geht, schönes Wetter um 12h für wahrscheinlicher hält als unter der gegenteiligen Bedingung. Ebenso kommt es in Newcombs Situation nur auf die absoluten Wahrscheinlichkeiten dafür, daß die zweite Urne gefüllt ist, an und nicht auf die dem Theoretisieren zuzurechnenden, handlungsbedingten Wahrscheinlichkeiten dafür.

So weit meine Erläuterung, inwiefern sich unser Prinzip aus Abschnitt 2.5 sehr wohl mit Wahrscheinlichkeiten für Handlungen, die eben nur theoretisch, nicht entscheidungsrelevant sind, verträgt; wie die Wahrscheinlichkeitsfamilien, die in Entscheidungsmodelle Eingang finden, sich zu umfassenderen theoreti-

---

[13] Gemäß Satz 11 (S. 123) ist dazu nur nachzuprüfen, daß für alle $i \in I^g$ und $j \in I$ mit $s_i \geq s_j$ $i \perp_{\mathrm{c}} j$ gemäß P genau dann gilt, wenn $i \perp_{\mathrm{c}} j$ gemäß Q. Und das ist nicht schwierig, lediglich etwas mühselig. Man schreibe dazu die Behauptung mittels der Definitionen von $\perp_{\mathrm{c}}$ und $\perp$ ausführlich aus; dann fällt sie fast ins Auge.

schen Einschätzungen erweitern und umgekehrt daraus wieder gewinnen lassen. Ich bin mir allerdings im klaren darüber, daß unsere Analysen insgesamt nur an der Oberfläche kratzen, daß sie dort abbrechen, wo sie noch mehr in die Breite und in die Tiefe gehen könnten und sollten. Um nur einige aktuelle Punkte zu nennen: Unsere dauernde Rede von Teilen und Resten von Wahrscheinlichkeitsmaßen und dergleichen war mathematisch sicherlich noch nicht sehr sauber. Ferner ist das Verhältnis zwischen der, wenn ich so sagen darf, praktischen Wahrscheinlichkeitsfamilie P für $\langle I, I^h, T, \leq, s\rangle$ und dem theoretischen Wahrscheinlichkeitsmaß Q auf $\mathcal{A}(I)$, wie ich es beschrieben habe, nicht so problemlos, wie es den Anschein haben mag. Eine Schwierigkeit ist, daß sich P und Q mit wachsender Erfahrung auseinander entwickeln. Damit ist folgendes gemeint: Mögen P und Q einander über (5.1) entsprechen. Werde nun die Erfahrung gemacht, daß C. Per einfacher Konditionalisierung ändert sich Q zu Q^C mit $Q^C(A) = Q(A \mid C)$ für alle $A \in \mathcal{A}(I)$ und P zu P^C mit $P_F^C(A) = P_F(A \mid C)$ für alle $A \in \mathcal{A}(I)$ und $F \in \mathcal{Z}(I^h)$. Es ist dann mitnichten – insbesondere dann nicht, wenn C eine Handlungsproposition ist – garantiert, daß auch P^C und Q^C im zu (5.1) analogen Verhältnis stehen. Ein zweites Problem ist dieses: Sei J eine Teilmenge von I mit $I^h \subseteq J$. Dann gilt nicht unbedingt, daß die Einschränkung von P auf $\mathcal{A}(J)$ (d.h. die Familie der Beschränkungen aller P_F auf $\mathcal{A}(J)$) und die Einschränkung von Q auf $\mathcal{A}(J)$ in dem durch (5.1) gegebenen Verhältnis stehen. Hier harrt also noch einiges der Aufklärung.

Schließlich wäre hier nun auch der rechte Ort, um den der Entscheidungstheorie angemessenen Handlungsbegriff etwas zu klären. Denn sicherlich ist unsere Feststellung, daß die Handlungen einer Person von ihr als Handelnder aus probabilistischen Einschätzungen und kausalen Abhängigkeiten ausgeklammert sind, für den Handlungsbegriff nicht folgenlos. So ergibt sich daraus, daß Handlungen, wie zu erwarten, Dinge sind, die – zumindest nach der Meinung des jeweiligen Handelnden – unter vollständiger Kontrolle des Handelnden stehen. Denn Ereignisse, die nicht unter der vollständigen Kontrolle des Handelnden sind, sind ja im Prinzip seiner probabilistischen und auch entscheidungsrelevanten Beurteilung zugänglich; z.B. ließen sich von den Handlungen verschiedene Faktoren nennen, die auch noch zu dem Zustandekommen dieser Ereignisse beitragen.

Man verstrickt sich hier noch nicht automatisch in das Problem, wie weit man in den Körper des Handelnden eindringen muß, um endlich etwas zu finden, was unter seiner vollständigen Kontrolle ist. Denn die Rede von der vollständigen Kontrolle ist hier immer relativ zu dem Entscheidungsmodell zu verstehen, mit dem der Handelnde beschrieben werden soll. So ist z.B. mein Gang vom Stachus zum Marienplatz nicht in einem absoluten Sinne unter meiner vollständigen Kontrolle; es gibt da eine Menge Hindernisse. Dennoch darf natürlich in einem Entscheidungsmodell, das sich um diese Hindernisse nicht kümmert, mein Gang vom einen zum anderen Platz als Handlung fungieren. Bereicherte man nun dieses Entscheidungsmodell um alle meine einzelnen Schritte, so träten selbstverständlich diese Schritte und nicht der ganze Gang als Handlungen auf. Es wäre also korrekter zu sagen, daß in einem Entscheidungsfeld diejenigen Spielräume als Handlungsspielräume fungieren, die von all den Spielräumen dieses Entscheidungsfeldes am *ehesten* unter der Kontrolle des Handelnden stehen. Erst wenn man auf die wirklich vollständige Kontrolle hinaus will, muß man sich in das Problem der sogenannten Basishandlungen vertiefen.[14]

Würgen wir jedoch dieses Thema mit der Bemerkung ab, daß sich von unseren Ausführungen, wie mir scheint, mittlerweile eine tragfähige Brücke zu dem, was unter Handlungstheorie firmiert, schlagen lassen müßte. Ich hoffe, daß trotz dieser, der früher erwähnten und der vielen unerwähnten Lücken durch unsere Ausführungen insbesondere in den Abschnitten 2.5, 3.3, 5.1 und 5.2 ein genügend kohärentes Bild gewachsen ist, um den Eindruck zu vermitteln, daß wir wenigstens an der richtigen Stelle gekratzt haben und man an dieser Stelle noch weiter bohren sollte.

Wo wir aber gerade noch am Kratzen waren, juckt es mich doch, einen letzten Gedanken in die Debatte zu werfen, das undurchsichtige Thema Willensfreiheit betreffend. Vielleicht ließe sich der leidige Gegensatz zwischen Freiheit und Notwendigkeit, d.h. zwischen Freiheit und kausaler Determiniertheit, damit aus der Welt schaffen, daß wir als Handelnde, wie festgestellt, unsere Handlungen nur als Anfangsglieder und nie als Zwischenglieder von Kausalketten betrachten, daß sich also für uns als Handelnde die Frage nach der kausalen Determiniertheit unserer Handlungen gar nicht sinnvoll stellt? Möge sich darüber jeder seine eigene Meinung bilden.

[14] Wie es in der Handlungstheorie diskutiert wird; vgl. etwa Meggle (1977), S. 89-162. Jeffreys Ausführungen über probabilistische Handlungen und Versuche in Jeffrey (1965), Abschnitt 11.9, scheinen mir ein bemerkenswerter, wenn auch noch zu wenig beachteter Beitrag dazu zu sein.

5.3 Zur empirischen Signifikanz der Entscheidungstheorie

Bisher haben wir im wesentlichen so getan, als ginge es darum, eine empirische Theorie über das Verhalten von Menschen und anderen Subjekten zu formulieren. Doch taten wir nur so; in Wirklichkeit hielten wir uns an die auf S. 37 abgesprochene Taktik, die Interpretation der Entscheidungstheorie als empirischer Theorie lediglich als Stellvertreter zumindest für die vier anderen dort genannten Interpretationsmöglichkeiten zu betrachten. Machen wir nun, die Interpretationsproblematik nach wie vor links liegen lassend, zu guter Letzt einmal mit der empirischen Interpretation der Entscheidungstheorie Ernst, um so wenigstens eine der vielen Ausklammerungen von Abschnitt 2.1, nämlich die der Überprüfungsproblematik, wieder etwas rückgängig zu machen. Erkundigen wir uns also nach dem empirischen Gehalt der Entscheidungstheorie, und fragen wir uns, wie er sich überprüfen ließe.

Da ist zunächst die ernüchternde Feststellung zu treffen, daß das Bisherige als empirische Theorie überhaupt nicht viel taugt. Das hat einen simplen Grund: Nehmen wir einfachheitshalber auf die etwas subtile Unterscheidung zwischen Handlungen und sonstigem Verhalten keine Rücksicht. Die einzigen Daten, die wir über ein Subjekt haben, solange wir von physiologischen Erhebungen Abstand nehmen, bestehen dann gerade aus all seinem tatsächlichen Verhalten. Und wie dieses Verhalten auch immer aussieht, es dürfte einleuchten, daß sich dazu immer ein Entscheidungsmodell erfinden und diesem Subjekt unterstellen läßt, in dem just dieses Verhalten maximalen subjektiven Wert erhält. Solange also die Bayessche Regel die einzige Beschränkung ist, die die Zuschreibung von Entscheidungsmodellen zu Subjekten zügelt, ist die Entscheidungstheorie noch völlig inhaltsleer. Oder anders gewendet: Es müssen noch all die Gründe untersucht und explizit gemacht werden, die uns jemandem eher dieses als jenes Entscheidungsmodell unterstellen lassen und die unserer Auswahl von unterstellbaren Entscheidungsmodellen derart einengen, daß sich das Prinzip von der Maximierung des erwarteten subjektiven Wertes nicht mehr trivial erfüllen läßt. Diese Gründe, die es ja in Hülle und Fülle gibt, will ich nun nicht systematisch untersuchen; ich könnte es auch gar nicht. Ich will hier lediglich etwas verdeutlichen, wieviel es da – immer noch – zu tun gibt. Dazu ist gut, sich klarzumachen, daß an einem unterstellten Entscheidungsmodell alles, aber auch wirklich alles hypothetisch ist:

Dies gilt erstens, das geht bei solchen Erörterungen häufig unter, für die Gesamt- und für die Handlungsspielraummenge eines Entscheidungsmodells. Welche Spielräume in einer Entscheidungssituation eine Rolle spielen, ist nicht von außen objektiv vorgebbar, sondern hängt ganz von dem Subjekt ab, das sich in dieser Situation befindet. Dies stellten wir schon im Abschnitt 3.5 fest, wo wir auf dem Wege zur Definition zusammenhängender Entscheidungsmodelle ermittelten, welche Spielräume gemäß den subjektiven Wahrscheinlichkeiten und Werten eines Entscheidenden für einen bestimmten Handlungsspielraum relevant sind.

Doch erschöpft sich darin die Subjektabhängigkeit der anzusetzenden Gesamtspielraummenge noch nicht ganz. Gingen wir von einer normativen Interpretation der Entscheidungstheorie aus, so könnten wir jetzt sagen, daß ein Individuum in seinem eigenen Interesse gefälligst alle Spielräume berücksichtigen *solle*, die gemäß seinen Überzeugungen und Wünschen eine Rolle spielen. Doch nicht einmal dieses Maß an objektiver Festlegung der Gesamtspielraummenge läßt sich bei einer empirischen Interpretation erreichen. Denn die Leute berücksichtigen in der Regel nicht alles, was sie berücksichtigen sollten. Und geht es um empirische Adäquatheit, so sollte man nur diejenigen Spielräume in Entscheidungsmodelle hineinschreiben, an die die Leute denken; denn in ihre Handlungen gehen nur ihre Überzeugungen und Wünsche bezüglich dieser Spielräume ein.

Diese Relativität auf das jeweils zu beschreibende Subjekt läßt sich nicht beseitigen, aber sie ließe sich zumindest in einer, wie ich meine, sehr sinnvollen Weise verschieben: Man verwende eine möglichst umfassende Gesamtspielraummenge, die man überdies zeitlich konstant halten kann, und bereichere dafür die Entscheidungsmodelle um so etwas wie einen Aufmerksamkeitsoperator. Dieser Aufmerksamkeitsoperator ist natürlich subjekt- und zeitabhängig, und er sucht aus der Gesamtspielraummenge diejenige Teilmenge heraus, auf die das fragliche Subjekt zum jeweiligen Zeitpunkt achtet. In den erwarteten subjektiven Wert von Handlungen gehen dann nur die subjektiven Werte und Wahrscheinlichkeiten bezüglich der Spielräume dieser Teilmenge ein. Man könnte den Aufmerksamkeitsoperator auch noch etwas raffinierter gestalten, so daß er nicht nur aus der Gesamtspielraummenge eine Teilmenge heraussucht, sondern auch noch jeden Spielraum aus dieser Teilmenge in geeigneter Weise einschränkt und vergröbert. Damit ließe sich berücksichtigen, daß Leute häufig Spielräume nicht in ihrer ganzen Bandbreite und Differenziertheit überschauen. In jedem Fall bedürfte all dies der sorgfältigen technischen Ausarbeitung.

Daß ein solcher Aufmerksamkeitsoperator, so weit ich weiß, in der Entscheidungstheorie noch nicht eingeführt wurde, wundert mich etwas. Schließlich ist die Idee mitnichten neu. Ähnliches wurde in der Psychologie, insbesondere in der Lerntheorie, von jeher für relevant gehalten und hat sich sogar mathematisch explizit in der sogenannten Stimulus-Sampling-Theorie niedergeschlagen.[15]

Übrigens ließe sich mit Hilfe eines solchen Aufmerksamkeitsoperators die Unterscheidung zwischen beschränkten und umsichtigen Individuen, wie wir sie im Kapitel 4 diskutierten, formal sehr einfach vornehmen. Die Gesamtspielraummenge könnte von vornherein auch so komplizierte Spielräume enthalten, wie sie die Rekursionsquellen von E2-Modellen darstellen. Der Umsichtige zeichnet sich dann gegenüber dem Beschränkten gerade dadurch aus, daß diese Rekursionsquellen im Bereich seiner Aufmerksamkeit liegen.

Schließlich ist darauf hinzuweisen, daß mit der Einführung eines Aufmerksamkeitsoperators die Subjektabhängigkeit der unterstellbaren Gesamtspielraummenge nicht unbedingt völlig zu eliminieren ist – jedenfalls dann nicht, wenn man nur solche Spielräume heranziehen will, bezüglich derer das jeweilige Subjekt wenigstens im Prinzip manifestierbare Überzeugungen und Wünsche hat. Und das variiert innerhalb eines Subjektes mit der Zeit; ein Kind etwa, das noch nicht gelernt hat, Wochentage zu unterscheiden, kann natürlich auch in seinen Überzeugungen und Wünschen den Spielraum der Wochentage nicht berücksichtigen. Erst recht variiert das von Subjekt zu Subjekt, und es gibt sogar Variationen kultureller und geschichtlicher Größenordnung. So waren z.B. die Menschen erst im 19. Jahrhundert dazu imstande, einen ganzen Spielraum möglicher Geometrien zu bedenken.

Gründe für die Unterstellung dieser oder jener Gesamtspielraummenge bzw. Aufschlüsse über die empirische Beschaffenheit dieses Aufmerksamkeitsoperators dürften freilich recht indirekter Natur sein. Wie immer man sie erhält, daß man sie suchen und finden muß, ist jedenfalls sehr wichtig. Dies machen insbe-

[15] Vgl. etwa Suppes (1969), S. 421-426.

sondere die in Abschnitt 3.6 vorgestellten Möglichkeiten zur Reduktion von Entscheidungsmodellen und der dort bewiesene Satz 19 (S. 145) deutlich, wonach ein Entscheidungsmodell und seine Reduktionen um *beliebige* Geschehensspielraummengen in bezug auf den erwarteten subjektiven Wert von Handlungsverläufen und damit empirisch völlig äquivalent sind, solange Entscheidungsmodelle nur über die Bayessche Regel mit den Daten in Verbindung stehen.

Neben den Spielräumen, die man in Anschlag bringt, sind zweitens natürlich auch die mit einem Entscheidungsmodell unterstellten subjektiven Wahrscheinlichkeiten und Werte hypothetisch. Dies war von jeher sonnenklar und wurde als gravierend bis fatal empfunden. So taten die in der Ökonomie in diesem Jahrhundert bis in die 30er Jahre hinein tonangebenden Ordinalisten, die, grob gesagt, nur ordinale, d.h. komparative Begriffe in der Ökonomie als sinnvoll anerkannten, quantitative subjektive Wahrscheinlichkeiten und Werte gar als pure Metaphysik ab. Daher war es wichtig und verständlich, daß die weniger Radikalen, um Oberhand zu gewinnen, nach möglichst starken Geschützen suchten – und tatsächlich auch wirksame fanden –, die dem hypothetischen Charakter quantitativer Glaubens- und Wünschensgrade wenigstens seine Fragwürdigkeit nahmen. Als Geschütze dienten dabei Metrisierungsresultate unterschiedlichen Kalibers.

A1s erste fuhren von Neumann und Morgenstern (1944) ein solches Geschütz auf, und es erwies sich als durchaus entwaffnend. Sie zeigten nämlich, daß subjektive Werte für eine Menge von Gegenständen dann existieren und im gewünschten Maße, d.h. bis auf positiv lineare Transiormationen, eindeutig bestimmt sind, wenn man von einer Präferenzrelation mit bestimmten Eigenschaften ausgeht, die nicht nur zwischen diesen Gegenständen, sondern auch zwischen beliebigen Lotterien, deren Preise diese Gegenstände sind, erklärt ist.[16] Und etwas später rückten noch kräftigere Kaliber in Stellung, in Gestalt von Savages Metrisierungstheorem und seinen Nachfolgern, deren Inhalt wir schon in Abschnitt 2.8 umrissen.

Daß man sich immer noch mit solcher Hingabe dem Metrisierungsproblem widmet, ist dagegen nicht mehr recht verständlich und eigentlich nur aus der Überzeugung heraus erklärlich, daß man sich *nur* so mit dem hypothetischen Charakter quantitativer Glaubens- und Wünschensgrade anfreunden, *nur* so ih-

[16] Für eine präzise und elegante Darstellung dieses Sachverhalts vgl. etwa Fishburn (1970), Kap. 8.

nen Sinn einflößen könne. Eine vergleichbare Überzeugung etwa hinsichtlich der Newtonschen Mechanik, die Meinung also, die Einführung eines quantitatiren Massebegriffs sei allein durch eine geeignete Metrisierung gerechtfertigt, stieße sicherlich auf entschieden mehr Skepsis und ziehe vor allem die Physiker früherer Jahrhunderte zumindest der Unbedachtheit. Es ist hier gewiß nicht mehr der Ort, in eine grundsätzliche Debatte über die Meßproblematik einzusteigen. Auch steht außer Zweifel, daß die vielen und immer weiter verbesserten Metrisierungstheoreme einen unschätzbaren Beitrag dazu liefern. Daß sie aber die einzige Methode zur Einführung quantitativer empirischer Begriffe darstellen, ist fragwürdig. So kommt es mir auch so vor, als drechsle man mit besonderem Eifer an einem Pfeiler der sehr komplizierten Konstruktion herum, die man zur Untermauerung einer empirisch aufgefaßten Entscheidungstheorie benötigt, ohne sich recht vergewissert zu haben, worauf seinerseits dieser Pfeiler fußt und ob dieser Pfeiler für eine tragfähige Konstruktion wirklich unerläßlich ist. Schauen wir uns einmal den Unterbau dieses Pfeilers, die Voraussetzungen entscheidungstheoretischer Metrisierungen, etwas näher an:

Da muß erstens natürlich eine Präferenzordnung für die jeweils betrachtete Menge von Alternativen, in der Regel Handlungen, gegeben sein. Nun ist aber die Präferenzordnung eines Individuum auch nichts, was sich direkt aus seinem Verhalten entnehmen ließe. Direkt manifestiert sich, so steht zu vermuten, nur, was in der Präferenzordnung ganz oben steht – eine Vermutung freilich, die auf ebenso triviale Weise wahr zu machen ist wie das Bayessche Prinzip, solange wir uns in der Unterstellung von Präferenzordnungen nicht weitere Beschränkungen auferlegen. Solche Beschränkungen können sich z.B. daraus ergeben, daß wir jenes Individuum zu Vergleichen von je zwei Alternativen auffordern. Wie man diese Vergleiche dann zu einer umfassenderen Präferenzordnung zusammensetzen kann, bedarf dann immer noch der Klärung. Oder sie können sich ergeben, indem man jenes Individuum nach seinen Präferenzen fragt – wobei es dann das Verhältnis zwischen berichteten und tatsächlichen Präferenzen zu untersuchen gälte. Man kann also schon hier, und nur darauf will ich hinaus, viele methodologische Skrupel entwickeln.

Es kommt noch hinzu, daß aus technischen Gründen stets eine Präferenzordnung auf einer fürchterlich aufgeblähten Menge von Alternativen gegeben sein muß. Die Experten bemühen sich zwar – erfolgreich – um Sparsamkeit; doch ist außer in glücklichen Spezialfällen immer nur eine Teilordnung davon empirisch ermittelbar.

Dabei müssen Metrisierungen entgegen äußerem Anschein noch mehr voraussetzen als nur geeignete Präferenzrelationen. Dies läßt sich z.B. an Savages Metrisierung sehr deutlich zeigen. In Anbetracht unseres Satzes 19(d) (S. 145) über die Reduktion von Entscheidungsmodellen kann es schon gar nicht sein, daß Savages Metrisierung nur eine Präferenzordnung für die möglichen Handlungen voraussetzt; daraus ließen sich allenfalls die erwarteten subjektiven Werte dieser Handlungen entnehmen. In der Tat ist ja auch mit Savages Repräsentation von Handlungen als Funktionen von einer bestimmten Menge von Umständen in eine bestimmte Menge von Konsequenzen bereits über das zu unterstellende Entscheidungsfeld entschieden. Nur dadurch lassen sich aus der Präferenzordnung mittels Metrisierung subjektive Wahrscheinlichkeiten und Werte für dieses Entscheidungsfeld herausquetschen. Metrisierungen setzen also gerade den ersten in diesem Abschnitt besprochenen Punkt als erledigt voraus.

Außerdem werden mit dieser Repräsentation von Handlungen schon gleich gewisse subjektive Wahrscheinlichkeiten unterstellt, nämlich daß die durch Handlungen und Umstände bedingten subjektiven Wahrscheinlichkeiten für gewisse Konsequenzen gleich 1 und für die anderen Konsequenzen gleich 0 sind. Diesen Sachverhalt notierten wir schon auf S. 48. Erst damit ist vollständig aufgezählt, wovon die Metrisierung von Savage und entsprechend die von Luce und Krantz ausgehen muß.

Die Metrisierung von v. Neumann und Morgenstern und in der Tat alle Metrisierungen, die mit Lotterien, mit „extraneous probabilities" und dergleichen arbeiten, machen eine noch stärkere Annahme, die nur ungenügend damit ausgedrückt ist, daß die objektiven Lotteriewahrscheinlichkeiten als gegeben betrachtet werden. Denn ausschlaggebend für die Präferenzen eines Individuums sind allein seine subjektiven Wahrscheinlichkeiten. Metrisierungen dieser Art müssen also unterstellen, daß sich die subjektiven Wahrscheinlichkeiten des betreffenden Individuums mit den objektiven Lotteriewahrscheinlichkeiten völlig decken, und nehmen so eine Menge subjektiver Wahrscheinlichkeiten als Startpunkt.

So weit die durchweg hypothetischen Grundlagen von Metrisierungen. Was nun die Entbehrlichkeit von Metrisierungen betrifft, so habe ich natürlich keine Alternative vorzuschlagen, sondern lediglich eine allgemeine Beobachtung anzuführen: Die Wissenschaftstheoretiker sind zwar immer noch über fast alles uneins, aber einer der wenigen Punkte, über den sie sich kaum mehr streiten, ist, daß theoretische Begriffe nicht prinzipiell nicht eliminierbar zu sein brauchen – was immer theoretische Begriffe genau sind und wie immer sie eingeführt werden und empirische Signifikanz gewinnen können.[17] Metrisierungen zeigen nun aber gerade die Eliminierbarkeit quantitativer Glaubens- und Wünschensgrade. Metrisierungen beinhalten ja, daß es zu jeder Präferenzordnung genau ein subjektives Wahrscheinlichkeitsmaß und – nach willkürlicher Festlegung zweier verschiedener subjektiver Werte – genau eine subjektive Wertfunktion gibt, die zusammen mit dieser Präferenzordnung übereinstimmen. Damit lassen sich diese zwei Funktionen durch die Präferenzordnung definieren (nämlich als diejenigen, die mit ihr übereinstimmen) und so auch wieder eliminieren. Nach praktisch einhelligem wissenschaftstheoretischem Urteil leisten also Metrisierungen lediglich sehr Nützliches, keineswegs aber Unerläßliches für die empirische Signifikanz quantitativer Glaubens- und Wünschensgrade.

Damit ist freilich nur gesagt, daß es Alternativen zu Metrisierungen geben müßte; worin sie bestehen, liegt ebenso im Dunkeln wie die gesamte Konstruktion zur Untermauerung der Entscheidungstheorie. Daß sie sehr kompliziert ist, unterliegt keinem Zweifel, und so dürfte sie nur sehr langsam immer mehr entschleiert werden. [18]

Wegen dieser Kompliziertheit sind alle direkten experimentellen Tests[19] zwar fraglos wertvoll, aber auch mit Vorsicht zu genießen. Schwerwiegende Folgerungen daraus können, wie mir scheint, nur voreilig sein. Um einen vielleicht nicht allzu unpassenden Vergleich zu wählen: Es habe jemand gewisse Grundprinzipien der Aerodynamik formuliert, und diese sollen nun anhand bestimmter Körper im Windkanal getestet werden. Leider stehen dazu aber nur einige Nachbildungen sehr unregelmäßiger Landschaftsformationen und Stadtbilder zur Verfügung. Ob dann die Windkanalversuche je scharfe Aussagen über diese Grundprinzipien zuließen? Und Menschen (oder auch Tiere) sind halt mindestens so vertrackte Gegenstände wie jene Nachbildungen.[20]

---

[17] Vgl. dazu etwa Stegmüller (1970), Kap. III-V, und (1973b), Kap. VII.

[18] All die genannten Punkte, der hypothetische Charakter der unterstellten Entscheidungsfelder und der unterstellten Präferenzen und subjektiven Werte und Wahrscheinlichkeiten sowie der nicht unbedingt benötigte Eliminierungseffekt von Metrisierungen, wurden mir durch einen Vortrag von Herrn Prof. Sneed in München im Januar 1975 nachdrücklich vor Augen geführt.

[19] Vgl. etwa den zusammenfassenden Überblick in Luce, Suppes (1965), Abschnitt 4.

[20] Natürlich verdiente, das ist mir klar, die subtile Methodologie psychologischer Tests eine detailliertere Diskussion.

Was also von theoretischer Seite aus zu geschehen hat – und daran arbeitet man ja intensiv, seit es die Entscheidungstheorie gibt –, ist, den Grundstock der Entscheidungstheorie, um den es hier immer nur ging, tüchtig aufzupäppeln. Dies kann erstens und vor allem dadurch geschehen, daß man mit Hilfe weiterer empirischer Gesetze die Menge der unterstellbaren Entscheidungsmodelle kräftig verringert. In bezug auf die unterstellbaren Entscheidungsfelder hieße das etwa, die empirische Beschaffenheit des vorhin erwähnten Aufmerksamkeitsoperators zu erhellen. Was die subjektiven Werte anlangt, so wären da auf empirischer Basis gewisse Formen subjektiver Wertfunktionen auszuschließen. Dies tun z.B. die Ökonomen, wenn sie das erste Gossensche Gesetz oder das Gesetz von der abnehmenden Grenzrate der Substitution[21] annehmen, stationäre Präferenzen untersuchen[22], etc. Dies tun die Psychologen, wenn sie verschiedene Triebe identifizieren, die dynamische Entwicklung der jeweiligen Triebstärke beschreiben, etwa in Form von Perioden, die mit einer triebreduzierenden Endhandlung enden, etc. Auch sind empirische Aussagen über wertmäßige Unabhängigkeit nützlich. Was schließlich die subjektiven Wahrscheinlichkeiten betrifft, so gibt es auch da eine Menge zusätzlicher empirischer Gesetze, nämlich erstens über die Einflüsse der Erfahrungen eines Subjektes auf seine Überzeugungen[23] und zweitens über das Gedächtnis, also darüber, wie gut die einmal gewonnenen Überzeugungen behalten werden. Beides wird ja in der Lerntheorie ausführlichst untersucht. All das trägt natürlich entscheidend zum empirischen Gehalt der Entscheidungstheorie bei.

Eine zweite wichtige Sache ist eine approximative Aufweichung der Entscheidungstheorie. Denn es ist eine nachgerade unzulässig starke Idealisierung anzunehmen, die in der Bayesschen Regel niedergelegte Mittelung oder Wägung funktioniere auf beliebig viele Stellen hinter dem Komma genau. Tatsächlich dürfte das meist nur eine sehr grobe Wägung sein, die zudem zahlreichen schwer feststellbaren und kontrollierbaren Zufallseinflüssen unterliegt. Und eben darauf gilt es sich einzustellen. Dem dienen erstens probabilistische Versionen der Entscheidungstheorie[24] und zweitens solche algebraische Versionen der Entscheidungstheorie, in denen die subjektiven Wahrscheinlichkeiten und Werte von Er-

---

[21] Vgl. etwa Krelle (1968), S. 16f.

[22] Wie es Koopmans (1960) tut.

[23] Diese haben wir ja schon im Abschnitt 4.2 etwas beschrieben.

[24] Für einen Überblick darüber s. Luce, Suppes (1965), Abschnitte 5-7.

eignissen oder Propositionen nicht auf einzelne reelle Zahlen festgelegt, sondern nur in mehr oder weniger breiten Intervallen eingegrenzt sind.[25] Diese Liste ließe sich leicht noch verlängern.

Hierher gehört auch die Behandlung all der – von einem strikten Standpunkt aus – Fehler, die bei nur grobschlächtigen Einschätzungen von Entscheidungssituationen typischerweise unterlaufen. Um nur drei davon zu nennen: Z.B. werden sehr unwahrscheinliche Ereignisse ständig völlig außer acht gelassen. Damit verwandt ist die Vergröberungsmethode, die wahrscheinlichste aller möglichen zukünftigen Entwicklungen, die dabei nicht einmal sehr wahrscheinlich zu sein braucht, als gegeben zu unterstellen und auf dieser Grundlage Entscheidungen zu treffen.[26] Schließlich ist es sehr beliebt, nach dem Motto „der Zweck heiligt die Mittel" über einer subjektiv wertmäßig herausragenden Sache alle sonst noch wertmäßig relevanten Dinge zu vernachlässigen. Natürlich ist es unfair, solche und andere Vergröberungsmethoden lediglich als Fehler zu betrachten; vielmehr werden sehr komplexe Entscheidungssituationen damit erst überschaubar und bewältigbar.[27]

Eine dritte Möglichkeit, die empirische Signifikanz der Entscheidungstheorie zu mehren, besteht darin, Anschluß an andere empirische Theorien herzustellen bzw. bestehende Anschlüsse auszubauen, und zwar am besten zu Theorien (eventuell geringeren Allgemeinheitsgrades), die schon auf eine breitere Erfahrungsbasis blicken können; natürlich eignen sich dazu vor allem Theorien, die zur Entscheidungstheorie begrifflich affin sind und möglichst ähnlich hohen Formulierungsstandards genügen wie die Entscheidungstheorie. Darunter fällt z.B. das am Ende von Abschnitt 4.1 erwähnte Projekt, die Spieltheorie nicht zur Entscheidungstheorie zu spezialisieren, sondern umgekehrt aus ihr zu gewinnen. Erst recht fallen darunter die vielfältigen Bemühungen der Ökonomen, etwa Gesetze des Marktes mikroökonomisch mit entscheidungstheoretischen Mitteln zu analysieren. Oder um zwei Beispiele aus der Psychologie zu wählen: Es dürfte keine allzu großen Schwierigkeiten bereiten, die TOTE-Hierarchien, die Miller, Galanter und Pribram (1973) erfunden und zur Grundlage ihrer Untersuchung

[25] S. etwa Suppes (1975).

[26] Diese Methode wird in der Politik anscheinend dauernd angewandt; diesen Eindruck muß jedenfalls der lediglich aus der Presse wohlinformierte Laie, wie ich es bin, bekommen.

[27] S. Gäfgen (1974), Kap. 9,II, Kap. 13,III, und Kap. 15,III.

gemacht haben, entscheidungstheoretisch nachzuzeichnen.[28] Und vor allem ist natürlich die mathematische Lerntheorie ein reizvoller, wenngleich vom Begrifflichen her nicht so willfähriger Vergleichsgegenstand.[29] Jedenfalls bin ich, ein bißchen demagogisch, guter Zuversicht, daß sich, wenn man Angefangenes fortführt und nicht Angefangenes anfängt, dabei die zentrale Stellung, die die Entscheidungstheorie im Bereich der Humanwissenschaften einnimmt, nur verfestigen kann.

---

[28] Eine Verbindung von den TOTE-Hierarchien zur Automatentheorie haben bereits Miller, Chomsky (1963), S. 485ff., geschlagen.

[29] Vgl. dazu die Ausführungen von Suppes (1969), S. 113f. und S. 139ff.

Literaturverzeichnis

Allais, M. (1953), „Le comportement de l'homme rationell devant le risque: critique des postulats et axiomes de l'ecole americaine“ *Econometrica* 21, 503-546.

Balch, M., und P.C. Fishburn (1974), „Subjective Expected Utility for Conditional Probabilities“, in: M. Balch, D.L. McFadden, S.Y. Wu (eds.), *Essays on Economic Behavior under Uncertainty*, Amsterdam u.a., S. 57-69.

Bauer, H. (1968), *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, Berlin.

Birkhoff, G. (1948), *Lattice Theory*, New York.

Blau, U. (1969), *Glauben und Wissen.* Eine Untersuchung zur epistemischen Logik, Dissertation München.

Blau, U. (1974), „Zur Situation der deontischen Logik“, *Papiere zur Linguistik* 6, 90-100.

Bolker, E.D. (1966), „Functions Resembling Quotients of Measures“, *Transactions of the American Mathematical Society* 124, 292-3l2.

Bühlmann, H., H. Loeffel und E. Nievergelt (1975), *Entscheidungs- und Spieltheorie*, Berlin u.a.

Carnap, R. (1947), *Meaning and Necessity*, Chicago, 2. erw. Aufl. 1956.

Carnap, R. (1950), *Logical Foundations of Probability*, Chicago.

Carnap, R. (1952), *The Continuum of Inductive Methods*, Chicago.

Carnap, R. (1971), „A Basic System of Inductive Logic, Part I“, in: R. Carnap, R.C. Jeffrey (eds.), *Studies in Inductive Logic and Probability*, Vol. I, Berkeley u.a., S. 33-165.

Chipman, J.S. (1960), „The Foundations of Utility“, *Econometrica* 28, 193-224.

Domotor, Z. (1972), „Causal Models and Space-Time-Geometries”, *Synthese* 24, 5-57.

Ellsberg, D. (1961), „Risk, Ambiguity, and the Savage Axioms”, *Quarterly Journal of Economies* 75, 645-669.

de Finetti B. (1964), „Foresight. Its Logical Laws, Its Subjective Sources“, in: H.E. Kyburg, H.E. Smokler (eds.), *Studies in Subjective Probability*, New York u.a., S. 93-158; engl. Übers. von '”La Prevision: Ses Lois Logique, Ses Sources Subjectives”, *Annales de L'Institut Henri Poincaré*, Bd. 7 (1937).

Fishburn, P.C. (1964), *Decision and Value Theory*, New York u.a.

Fishburn, P.C. (1968), „Utility Theory“, *Management Science* 14, 335-378.

Fishburn, P.C. (1970), *Utility Theory for Decision Making*, New York.

Fisher, M. (1962), „A System of Deontic-Alethic Modal Logic“, *Mind* 71, 231-236.

Gäfgen, G. (1974), *Theorie der wirtschaftlichen Entscheidung*, Tübingen, 3. Aufl.

Gärdenfors, P. (1975), „Qualitative Probability as an Intensional Logic“, *Journal of Philosophical Logic* 4, 171-185.

Gibbard, A., und W.L. Harper (1976), „Counterfactuals and Two Kinds of Expected Utility“, The Center for Mathematical Studies in Economics and Management Science, Northwestern University, Evanston, Ill., Discussion Paper No. 194.

Grice, H.P. (1957), „Meaning“, *Philosophical Review* 66, 377-388.

Hacking, I. (1965), *Logic of Statistical Inference*, Cambridge.

Hansson, B. (1971), „An Analysis of Some Deontic Logics“, in: R. Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings*, Dordrecht, S. 121-147.

Hilgard, E.R., und G.H. Bower (1975), *Theories of Learning*, Englewood Cliffs, 4. Aufl.

Hintikka, J. (1962), *Knowledge and Belief*, Ithaca.

Hintikka, J. (1971), „Some Main Problems of Deontic Logic“, in: R. Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings*, Dordrecht, S. 59-104.

Hintikka, J., und R. Hilpinen (1966), „Knowledge, Acceptance and Inductive Logic“, in: J. Hintikka, P. Suppes (eds.), *Aspects of Inductive Logic*, Amsterdam, S. 1-20.

Jeffrey, R.C. (1965), *The Logic of Decision*, New York u.a., 2. Aufl. 1967, dt. Übers.: *Logik der Entscheidungen*, Wien 1967.

Jeffrey, R.C. (1977), „A Note on the Kinematics of Preferences“, *Erkenntnis* 11, 135-141.

Jeffrey, R.C. (1978), „Higher Order Preferences“, im Erscheinen.

Koopmans, T.C. (1960), „Stationary Ordinal Utility and Impatience“, *Econometrica* 28, 287-309.

Krantz, D.H., R.D. Luce, P. Suppes und A. Tversky (1971), *Foundations of Measurement*, vol. I, New York, London.

Krelle, W. (1968), *Präferenz- und Entscheidungstheorie*, Tübingen.

v. Kutschera, F. (1969), „Der subjektive Wahrscheinlichkeitsbegriff“, *Synthese* 20, 84-103.

v. Kutschera, F. (1975), „Semantic Analyses of Normative Concepts“, *Erkenntnis* 9, 195-218.

Levi, I. (1967), „Probability Kinematics“, *British Journal for the Philosophy of Science* 18, 197-209.

Levi, I. (1975), „Newcomb's Many Problems“, *Theory and Decision* 6, 161-175.

Lewis, D. (1969), *Convention*, Cambridge, Mass.

Lewis, D. (1973), *Counterfactuals*, Oxford.

Lewis, D. (1976), „Probabilities of Conditionals and Conditional Probabilities“, *The Philosophical Review* 85, 297-315.

Luce, R.D. (1959), *Individual Choice Behavior*, New York.

Luce, R.D., und D.H. Krantz (1971), „Conditional Expected Utility“, *Econometrica* 39, 253-271.

Luce, R.D., und D.H. Krantz (1974), „The Interpretation of Conditional Expected-Utility Theories“, in: M. Balch, D.L. McFadden, S.Y. Wu (eds.), *Essays on Economic Behavior under Uncertainty*, Amsterdam u.a., S. 70-73.

Luce, R.D., und H. Raiffa (1957), *Games and Decisions*, New York u.a.

Luce, R.D., und P. Suppes (1965), „Preference, Utility, and Subjective Probability“, in: R.D. Luce, R.R. Bush, E. Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 3, New York, S. 249-410.

Meggle, G. (Hrsg.) (1977), *Analytische Handlungstheorie*, *Band* 1, *Handlungsbeschreibungen*, Frankfurt a.M.

Miller, G.A., und N. Chomsky (1963), „Finitary Models of Language Users“, in: R.D. Luce, R.R. Bush, E. Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 2, New York, S. 419-491.

Miller, G.A., E. Galanter und K.H. Pribram (1973), *Strategien des Handelns. Pläne und Strukturen des Verhalten*s, Stuttgart; dt. Übers. von: *Plans and the Structure of Behavior*, New York 1960.

Mine; H., und S. Osaki (1970), *Markovian Decision Processes*, New York.

Montague, R. (1970), „Pragmatics and Intensional Logic“, Synthese 22, 68-94.

v. Neumann, J., und 0. Morgenstern (1944), *Theory of Games and Economic Behavier*, Princeton; dt. Übers.: *Spieltheorie und wirtschaftliches Verhalten*, Würzburg 1967.

Nozick, R. (1969), „Newcomb's Problem and Two Principles of Choice“, in: N. Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel*; Dordrecht, S. 114-146.

Peleg, B., und M.E. Yaari (1973), „On the Existence of a Consistent Course of Action When Tastes are Changing“, *Review of Economic Studies* 40, 391-401.

Pfanzagl, J. (1967), „Subjective Probability Derived from the Morgenstern-von Neumann Utility Concept“, in: M. Shubik (ed.), *Essays in Mathematical Economies in Honor of Oskar Morgenstern*, Princeton, S. 237-251.

Pollak, R.A. (1968), „Consistent Planning“, *Review of Economic Studies* 35, 201-208.

Quine, W.V.O. (1951), „Two Dogmas of Empiricism“, *Philosophical Review* 60, 20-43.

Raiffa, H. (1973), *Einführung in die Entscheidungstheorie*, München; dt. Übers. von: *Decision Analysis*, Reading, Mass., 1968.

Reichenbach, H. (1956), *The Direction of Time*, Berkeley u.a.

Renyi, A. (1973), *Wahrscheinlichkeitsrechnung*, Berlin, 4. Aufl.

Rescher, N. (1964), *Hypothetical Reasoning*, Amsterdam.

Rescher, N. (1966), „Semantic Foundations for the Logic of Preference“, in: N. Rescher (ed.), *The Logie of Decision and Action*, Pittsburgh, S. 37-62.

Rohracher, H. (1971), *Einführung in die Psychologie*, Wien u.a., 10. Aufl.

Ryle, G. (1949), *The Concept of Mind*, London.

Salmon, W.C. (1975), „Confirmation and Relevance“, in: G. Maxwell, R.M. Anderson (eds.), *Minnesota Studies in the Philosophy of Science*, vol. VI, Minneapolis, S. 3-36.

Savage, L.J. (1954), *The Foundations of Statistics*, New York, 2. Aufl. 1972.

Savage, L.J. (1967), „Difficulties in the Theory of Personal Probability“, *Philosophy of Science* 34, 305-310.

v. Savigny, E. (1974), *Die Philosophie der normalen Sprache*, Frankfurt a.M., 2. Ausg.

Sneed, J. (1966), „Strategy and the Logic of Decision“, *Synthese* 16, 270-283.

Spohn, W. (1973), *Zur deontischen Logik und Wollenslogik*, Magisterarbeit München.

Spohn, W. (1977), „Where Luce and Krantz Do Really Generalize Savage's Decision Model“, *Erkenntnis* 11, 113-134

Spohn, W. (1980), „Stochastic Independence, Causal Independence, and Shieldability“, Journal of Philosophical Logic 9, 73-99.

Stegmüller, W. (1970), *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band II, Theorie und Erfahrung, 1. Halbband*, Berlin u.a.

Stegmüller, W. (1973a), *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band IV, Personelle und Statistische Wahrscheinlichkeit*, Berlin u.a.

Stegmüller, W. (1973b), *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band II, Theorie und Erfahrung, 2. Halbband*, Berlin u.a.

Strotz, R.H. (1955/56), „Myopia and Inconsistency in Dynamic Utility Maximation“, *Review of Economic Studies* 23, 165-180.

Suppes, P. (1966), „Concept Formation and Bayesian Decisions“, in: J. Hintikka, P. Suppes (eds.), *Aspects of Inductive Logic*, Amsterdam, S. 21-48.

Suppes, P. (1969), *Studies in the Methodology and Foundation of Science. Selected Papers from 1951 to 1969*, Dordrecht.

Suppes, P. (1970), *A Probabilistic Theory of Causality*, Amsterdam.

Suppes, P. (1975), „Approximate Probability and Expectation of Gambles“, *Erkenntnis* 9, 153-161.

Teller, P. (1976), „Conditionalization, Observation, and Change of Preference“, in: W.L. Harper, C.A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, ond Statistical Theories of Science*, vol. I, Dordrecht, S. 205-259.

Thomae, H. (1965), „Die Bedeutungen des Motivationsbegriffes“, in: H. Thomae (Hrsg.), *Handbuch der Psychologie*, *2. Band, Allgemeine Psychologie, Teil II, Motivation*, Göttingen, S. 3-44.

Tversky, A. (1975), „A Critique of Expected Utility Theory: Descriptive and Normative Considerations“, *Erkenntnis* 9, 163-173.

Wald, A. (1950), *Statistical Decision Functions*, New York.

v. Weizsäcker, C.C. (1971), „Notes on Endogenous Changes of Taste“, *Journal of Economic Theory* 3, 345-372.

v. Wright, G.H. (1951), „Deontic Logic“, *Mind* 60, 1-15.

v. Wright, G.H. (1963), *The Logic of Preference*, Edinburgh.

Yaari, M.E. (1977), „Endogenous Changes in Tastes: A Philosophical Discussion“, *Erkenntnis* 11, 157-196.

Namensverzeichnis

Sachverzeichnis

Kursive Seitenangaben weisen darauf hin, daß der fragliche Begriff an der betreffenden Stelle eingeführt oder definiert wird.

Symbolverzeichnis

Es wird die folgende mathematische und mengentheoretische Standardsymbolik verwandt:

R	die Menge der reellen Zahlen
$\varnothing$	die leere Menge
$x \in A$	das Ding x ist Element der Menge A
$\{x_1, ..., x_n\}$	die Menge aus den Elementen $x_1, ..., x_n$
$\{x \mid ... x ...\}$	die Menge aller Dinge x, für die ... x ...
$\langle x_1, ..., x_n \rangle$	das geordnete n-Tupel aus den Dingen $x_1, ..., x_n$
$A \subseteq B$	die Menge A ist (echte oder unechte) Teilmenge der Menge B
$A \subset B$	die Menge A ist echte Teilmenge der Menge B
$\mathrm{A} \cap \mathrm{B}$, $\bigcap_{r=1}^{n} A_r$, $\bigcap_{r \in R} A_r$	der Durchschnitt der Mengen A und B bzw. der Mengen $A_1, ..., A_n$ bzw. aller A_r ($r \in R$)
$A \cup B$, $\bigcup_{r=1}^{n} A_r$, $\bigcup_{r \in R} A_r$	die Vereinigung der Mengen A und B bzw. der Mengen $A_1, ..., A_n$ bzw. aller A_r ($r \in R$)
$A \setminus B$	die Differenz der Mengen A und B
$A \times B$	das kartesische Produkt der Mengen A und B
Pot(A)	die Potenzmenge der Menge A
$g \circ f$	die Komposition oder Hintereinanderausführung der Funktionen f und g
$P(.\mid A)$	die Funktion, die jedem (geeigneten) Ding B die Zahl $P(B \mid A)$ zuordnet

Im Kapitel 1 wird verwandt und an den angegebenen Stellen eingeführt:

Im Kapitel 2 wird verwandt:

C	die Menge möglicher Konsequenzen (eines Entscheidungsmodells)
H	die Menge möglicher Handlungen (eines Entscheidungsmodells)
P	die subjektive Wahrscheinlichkeitsfunktion (eines Entscheidungsmodells)
U	die erwartete subjektive Wertfunktion (eines Entscheidungsmodells)
V	die subjektive Wertfunktion (eines Entscheidungsmodells)
W	die Menge möglicher Umstände (eines Entscheidungsmodells)

In den Kapiteln 3- 5 wird verwandt und an den angegebenen Stellen eingeführt:

$\mathcal{A}$, $\mathcal{A}(J)$, $\mathcal{A}(i)$	S. 96
$\mathcal{F}$	S. 93f.
$\mathcal{H}$	S. 96
I, I^h, I^g	S. 96
Ind	S. 92
s, s_i	S. 98
T	S. 92, 97
W	S. 96
$\mathcal{Z}$, $\mathcal{Z}(J)$, $\mathcal{Z}(i)$	S. 96
$\leq, \geq, <, >$	S. 97
$J_{\leq t}, J_{\geq t}, J_{<t}, J_{>t}, J_{=t}$	S. 98
$F \approx_H G$	S. 97
$K \perp J / L$, $K \perp J$	S. 104, 105
$\perp$-Zerlegung	S. 107
$K \perp_{\mathrm{a}} J$	S. 117, 121
$\perp_{\mathrm{a}}$-Zerlegung	S. 124
$K \perp_{\mathrm{c}} J$	S. 114, 120
$\perp_{\mathrm{c}}$-Zerlegung	S. 124
$K \perp_{\mathrm{w}} L$	S. 128
$\perp_{\mathrm{w}}$-Zerlegung	S. 128
$\perp_{\mathrm{cw}}$-Zerlegung	S. 134
$[t]$	S. 114