

Die Gewichtung von Wahrscheinlichkeitsinformation in Abhangigkeit von deren Genauigkeit bei Entscheidungen unter Unsicherheit

Dissertationsschrift zur Erlangung des Doktorgrades des Fachbereichs 2 der
Universitat-Gesamthochschule Paderborn, vorgelegt im Dezember 2000 von

Johannes Honekopp

Müßiger Leser! Ohne Eidesschwur kannst du mir glauben,
dass ich wünschte, dieses Buch, als der Sohn meines Geistes,
wäre das schönste, stattlichste und geistreichste, das sich
erdenken ließe. Allein ich konnte nicht wider das Gesetz der
Natur aufkommen, in der ein jedes Ding seinesgleichen
erzeugt.

Cervantes

VORWORT

Auf der Suche nach einem konkreten Promotionsvorhaben, dass in dem Bereich angesiedelt sein sollte, wie Menschen Wahrscheinlichkeiten bzw. Grade subjektiver Sicherheit sprachlich ausdrücken, stieß ich auf eine andere Fragestellung, die ich in rührender Naivität mal eben mit einem kleinen Experiment zwischendurch beantworten wollte. Dieses Unterfangen wuchs sich dann zu der vorliegenden Arbeit aus. Hätte ich zu Beginn gewusst, mit welchen methodologischen Fall-stricken die Untersuchung dieses Themas behaftet ist, hätte ich vermutlich die Finger davon gelassen. Nicht in erster Linie, weil ich die Mühen bereue, die mir dieses Thema bereitet hat, sondern eher weil ich einen gut Teil derselben Lesern nicht ersparen kann. In meinen Augen sollte Wissenschaft einfach sein – diese Arbeit ist es leider nicht. Unter anderem deswegen, weil es notwendig war, eine Vielzahl neuer Konzepte und Größen einzuführen. Das Glossar möge Ihnen ein Ariadnefaden durch deren Labyrinth sein.

Wenn diese Arbeit ein glückliches Ende genommen hat, so verdanke ich dies auch den Menschen, die mich unterstützt haben. Mein besonderer Dank gilt (in alphabetischer Reihenfolge) Wolfgang Hell, Bianca Hönekopp, Romana Podgorsek, Frank Renkewitz, Peter Sedlmeier, Roland Stewen und Manfred Wettler.

Widmen möchte ich diese Arbeit meinen Eltern.

Münster, im Dezember 2000

Johannes Hönekopp

INHALT

0. ZUSAMMENFASSUNG	1
1. EINLEITUNG	3
2. EIN HISTORISCHER ABRISS DER THEORETISCHEN AUSEINANDERSETZUNGEN MIT ENTSCHEIDUNGEN UNTER UNSICHERHEIT: VOM BEGINN DER WAHRSCHEINLICHKEITSRECHNUNG ZUR EXPECTED UTILITY THEORIE . . .	7
2.1 Der Beginn der Wahrscheinlichkeitsrechnung (7)	
2.2 Die Erwartungswertregel und die Probleme mit dieser (8)	
2.3 Die Orientierung am erwarteten Nutzen als Entscheidungsstrategie (10)	
2.4 Probleme des klassischen Nutzenkonzeptes (11)	
2.5 Eine neue Fundierung des Nutzenkonzeptes – von Neumann und Morgenstern (13)	
2.6 Verschiedene Konzeptualisierungen von Wahrscheinlichkeit und deren Implikationen für die EU-Regel (14)	
2.7 Zusammenfassung (16)	
3. WEITERE THEORETISCHE ENTWICKLUNGEN: VON DER SUBJECTIVE EXPECTED UTILITY THEORIE ZUM KONZEPT DER ENTSCHEIDUNGSGEWICHTE	17
3.1 Die Messbarkeit subjektiver Wahrscheinlichkeit: Savages Subjective Expected Utility Theorie (17)	
3.2 Drei Phänomene, die Unzulänglichkeiten der SEUT exemplarisch aufzeigen (20)	
3.2.1 Ambiguität (20)	
3.2.2 Das Phänomen des Preference-Reversal und das Contingent Weighting Modell (22)	
3.2.3 Das Allais-Paradox (24)	
3.3 Das Konzept der Entscheidungsgewichte (25)	
3.4 Die Venture Theorie (28)	
3.5 Zusammenfassung (31)	
4. FRAGESTELLUNG	31
4.1 Die Gewichtungshypothese (31)	
4.2 Ambiguität – was ist das eigentlich genau? (32)	
4.3 Simulation I: Der Zusammenhang zwischen Erwartungswerten und Wahrscheinlichkeitsinformation bei zunehmender Ambiguität (33)	

4.4	Die Relevanz der Gewichtungshypothese (35)	
4.5	Zusammenfassung (36)	
5.	WAS BISHER GESCHAH	36
5.1	Die Studie von Gonzalez-Vallejo, Erev und Wallsten (1994) (36)	
5.2	Die Studie von Gonzalez-Vallejo und Wallsten (1992) (40)	
5.3	Zusammenfassung (42)	
6.	... WARUM DIES UNZUREICHEND IST,	43
6.1	Warum die Studie von Gonzalez-Vallejo, Erev und Wallsten (1994) nichts über die Gewichtungshypothese aussagt (45)	
6.2	Warum die Studie von Gonzalez-Vallejo und Wallsten (1992) nichts über die Gewichtungshypothese aussagt (48)	
6.3	Zusammenfassung (48)	
7.	... UND WIE MAN ES BESSER MACHEN KANN,	50
7.1	Das grundlegende Paradigma der vorliegenden Arbeit (50)	
7.2	Simulation II: Ist in diesem Paradigma $V(C, \epsilon)$ eine geeignete abhängige Variable? (51)	
7.2.1	Analyse zweier Strategien (53)	
7.3	Eine geeignete abhängige Variable (54)	
7.4	Zusammenfassung (55)	
8.	... WOFÜR ALLERDINGS ZWEI PROBLEME ZU LÖSEN SIND.	56
8.1	Simulation III: Welches Kriterium ist am geeignetsten, um θ zu schätzen? (56)	
8.2	Die Spezifikation von Basismodellen (58)	
8.2.1	Ein empirisch begründetes Basismodell (59)	
8.2.2	Theoriegeleitete und minimalistische Basismodelle (61)	
8.3	Zusammenfassung (64)	
9.	EXPERIMENT I	65
9.1	Materialien und Aufbau (65)	
9.2	Versuchsteilnehmer und Versuchsablauf (67)	
9.3	Ergebnisse (67)	
9.3.1	Ausschluss von Daten von der weiteren Analyse (67)	
9.3.2	Die Leistung der Vpn (68)	
9.3.3	Die Identifikation eines geeigneten Basismodells (72)	
9.3.4	Die relative Gewichtung der Wahrscheinlichkeit (0) (75)	
9.3.5	Eine alternative Überprüfung der Gewichtungshypothese (76)	

9.3.6 Wodurch wird die Leistung der Vpn determiniert? (78)	
9.4 Fazit und weiteres Vorgehen (81)	
10. EXPERIMENT II	84
10.1 Materialien und Aufbau (84)	
10.2 Versuchsteilnehmer und Versuchsablauf (85)	
10.3 Ergebnisse (85)	
10.3.1 Fehlende Daten und Ausschluss von Daten von der weiteren Analyse (85)	
10.3.2 Die Leistung der Vpn (86)	
10.3.3 Die Identifikation eines geeigneten Basismodells (88)	
10.3.4 Die relative Gewichtung der Wahrscheinlichkeit (θ) (92)	
10.3.5 Analyse des Einflusses von P_{Sicht} auf die Attraktivität der Lotterien (93)	
10.3.6 Bearbeitungszeiten (95)	
10.3.7 Gedächtnistest (96)	
10.3.8 Wodurch wurde die Leistung der Vpn determiniert? (97)	
10.4 Fazit und weiteres Vorgehen (98)	
11. EXPERIMENT III	100
11.1 Materialien (100)	
11.2 Aufbau (103)	
11.3 Versuchsteilnehmer und Versuchsablauf (105)	
11.4 Ergebnisse (106)	
11.4.1 Ausschluss von Daten von der weiteren Analyse (106)	
11.4.2 Die Leistung der Vpn (106)	
11.4.3 Die Identifikation eines geeigneten Basismodells (107)	
11.4.4 Die relative Gewichtung der Wahrscheinlichkeit (θ) (111)	
11.4.5 Bearbeitungszeiten (113)	
11.4.6 Wodurch wurden die Leistungen der Vpn bestimmt? (116)	
12. FAZIT AUS DEN DREI EXPERIMENTEN	117
13. WAS KOSTET AMBIGUITÄT, WAS KOSTEN SUBOPTIMALE ENTSCHEIDUNGSSTRATEGIEN? SIMULATION VIII	119
13.1 Ein realistisches Modell menschlichen Verhaltens – Simulation VII (120)	
13.2 Beschreibung der Simulation VIII (122)	
13.2.1 Berücksichtigte unabhängige Variablen (122)	
13.2.2 Design (125)	
13.3 Ergebnisse (125)	
13.3.1 Überblick über die Wirkung der unabhängigen Variablen (126)	
13.3.2 Andere Parameterwerte andere Ergebnisse? (132)	

13.3.3 Die Beantwortung der Kernfragen und aus diesen resultierende Implikationen (133)	
13.4 Die Übertragbarkeit auf natürliche Entscheidungssituationen (139)	
13.4.1 Fehlende „wahre Wahrscheinlichkeiten“ (140)	
13.4.2 Faktoren, die nicht berücksichtigt wurden (141)	
14. FAZIT UND AUSBLICK	143
ANHANG	145
GLOSSAR	156
LITERATUR	159

0. ZUSAMMENFASSUNG

Oftmals müssen Menschen Entscheidungen treffen, von denen sie nicht genau absehen können, welche Konsequenzen diese haben werden. Man nennt solche „Entscheidungen unter Unsicherheit“. Mit ihnen befasst sich die vorliegende Arbeit.

Theorien, die menschliches Verhalten in solchen unsicheren Situationen zu beschreiben suchen, legen einheitlich besonderen Wert auf zwei Aspekte. Diese sind, vereinfacht gesprochen, in zwei Fragen zu fassen: „Wie erstrebenswert wird meine Situation sein, falls meine Entscheidung zum erhofften Resultat führt?“ und „Wie groß ist die Chance dafür, dass das angestrebte Ergebnis eintritt?“. Die Erfolgswahrscheinlichkeit, der zweite Aspekt, kann dabei mehr oder minder genau bezifferbar sein. So lässt sich bei einer Wette auf einen Münzwurf die Gewinnwahrscheinlichkeit präzise angeben, bei einer Wette auf ein Fußballspiel nicht.

Die vorliegende Arbeit untersucht, inwieweit die Genauigkeit der beteiligten Wahrscheinlichkeiten Entscheidungen unter Unsicherheit beeinflusst. So erscheint es zunächst plausibel anzunehmen, dass diffuse Wahrscheinlichkeiten dazu führen, dass sich die Entscheidungen stärker auf den erstgenannten Wertaspekt stützen (Gewichtungshypothese). Tatsächlich wird diese Meinung in der Literatur vertreten (Wallsten, Budescu & Tsao, 1997). Die Untersuchungsmethodik, auf Grund derer die Autoren zu diesem Schluss kamen, ist jedoch, wie hier gezeigt wird, unbrauchbar. Eine angemessene Methodik wird entwickelt.

Das experimentelle Paradigma sieht dabei so aus, dass die Vpn sechs gleichzeitig präsentierte Lotterien, bei denen entweder ein Gewinn ϵ mit einer Wahrscheinlichkeit P gewonnen wird oder andernfalls nichts passiert, gemäß ihrer Attraktivität in eine Rangfolge bringen müssen. Dabei wird in drei Experimenten die Genauigkeit der Wahrscheinlichkeitsinformation variiert. Dies geschieht in der Regel dadurch, dass die Gewinnwahrscheinlichkeiten in Form eines Glücksrades dargestellt sind – größere Gewinnflächen entsprechen dabei höheren Gewinnwahrscheinlichkeiten; bei einem Teil der Lotterien ist ein Stück des Glücksrades derart verdeckt, dass sich die Größe der Gewinnfläche nicht mehr genau einschätzen lässt. In Experiment III wird alternativ die Genauigkeit der Wahrscheinlichkeitsinformation auch dadurch herabgesetzt, dass die Gewinnwahrscheinlichkeiten in Worten beschrieben werden (etwa „eher geringe Wahrscheinlichkeit“), die nur einen ungenauen Rückschluss auf die tatsächlichen Wahrscheinlichkeitswerte erlauben.

Aus den von den Vpn hergestellten Attraktivitätsrangfolgen lässt sich schätzen, wie stark sie sich bei ihrer Entscheidung auf die Wert- und wie stark auf die Wahrscheinlichkeitsinformation stützen.

Die Gewichtungshypothese kann nicht bestätigt werden; im Gegenteil stützen sich die Entscheidungen bei Gewinnaufgaben mit abnehmender Informationsgenauigkeit verstärkt auf die Wahrscheinlichkeitsinformation; eine schwache gegenläufige Tendenz bei Verlustraufgaben deutet darauf hin, dass der Grund hierfür darin liegt, dass die Vpn bei sinkender Informationsgenauigkeit insgesamt vorsichtiger entscheiden.

Mit Hilfe einer Computersimulation, die eine Vielzahl unterschiedlicher Randbedingungen berücksichtigt, wird geklärt, welche Folgen für Entscheidungen unter Unsicherheit dieser Effekt im speziellen und mangelnde Informationsgenauigkeit im allgemeinen hat. Es zeigt sich dabei, dass der

gefundene Effekt von geringer praktischer Relevanz ist. Aus der Simulation ergeben sich jedoch ein Reihe wichtiger Ergebnisse und Schlussfolgerungen: (i) Entscheidungen unter Unsicherheit sind äußerst robust; sowohl unpräzise Information wie auch suboptimales Entscheidungsverhalten beeinträchtigen die Entscheidungsgüte nur geringfügig. (ii) Eine einfache Entscheidungsstrategie führt zu sehr guten Ergebnissen, wenn diese konsequent befolgt wird; Trainingsmaßnahmen sollten Entscheider daher weniger dazu ermutigen, eine besonders ausgefeilte und komplizierte Strategie zu entwickeln, als vielmehr einer Strategie mit handhabbarer Komplexität konsequent treu zu bleiben. (iii) Bemühungen, die Präzision von Wahrscheinlichkeitsinformation zu vergrößern, haben einen steil abfallenden Grenznutzen. (iv) Der Grad der Präzision der Wahrscheinlichkeitsinformation sollte bei der Entscheidung in der Regel nicht berücksichtigt werden.

1. EINLEITUNG

Hier wird erklärt, was unter Entscheidungen unter Unsicherheit, die Gegenstand der vorliegenden Arbeit sind, zu verstehen ist. Darauf hinaus wird ein Klassifikationsschema vorgestellt, anhand dessen sich die Mannigfaltigkeit des Phänomens „Entscheidungen“ ordnen lässt.

Menschen haben die Möglichkeit Entscheidungen zu treffen – und sie sind gleichzeitig dazu gezwungen. Bei einem Teil solcher Entscheidungen ist ein wesentliches Bestimmungsstück, dass die Konsequenzen der Entscheidung nicht eindeutig kalkulierbar sind. In einem solchen Fall spricht man von *Entscheidungen unter Unsicherheit* – diese sind Gegenstand der vorliegenden Arbeit.

Ein Beispiel mag illustrieren, was damit gemeint ist: Denken wir uns ein junges Paar, das nach einer stressigen Zeit zehn Tage im Mai Urlaub in sonniger Beschaulichkeit verbringen möchte. Der Urlaub soll in zwei Wochen beginnen. Es besteht die Möglichkeit, ein elterliches Ferienhaus auf Rügen zu nutzen. Da die beiden Inselliebhaber sind und ihre finanzielle Situation zudem angespannt ist, wäre diese Lösung eigentlich optimal – falls das Wetter mitspielt. Durch Zufall kommt der Frau nun ein einmalig günstiges Angebot für zehn Tage Lanzarote in die Hände – als Insel für die beiden ähnlich schön wie Rügen, nicht so billig wie dieses aber mit Schönwettergarantie. Gebucht werden muss sofort. Wie sollten sie sich entscheiden? Die Schwierigkeit dieser Entscheidung gründet sich genau auf die nicht sicher zu kalkulierende Komponente – das Wetter auf Rügen.

In diesem Beispiel röhrt die Unsicherheit daher, dass sich ein zukünftiger Zustand nur schwer prognostizieren lässt. Quelle der Unsicherheit kann aber etwa auch die Lückenhaftigkeit des eigenen Wissens sein: Wenn Sie schätzen sollen, ob Augsburg oder Braunschweig mehr Einwohner hat, und Ihnen für die richtige Antwort 100,— DM winken, so kann die Antwort prinzipiell gewusst werden (im Gegensatz dazu ist das Maiwetter auf Rügen per se unbekannt). Dennoch bin ich mir sicher, dass dies für Sie eine Entscheidung unter Unsicherheit darstellt. Einen Überblick über mögliche Quellen von Unsicherheit geben Kahneman & Tversky (1982).

Man kann argumentieren, dass alle Entscheidungen solche unter Unsicherheit sind – schließlich kann auch die lapidarste Wahl ungeahnte Folgen haben: Hätte ich gestern in der Pommesbude Mayonnaise anstelle von Ketchup genommen, wäre diese vielleicht schlecht gewesen, die anschließende Lebensmittelvergiftung hätte mich ins Krankenhaus gebracht, wo dann möglicherweise eine Geschäftspartnerschaft mit einem jungen Arzt ihren Beginn hätte nehmen können, die schließlich dazu führen würde, dass ich Teilhaber einer neugegründeten Firma für Biotechnologie würde, deren kometenhafter Aufstieg den von Microsoft würde verblassen lassen. Vielleicht bin ich also ganz knapp daran vorbeigeschlittert eines Tages reichster Mann der Welt zu sein. Aber alle diese Möglichkeiten spielten zum Zeitpunkt der Wahl keine Rolle (glücklicherweise: denn hätte ich versucht alle Unwägbarkeiten abzuwägen, hätte ich bei Geschäftsschluss immer noch nicht gewusst, was ich nehmen soll). Obschon die Entscheidung also in einem gewissen Sinne Unsicherheit barg, war diese doch für den Entscheider nicht relevant und somit „psychologisch“ (sowohl aus Sicht des Entscheiders wie auch dessen, der diese Entscheidung zu beschreiben und verstehen sucht) irrelevant.

Mit Entscheidungen unter Unsicherheit beschäftigt sich eine Reihe wissenschaftlicher Disziplinen. Besondere Nähe zur psychologischen Entscheidungsforschung besteht dabei seitens der Statistik, die einen normativen Ansatz verfolgt – wie sollte man generell mit Unsicherheit umgehen? konkret zumeist: wie soll man Daten sammeln, um Hypothesen zu überprüfen; wie soll man Hypothesen im Lichte bestehender Daten einschätzen? – sowie zur Ökonomie, die sowohl das deskriptive Anliegen hat, das Entscheidungsverhalten von Konsumenten, Investoren und anderen Wirtschaftssubjekten zu beschreiben, als auch ein präskriptives, insoweit sie sich bemüht, Entscheidungsträgern Hilfsmittel an die Hand zu geben.

Entscheidungen sind vielgestaltig und können in einer Vielzahl von Aspekten höchst verschieden voneinander sein. Schneisen der Ordnung in dieses Dickicht der Mannigfaltigkeit schlägt eine Taxonomie, die Jungermann, Pfister & Fischer (1998) entwickelt haben. Sie charakterisieren Entscheidungen anhand dreier Dimensionen, die jedoch nicht als voneinander völlig unabhängig anzusehen sind: (i) Art und Umfang des erforderlichen kognitiven Aufwands, (ii) Merkmale von Entscheidungssituationen und (iii) verwendete Entscheidungsstrategien.

Zunächst stellt sich die Frage, ob nicht *alle* Handlungen eine Entscheidung repräsentieren – schließlich hätte man stets auch anders gekonnt. Ohne eine genaue Grenze angeben zu wollen, ab wann wir von einer Entscheidung sprechen sollten, sei darauf hingewiesen, dass zu einer Entscheidung ein Entscheidungsbewusstsein gehört und ein zumindest minimaler kognitiver Aufwand. Ob in einem Einzelfall eine Entscheidung vorliegt oder nicht, wird also nicht durch äußere Faktoren bestimmt (hier ist allein die Existenz von Alternativen wichtig), sondern durch deren innere Repräsentation und durch den Umgang mit der Situation.

Kognitiver Aufwand

Jungermann et al. unterscheiden hier vier Arten von Entscheidungen. Auf der untersten Stufe *routinierter* Entscheidungen handelt es sich lediglich um den immer wiederkehrenden Nachvollzug einer früher einmal getroffenen Entscheidung. Dies wäre etwa dann der Fall, wenn sich eine vegetarisch lebende Studentin in der Mensa automatisch in die Schlange für das fleischlose Gericht einreihet. Diese Entscheidung verlangt praktisch nur ein Matching der gegebenen Situation auf eine allgemeine Entscheidungsregel. Der kognitive Aufwand ist somit minimal, die Entscheidungszeit verschwindend gering, Aufmerksamkeit wird fast nicht benötigt. Das entgegengesetzte Ende des Anforderungskontinuums markieren *konstruktive* Entscheidungen. Hier denkt der Entscheider explizit über seine Präferenz nach, Informationen müssen aktiv gesucht und Entscheidungsalternativen selbstständig generiert werden. Weiterhin ist es nötig, sich Klarheit über die eigenen persönlichen Werte zu machen. Entsprechend vielgestaltig und hoch sind die kognitiven Anforderungen hierfür. Ein Beispiel für eine solche Entscheidung kann etwa die Berufswahl sein.

Merkmale von Entscheidungssituationen

Außer nach Art und Umfang des kognitiven Aufwandes lassen sich Entscheidungen auch anhand von Merkmalen der Entscheidungssituation klassifizieren. Zu diesen gehört, ob die Menge der Optionen geschlossen oder offen ist. Ersteres ist der Fall, wenn Sie im Supermarkt eine Tafel Vollmilchschokolade kaufen wollen – das Angebot ist begrenzt und gut überschaubar. Anders verhält es sich, wenn Sie überlegen, was Sie einer Freundin zum Geburtstag schenken – hier gilt es, erst einmal Optionen zu generieren. (Wie wir dies tun, ist ein Aspekt, der von der Entscheidungsforschung bislang kaum thematisiert worden ist. In Experimenten sind die Optionen, zwischen denen gewählt werden soll, quasi immer vorgegeben. Sinnvolle Verfahren zur Alternativengenerierung stellen Eisenführ und Weber, 1993, vor). Weiter können Entscheidungen danach unterschieden werden, ob sie ein- oder mehrstufig sind. Dem Kauf einer Tafel Schokolade liegt in der Regel eine einstufige Entscheidung zu Grunde, bei einer Flasche Wein kann dies anders aussehen: Prüfung von Etikett und Preis führen vielleicht zunächst zu dem Entschluss, dass man den Wein probieren möchte, das Resultat dieser Probe führt dann erst zu Kauf oder nicht Kauf. Jungermann et al. folgend sei noch ein weiteres Merkmal von Entscheidungen erwähnt, welches darin besteht, dass sie einmalig sein oder wiederholt auftreten können. Stellen Sie sich etwa zwei Gangster vor, die sich an einem entlegenen Ort treffen, um einen Deal abzuwickeln, etwa den Verkauf gestohlener Diamanten. Jeder steht nun vor der Entscheidung, das Geschäft einfach durchzuziehen, oder aber den anderen niederrzuschießen und somit in den Besitz von Schmuck *und* Geld zu kommen. Stehen noch weitere ähnliche Geschäfte zwischen den beiden an, so ist der Anreiz für kooperatives Verhalten sicherlich höher, als wenn es sich um eine einmalige Begegnung handelt.

Entscheidungsstrategien

Dass auch Entscheidungen ohne Unsicherheit schwierig sein können, dürfte jedermann aus eigener Erfahrung bekannt sein. So mögen etwa zwei gleichzeitig bestehende Stellenangebote, der gutbezahlte Traumjob in Hintertupfingen und die akzeptable aber wenig Begeisterung weckende Offerte in der Stadt, an die man vielfältig gebunden ist, einen vor eine schwierige Wahl stellen. Die Schwierigkeit wird hier dadurch verursacht, dass sich die beiden Optionen hinsichtlich einer Vielzahl von Attributen unterscheiden, die untereinander schwer zu vergleichen sind. So müssen etwa die Qualität des Lebensumfeldes, die Höhe des Einkommens, die berufliche Zufriedenheit und die Nähe zu den Menschen, die einem wichtig sind, gegeneinander abgewogen werden. Diese Schwierigkeit rückt eine dritte Dimension, anhand derer sich Entscheidungen charakterisieren lassen, in den Blickpunkt: die Wahl einer Entscheidungsstrategie. Wenn wir davon ausgehen, dass für die Entscheidung der wahrgenommene Nutzen der Optionen relevant ist, werden dadurch zwar eine Menge möglicher Entscheidungsprinzipien ausgeklammert (denkbar wären auch: eine Entscheidung per Zufall; nach einer festen, von der momentanen Situation unabhängigen Regel verfahren; die Entscheidung treffen, die den Status quo am wenigsten verändert etc.), es bleibt aber noch immer eine Vielzahl möglicher Entscheidungsstrategien bestehen.

Eine denkbar einfache und befriedigende ist die *Dominanzstrategie*: Wähle die Option, die auf allen Attributen mindestens so gut ist wie alle anderen Optionen und auf mindestens einem Attribut besser. Diese Regel hat leider den Haken, dass sie oftmals nicht anwendbar ist. Entscheidungen, die als schwierig empfunden werden, zeichnen sich gerade dadurch aus, dass keine dominierende Option vorliegt – so auch im obigen Beispiel. Eine Entscheidungsstrategie, die auch im obigen Beispiel anwendbar wäre, ist etwa die *lexikographische*: Dabei wird diejenige Option gewählt, die auf dem wichtigsten Attribut den besten Wert hat. Sind diesbezüglich alle Wahlmöglichkeiten gleich, so wird das zweitwichtigste Attribut betrachtet usw. Ein wesentliches Merkmal der lexikographischen Regel ist, dass sie *nicht-kompensatorisch* ist, d. h. die Schwäche, die eine Option auf einem wichtigen Attribut aufweist, lässt sich auch durch noch so günstige Ausprägungen weniger wichtiger Attribute nicht kompensieren. Würde also das oben beschriebene Problem lexikographisch gelöst und erwiese sich die Erreichbarkeit vom jetzigen Wohnort als wichtigstes Attribut zur Bewertung von Stellenangeboten, so würde die Wahl auf jeden Fall gegen Hintertupfingen ausfallen, selbst wenn der dortige Arbeitgeber seine Gehaltsofferte um 50% steigern würde. Ein anderes wichtiges Beispiel für eine nicht kompensative Strategie ist *Elimination by Aspects*: Grundgedanke ist hier, dass es für jedes Attribut einen Schwellenwert gibt, der nicht unterschritten werden darf. So muss etwa, zumindest für die meisten Menschen, jede beliebige Arbeit einen Mindestverdienst sicherstellen. Elimination by Aspects bedeutet nun, dass für alle Optionen zunächst anhand des wichtigsten Attributs geprüft wird, ob der spezifische Schwellenwert überschritten wird. Optionen, die diesen nicht erreichen, werden aussortiert, die verbleibenden anhand des nächsten Attributes untersucht usw. Gewählt wird die Option, die zuletzt übrigbleibt.

Ein Beispiel für eine sehr komplexe und voraussetzungsreiche *kompensatorische* Strategie besteht darin, nach dem *multiattributiven Nutzen* zu entscheiden. Hierzu ist es erstens notwendig, die einzelnen Attribute zu gewichten („die Höhe des Einkommens ist mir doppelt so wichtig wie der Freizeitwert der Stadt“) und zweitens, die Ausprägung eines jeden Attributes einer jeden Option auf einer Intervallskala zu bewerten („wenn ich den Freizeitwert von Hintertupfingen mit 1 bewerte, dann ist der der anderen Stadt 7“). Für jede Option wird nun für jedes Attribut das Produkt aus der Attributausprägung und der Attributgewichtung gebildet, und für jede Option wird dann die Summe dieser Partialnutzenwerte ermittelt. Die Wahlmöglichkeit mit der höchsten Summe gewinnt.

Eine wichtige Untergruppe der nicht-kompensatorischen Strategien stellen die sogenannten *satisficing* Regeln dar (ein Kunstwort, das aus *satisfy* und *suffice* zusammengesetzt ist). Hierbei werden die Optionen in unsystematischer Reihenfolge geprüft. Diese Prüfung wird abgebrochen, sobald eine Option gefunden ist, die einem gewissen Kriterium Genüge tut. Der Vorteil eines solchen Vorgehens besteht darin, dass es mit geringem Aufwand zu einem akzeptablen Ergebnis führt. Generell handelt es sich bei der Wahl einer Entscheidungsstrategie auch um ein Entscheidungsproblem; bei diesem muss der Aufwand, der mit der Anwendung einer Regel verbunden ist, gegen den Grad der Ausbeute der gegebenen Informationen abgewogen werden.

Einen guten Überblick mit einer Fülle von Entscheidungsstrategien geben Jungermann, Pfister & Fischer (1998). Die verschiedenen Strategien schließen einander nicht zwingend aus. Existiert eine große Menge an Optionen – wir begegnen ihr etwa im Reisebüro oder im Buchladen – so bietet es

sich an, die Zahl der Optionen zunächst mit einer nicht-kompensatorischen Regel zu reduzieren, was wenig Aufwand erfordert, und die wenigen verbleibenden Alternativen dann eingehender anhand einer kompensatorischen Strategie zu prüfen.

Zusammenfassung: Entscheidungen lassen sich sowohl anhand von Situations- als auch anhand von Verhaltensmerkmalen kategorisieren. Entscheidungen unter Unsicherheit zeichnen sich dadurch aus, dass deren Konsequenzen aufgrund eines Mangels an Wissen nicht klar voraussehbar sind. Dieser kann aus einer Wissenslücke des Entscheiders resultieren oder auch prinzipieller Natur sein. Letzteres ist etwa dann der Fall, wenn der Wissensmangel aus einer grundsätzlich nicht zu überbrückenden Unkenntnis der Zukunft resultiert.

2. EIN HISTORISCHER ABRISS DER THEORETISCHEN AUSEINANDERSETZUNGEN MIT ENTSCHEIDUNGEN UNTER UNSICHERHEIT: VOM BEGINN DER WAHRSCHEINLICHKEITSRECHNUNG ZUR EXPECTED UTILITY THEORIE

Die Kapitel 2 und 3 geben einen chronologischen Überblick über Theorien der Entscheidung unter Unsicherheit. Diese sind teilweise normativer Natur, versuchen also festzulegen, welches Verhalten angesichts Unsicherheit vernünftig ist, und teilweise deskriptiver Natur, indem sie nämlich zu beschreiben suchen, wie sich Menschen in solchen Situationen tatsächlich verhalten.

2.1 Der Beginn der Wahrscheinlichkeitsrechnung

Die ersten theoretischen Auseinandersetzungen mit Entscheidungen unter Unsicherheit, die noch heute Relevanz haben, finden sich im 16. Jahrhundert. Erstmals wird Unsicherheit formalisiert, indem nämlich die Wahrscheinlichkeitsrechnung entsteht. Diese kann sich jedoch zunächst nur einem kleinen Teil der Wirklichkeit zuwenden, Glücksspielen. Der Spieler Cardano analysierte, wie wahrscheinlich verschiedene Augensummen bei einem Wurf mit drei Würfeln sind und formulierte dabei, was heute als die klassische Definition der Wahrscheinlichkeit bezeichnet wird, dass nämlich Wahrscheinlichkeit durch die Proportion gleichwahrscheinlicher Ausgänge bestimmt ist (Good, 1959). Wenn also ein ungezinkter Würfel einmal geworfen wird, so gibt es sechs gleichwahrscheinliche Ausgänge. Diese werden auch als *Elementarereignisse* bezeichnet. In der Regel interessiert uns jedoch die Wahrscheinlichkeit eines *Ereignisses*. Dieses kann durch verschiedene Elementarereignisse verwirklicht werden – das Ereignis „gerade Zahl“ etwa durch die Elementarereignisse *zwei, vier oder sechs*. Die Wahrscheinlichkeit eines bestimmten Ereignisses wie „gerade Zahl“ ergibt

sich als Verhältnis der Zahl der Elementarereignisse, die das Ereignis herbeiführen, zur Gesamtzahl der Elementarereignisse. Somit ergibt sich hierfür also eine Wahrscheinlichkeit von $3/6 = .5$.

Am Rande sei bemerkt, dass die klassische Konzeptualisierung von Wahrscheinlichkeit nicht unproblematisch ist, denn sie setzt das Wissen um die Gleichwahrscheinlichkeit der Elementarereignisse voraus. Wie kommen wir aber zu diesem Wissen? Die Gleichwahrscheinlichkeit wird aufgrund des *Prinzips des unzureichenden Grundes* postuliert. Dieses besagt, dass die gesamte Wahrscheinlichkeitsmasse gleichmäßig auf alle Optionen verteilt werden sollte, wenn kein Grund vorliegt, dies nicht zu tun. Die Anwendung dieses Prinzips kann jedoch zu Schlüssen führen, die für uns nicht akzeptabel sind. So ist zumindest hypothetisch möglich, dass im Mond ein Männchen sitzt, das sich von Fruchtdrops ernährt. Die Wahrscheinlichkeit hierfür mit $\frac{1}{2}$ anzugeben (andere Alternative: es gibt kein solches Männchen), mutet bereits absurd an. Noch aberwitziger ist, dass wir die Wahrscheinlichkeit dafür, dass sich dort überhaupt ein Männchen befindet, mühelos steigern können, indem wir uns weitere mögliche Alternativen denken: ein Männchen, das sich von Wurstbroten ernährt, ein Männchen, das sich von marsianischen Buchhändlerinnen ernährt, usw. Gegen eine solche Argumentation hält sich das Prinzip des unzureichenden Grundes zwar ein Hintertürchen offen („... wenn kein Grund vorliegt dies nicht zu tun“); ist man aber bereit dieses anzuerkennen, so wird das Prinzip dahingehend unbefriedigend, dass nicht definiert werden kann, wann denn ein solches Entschlüpfen angebracht ist und wann nicht.

Mathematische Betrachtungen von Wahrscheinlichkeit waren lange fast ausschließlich mit Glücksspielen befasst, da hier aufgrund der Gleichförmigkeit von Würfeln, Spielkarten, Losen usw. Elementarereignisse existierten, denen a priori spezifische Wahrscheinlichkeiten zugeschrieben werden konnten. Andere Arten von Wahrscheinlichkeit, eine kommt in der Formulierung „es ist sehr unwahrscheinlich, dass es Leben auf dem Mond gibt“ zum Ausdruck, können von der klassischen Konzeption nicht erfasst werden. Diese Beschränkung wurde, wie wir noch sehen werden, erst später überwunden.

2.2 Die Erwartungswertregel und die Probleme mit dieser

Von Anfang an waren Überlegungen zur Wahrscheinlichkeit mit Überlegungen bezüglich Handlungen verknüpft. Wenn zwei Spieler mit einem Wurf um zehn Taler würfeln, der erste eine zwei gewürfelt hat und das Spiel dann abgebrochen werden muss, wie sollte das Geld dann aufgeteilt werden? Wieviel sollte man bereit sein, für eine Wette zu bezahlen, bei der man einen Taler gewinnt, wenn man eine vorher bestimmte Zahl würfelt? Sollte man eine Wette annehmen, bei der man zwei Goldstücke gewinnt, wenn man mit zwei Würfeln mindesten neun Augen würfelt, andernfalls aber ein Goldstück zahlen muss? Solche Fragen wurden mit Hilfe des Konzeptes des *Erwartungswertes* (abgekürzt EV für „expected value“) entschieden. Dieser ist für jede Handlungsalternative definiert als

$$EV = \sum (p_i \cdot \epsilon_i)$$

wobei p_i die Wahrscheinlichkeit des i ten Handlungsausgangs bezeichnet und ϵ_i dessen Wert. Verluste werden durch ein negatives Vorzeichen gekennzeichnet.

Die Erwartungswertanalyse sieht für das letztgenannte Problem so aus: Die Chance, mit zwei Würfeln mindestens eine neun zu werfen beträgt $10/36$ oder etwa .278. Der Erwartungswert für die oben beschriebene Wette beträgt also $.278 \cdot 2$ Goldstücke $- .722 \cdot 1$ Goldstück = $-.166$ Goldstücke. Die Erwartungswertregel sagt, dass man sich stets für die Option mit dem höchsten Erwartungswert entscheiden sollte (zumindest unter der Voraussetzung, dass man lieber mehr als weniger Geld, Kühle, Mäntel, oder worum man sonst gerade spielt, hat). Da die vorliegende Wette einen negativen Erwartungswert hat, nicht zu Wetten aber den Erwartungswert null (weil sich hierbei keine Veränderung ergibt), sollte man also dieser Regel folgend die Wette nicht annehmen. Die Bestimmung des korrekten Preises für ein Glücksspiel folgt dem gleichen Prinzip; dieser gleicht dem Erwartungswert des Spiels.

Ein Vorteil der EV-Regel (sowie auch nachfolgender Abwandlungen derselben) besteht darin, dass diese auf einfache Weise gestattet, verschiedene Handlungsalternativen gleichzeitig bezüglich grundverschiedener Aspekte (Geld und Wahrscheinlichkeit) zu vergleichen.

Warum diese Verknüpfung multiplikativ sein sollte, ergibt sich aus dem *Gesetz der großen Zahl*, das in seinen Grundzügen von Jacob (teilweise auch „James“) Bernoulli (1654 – 1705) formuliert wurde. Es besagt sinngemäß Folgendes: Wenn ein Zufallsexperiment, bei dem ein bestimmtes Ereignis jedesmal die Auftretenswahrscheinlichkeit p hat, n mal wiederholt wird, dann wird bei großem n dieses Ereignis sehr wahrscheinlich annähernd pn mal eintreffen. Oder anders gesagt, die relative Auftretenshäufigkeit wird sehr wahrscheinlich der Wahrscheinlichkeit des Ereignisses sehr ähnlich sein. Wenn eine Münze also oft geworfen wird, so ist davon auszugehen, dass der relative Anteil der Kopf-Würfe annähernd 50% beträgt, was genau der Wahrscheinlichkeit des Ereignisses „Kopf“ entspricht. So beträgt z. B. nach zwei Würfen die Chance für das Ereignis „nur Kopf oder nur Zahl“ $\frac{1}{2}$. In diesem Fall liegt der Anteil der Kopf-Würfe bei 0% oder 100% und ist somit maximal verschieden von der Wahrscheinlichkeit für einen Kopf-Wurf, die $\frac{1}{2}$ beträgt. Nach fünf Würfen beträgt die Wahrscheinlichkeit für diese extreme Abweichung jedoch nur noch $1/16$. Der im Gesetz der großen Zahl beschriebene Sachverhalt führt dazu, dass die Erwartungswertregel auf lange Sicht den eigenen Gewinn maximiert.

Das Gebot, die eigenen Handlungen am Erwartungswert der gegebenen Optionen zu orientieren, kann jedoch in Schwierigkeiten führen. So würde es sich demnach verbieten, Versicherungen abzuschließen. Denn diese haben für den Versicherungsnehmer einen negativen Erwartungswert, da die Versicherung insgesamt immer weniger an Zahlungen ausschüttet als sie an Prämien einnimmt (davon lebt das Unternehmen schließlich). Noch deutlicher veranschaulicht die Problematik der EV-Regel das sogenannte Petersburg-Paradox. Stellen Sie sich vor, dass Ihnen das folgende Glücksspiel zum Kauf angeboten wird. Eine Münze wird so lange geworfen, wie Kopf fällt, bei Zahl endet das Spiel. Sie bekommen 2^n DM ausgezahlt, wobei n die Zahl der Kopfwürfe ist. Fällt also nur einmal Kopf, erhalten Sie zwei Mark, bei zwei Köpfen vier Mark, bei drei Köpfen acht Mark usw. Wieviel wäre es Ihnen wert, dieses Spiel spielen zu dürfen? Vermutlich nur einen relativ geringen Betrag.

Der EV-Regel folgend sollten Sie jedoch bereit sein, für dieses Spiel ihr ganzes Vermögen einzusetzen, da der Erwartungswert dieses Spieles nämlich *unendlich* groß ist! Denn er ist das Ergebnis einer unendlich langen Summe, bei der jeder Summand eine Mark beträgt: $EV = \frac{1}{2} \cdot 2 \text{ DM} + \frac{1}{4} \cdot 4 \text{ DM} + \frac{1}{8} \cdot 8 \text{ DM} \dots$

2.3 Die Orientierung am erwarteten Nutzen als Entscheidungsstrategie

Solche Überlegungen führten Daniel Bernoulli dazu (1700–1782), das Konzept des Erwartungswertes als Handlungsanleitung zu verwerfen. Statt dessen versuchten, so Bernoulli, vernünftige Leute die Erwartung des „moralischen Wertes“ zu maximieren. An Stelle dieses Begriffs hat sich hier jedoch der Ausdruck des *erwarteten Nutzens* (abgekürzt EU für „expected utility“) durchgesetzt. Von seiner Intuition ausgehend legte Bernoulli dar, dass ein bestimmter Geldbetrag, sagen wir eine Mark, nicht stets denselben subjektiven Wert oder Nutzen hat. Sondern dieser wird für einen Bettler höher sein als für einen Millionär. Mehr noch, der Nutzen wird sich für ein und dieselbe Person ändern, wenn diese vom Bettler zum Millionär wird. Somit erscheint es sinnvoll, eine Entscheidung zwischen zwei Glücksspielen nicht von deren Geldbeträgen, sondern von den mit diesen Beträgen assoziierten Nutzen abhängig zu machen. Vom Vergleich des Armen und des Reichen ausgehend argumentierte Bernoulli weiter, dass die Mehrung des persönlichen Reichtums um einen fixen Betrag das Vermögen um einen höheren Prozentsatz mehrt, als den damit verbundenen Nutzen (Savage, 1954). In anderen Worten, der Zugewinn einer Geldsumme ϵ erbringt auf einem niedrigen Vermögenslevel einen höheren Nutzenzuwachs als auf einem hohen. Dieses *Prinzip des abnehmenden Grenznutzens*, unter diesem Namen ist es später in die Ökonomie eingegangen, ist in Abbildung 1 illustriert¹. Ökonomen haben dieses Prinzip auch auf den Nutzen von Gütern ausge-

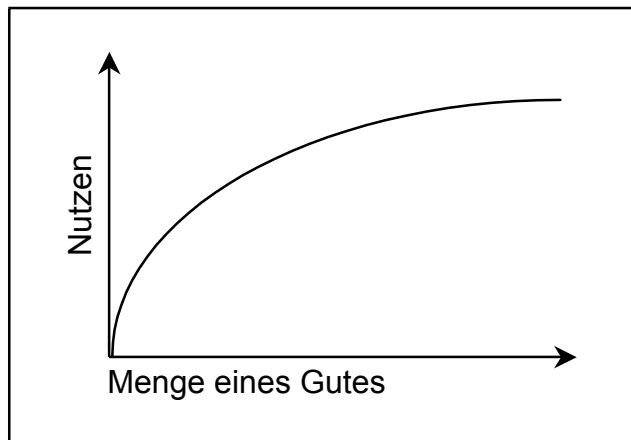


Abb. 1 Prinzip des abnehmenden Grenznutzens.

¹ Bernoulli hat zur Beschreibung des Nutzenzuwachses ΔU , der sich aus der Mehrung des Vermögens V um die Geldsumme ΔV ergibt, folgenden Zusammenhang vorgeschlagen: $\Delta U = k \Delta V / V$, wobei k eine Konstante ist. Die selbe Funktion, die heute als *Fechnersches Gesetz* bekannt ist, hat über einhundert Jahre später eben dieser zur Beschreibung des Verhältnisses von physikalischen Reizen zu daraus resultierenden Sinneseindrücken postuliert (Edwards, 1954).

weitert. Somit ist also die erste Million zwar die schwerste, dafür aber auch die schönste und der Kauf des neunzigsten Paar Schuhe bereitet nicht soviel Freude wie der des vierten².

Das Prinzip des abnehmenden Grenznutzens beansprucht keine absolute Allgemeingültigkeit. In spezifischen Situationen kann der Nutzen auch plötzlich ansteigen. So sind die letzten tausend Mark, die zur Vervollständigung der Lösegeldsumme führen, von besonderer Bedeutung, ebenso die dreitausendundachte Schiene, die das neue Bahngleis vollendet.

Die EU-Regel besagt, dass genau die Handlungsoption gewählt werden soll, die den höchsten erwarteten Nutzen hat. Dieser berechnet sich in Analogie zum Erwartungswert für jede Option als

$$EU = \sum (p_i \cdot u_i)$$

wobei p_i die Wahrscheinlichkeit des i ten Handlungsausgangs bezeichnet und u_i dessen Nutzen. Verluste werden wiederum durch ein negatives Vorzeichen gekennzeichnet. Die EU- hat gegenüber der EV-Regel nicht nur den Vorteil, dass sie die Schwierigkeiten, in die letztere führt (s. o.), nicht aufweist, sondern sie gestattet generell auch den Vergleich verschiedenartiger Güter.

Ein radikales Beispiel für die EU-Analyse eines Entscheidungsproblems ist Pascals aus der Vernunft begründete Empfehlung, an Gott zu glauben. Die Wahrscheinlichkeit für dessen Existenz bezifferte er vom Prinzip des unzureichenden Grundes argumentierend auf $\frac{1}{2}$ (es kann Gott geben oder es kann ihn nicht geben). Falls es Gott gibt, so Pascal, gewinne ich alles, wenn ich an ihn glaube, und ich verliere alles, wenn ich nicht an ihn glaube. Gibt es Gott nicht, so macht es keinen nennenswerten Unterschied ob ich geglaubt habe oder nicht. Daher sollte man glauben – der erwartete Nutzen dieses Verhaltens ist unendlich viel höher.

2.4 Probleme des klassischen Nutzenkonzeptes

Auch in ökonomischen Theorien der Entscheidung fand das Konzept des Nutzens lange Zeit weite Verbreitung. So differenzierte Adam Smith zwischen dem Gebrauchswert („value in use“) und dem Tauschwert („value in exchange“) von Gütern. Sehr häufig seien beide völlig verschieden voneinander. So gebe es nichts nützlicheres als Wasser, für Wasser ließe sich jedoch fast nichts eintauschen, während ein Diamant quasi nutzlos sei, aber einen hohen Tauschwert habe. Das Konzept des Gebrauchswertes wurde jedoch dahingehend kritisiert, dass dieser nicht messbar sei. So kommentierte der Ökonom David Ricardo: „Value in use cannot be measured by any known standard; it is differently estimated by different persons“ (zit. nach Stigler, 1950). Wenn der Gebrauchswert jedoch nicht messbar ist, ist das ganze Konzept wenig tauglich; so lässt sich beispielsweise ein Unterschied zum Tauschwert schlicht nicht feststellen. Trotz des Problems der Messbarkeit blieb Nutzen lange ein wichtiges Konzept. So konnte Walras mit Hilfe des abnehmenden Grenznutzens erklären, warum ein Sinken der Preise zu höherer Nachfrage führt, während umgekehrt steigende Preise die Nachfrage dämpfen (Stigler, 1950). Wichtig ist die Idee des Nutzens vor allem auch in der Wohlfahrtsökonomie, die sich mit der adäquaten Verteilung von Einkommen und Gütern beschäftigt. So

² Aus eigener Erfahrung weiß ich jedoch, dass diesem Prinzip beim Kauf von CDs keine Geltung zukommt.

führt etwa die Idee des abnehmenden Grenznutzens in Verbindung mit der Maxime, dass jeder Steuerzahler den gleichen Anteil seines ökonomischen Nutzens abführen soll, zur progressiven Besteuerung von Einkünften.

Problematisch wird das Konzept des Nutzens vor allem dann, wenn nicht der Nutzen eines einzelnen Gutes betrachtet wird, sondern eines Korbes von Gütern (Edwards, 1954). Der Gesamtnutzen des Korbes kann nämlich nicht mit der Summe der Nutzen der einzelnen Güter gleichgesetzt werden, da sich manche Produkte komplementär zueinander verhalten (z. B. Brot und Käse), andere hingegen miteinander konkurrieren (etwa Käse und Wurst).

Die Schwierigkeiten bei der Messung von Nutzen sowie die Entwicklung von *Indifferenzkurven* und die Erkenntnis von Pareto, dass sich auf deren Grundlage dieselben Ableitungen ergeben wie aus den Nutzenkurven, führte schließlich zur Marginalisierung des klassischen Nutzenkonzeptes (Edwards, 1954).

Das Konzept der Indifferenzkurven stammt von Edgeworth. In seinen Worten beschreibt eine Indifferenzkurve verschiedene Warenkörbe, in denen sich dieselben Güter (im einfachsten Fall zwei) in solchen Mischungsverhältnissen befinden, dass alle Warenkörbe denselben Nutzen haben. Abb. 2 zeigt ein (fiktives) Beispiel für ein Mondmännchen: Dieses zieht denselben Nutzen aus drei Drops und einer Buchhändlerin (Punkt *a*) wie aus einem Drop und acht Buchhändlerinnen (Punkt *b*); ebenso haben sieben Drops und eine Buchhändlerin denselben Nutzen wie vier Drops und sechs Buchhändlerinnen (Punkte *c* und *d*). Wie leicht zu sehen ist ergibt sich beim Vergleich zweier Güter also eine ganze Familie von Indifferenzkurven (drei sind in der Abbildung exemplarisch abgetragen), die zusammen eine Indifferenzkarte bilden. Höhergelegene Kurven repräsentieren dabei einen höheren Nutzen.

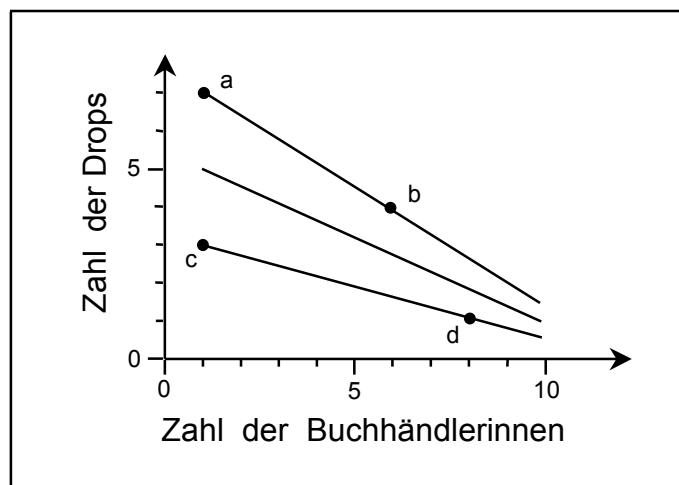


Abb. 2 Indifferenzkurven. Alle Punkte auf einer Linie repräsentieren den gleichen Nutzen. Höhergelegene Kurven zeigen einen höheren Nutzen an.

Solche Kurven lassen sich für alle Güter bilden, auch wenn diese komplementär oder konkurrierend sind, und sie lassen sich unmittelbar aus der Beobachtung von Wahlen ableiten. Eine Interpretation im Sinne von Nutzen ist nicht notwendig, um fundamentale ökonomische Gesetzmäßigkeiten

aus ihnen abzuleiten, so dass Ökonomen auf das bis dato relativ weiche und letztlich introspektiv begründete Konzept des Nutzens zur Beschreibung der Nachfrage nach Konsumgütern zwischenzeitlich verzichteten (Edwards, 1954).

2.5 Eine neue Fundierung des Nutzenkonzeptes – von Neumann und Morgenstern

Das Nutzenkonzept wurde 1944 durch die Arbeit des Mathematikers John von Neumann und des Ökonomen Oskar Morgenstern auf eine neue messtheoretische Grundlage gestellt. In ihrem Buch *Theory of Games and Economic Behavior* zeigten sie Folgendes. Wenn die Wahlen, die ein Entscheidungs-finder zwischen verschiedenen unsicheren Optionen trifft, bestimmten Axiomen folgen, so lassen sich aus diesen Wahlen intervallskalierte Nutzenwerte ableiten, so dass eine Alternative gegenüber einer anderen genau dann bevorzugt wird, wenn deren erwarteter Nutzen größer ist als der der konkurrierenden Option. Die Wahlen beziehen sich dabei auf Optionen der Form (A, p, B) . D. h. es handelt sich hier um „Lotterien“ bei denen die Konsequenz A mit der Wahrscheinlichkeit p eintritt, andernfalls aber die Konsequenz B (deren Wahrscheinlichkeit somit $1 - p$ beträgt). A und B können dabei wiederum Lotterien sein. (Dawes, 1988).

Ein Beispiel mag diese Art der Nutzenkonstruktion veranschaulichen. Betrachten wir einen Entscheider Felix, der in der glücklichen Situation ist, zwischen den folgenden drei Optionen wählen zu können: A erhält 10 000 DM, B erhält 6 000 DM und C erhält 2 000 DM. Felix zeige dabei eine Präferenz von A gegenüber B und von B gegenüber C . Die Zuweisung von Nutzenwerten u zu A und C ist zunächst arbiträr, da der Nullpunkt einer Intervallskala frei wählbar und diese beliebig streck- oder stauchbar ist. Das heißt wenn U eine Nutzenfunktion ist, dann ist $U' = aU + b$ mit $a, b \in \mathbb{R}$ eine äquivalente Funktion.

Wir wollen willkürlich festlegen, es sei $u_A = 1$ und $u_C = 0$. Um u_B zu ermitteln, kreieren wir eine vierte Option D . Dabei handelt es sich um eine Lotterie, bei der Felix mit einer Wahrscheinlichkeit von .6 A erhält (also 10 000 DM), andernfalls aber C (2 000 DM). Der Nutzenwert dieser neuen Option berechnet sich nun als $u_D = .6 \cdot u_A + .4 \cdot u_C = .6$. Felix ist genau dann zwischen den beiden Optionen B und D indifferent, wenn gilt $u_B = u_D$. Wir wollen annehmen, dass Felix die Wahl zwischen diesen beiden Möglichkeiten tatsächlich gleichgültig ist – und wissen somit $u_B = .6$. Die Viertausendmarkdifferenz zwischen 2 000 DM und 6 000 DM entspricht also einer Nutzenwertdifferenz von .6. Die Viertausendmarkdifferenz zwischen 6 000 DM und 10 000 DM führt aber nur zu einer Nutzenwertdifferenz von .4. Für Felix würde sich somit, zumindest im hier betrachteten Bereich, ein abnehmender Grenznutzen von Geld ergeben.

Diese Nutzenwerte sind zunächst rein mathematische Größen, deren Existenz sich aus den Axiomen, die später noch genauer beschrieben werden, ergibt. D. h. dem Entscheider braucht dieser „von Neumann und Morgenstern Nutzen“ nicht gewärtig zu sein. In der Regel wird er jedoch persönliche Werte widerspiegeln, die der Introspektion zugänglich sind.

2.6 Verschiedene Konzeptualisierungen von Wahrscheinlichkeit und deren Implikationen für die EU-Regel

Mit der von von Neumann und Morgenstern geschaffenen Fundierung des Nutzenkonzeptes gewinnt die EU-Regel an Substanz. Da diese Regel aber auf das Vorliegen von Wahrscheinlichkeiten angewiesen ist, bleibt ihr Anwendungsspielraum stark begrenzt, solange man das klassische Verständnis von Wahrscheinlichkeit aufrechterhält. Denn in diesem Fall gibt es nur sehr wenige Ereignisse, denen überhaupt feste Wahrscheinlichkeiten zugeordnet werden können. Diese wenigen sind obendrein für unsere Entscheidungen weitgehend irrelevant, denn die wenigsten Entscheidungen unter Unsicherheit, die wir treffen, sind abhängig von Würfeln, Münzen, Spielkarten oder dergleichen.

Die klassische Konzeption von Wahrscheinlichkeit ist jedoch nicht die einzige. Zur Zeit existiert kein einheitliches Verständnis von Wahrscheinlichkeit, vielmehr gibt es eine Reihe unterschiedlicher Konzeptionen und Schulen. Diese sind sich einig, was die mathematische Definition von Wahrscheinlichkeiten betrifft, die durch die *Kolmogoroff Axiome* festgelegt ist: Es lässt sich jedem Ereignis E eine reelle Zahl $p(E)$ zuordnen, wobei den folgenden Bedingungen genüge zu tun ist (vgl. beispielsweise Bortz, 1999):

1. $p(E) \geq 0$.
2. Die Wahrscheinlichkeit des sicheren Ereignisses beträgt 1.
3. Wenn sich n verschiedene Ereignisse E_1, E_2, \dots, E_n wechselseitig ausschließen, so ist die Wahrscheinlichkeit dafür, dass irgendeins dieser Ereignisse auftritt, gleich der Summe der Einzelwahrscheinlichkeiten: $p(E_1 \cup E_2 \cup \dots \cup E_n) = p(E_1) + p(E_2) + \dots + p(E_n)$.

Aus diesen Axiomen lassen sich dann Regeln ableiten, wie man mit Wahrscheinlichkeiten zu rechnen hat. Auch in diesem Punkt gibt es keinen Dissenz zwischen den verschiedenen Schulen. Sind also Ereignissen erst einmal Wahrscheinlichkeiten zugeordnet, so ist klar, wie mit diesen zu verfahren ist. Das Kernproblem besteht aber darin, wie eine solche Zuweisung möglich ist. Die Antworten hierauf sind eng damit verknüpft, welche Haltung man bezüglich der Frage einnimmt, was die Bedeutung von Wahrscheinlichkeit ist. In diesen beiden Fragen gibt es keine Übereinstimmung zwischen verschiedenen Schulen.

Die klassische Antwort zu diesen Fragen wurde bereits beschrieben; aufgrund der symmetrischen Eigenschaften von Objekten (Gleichheit von Würfel- und Münzseiten, Spielkarten und Vertiefungen in einem Rouletterad) werden Elementarereignissen, die sich in unvorhersehbarer Weise aus der Manipulation dieser Objekte (werfen, mischen und ziehen, rollen) ergeben, a priori Wahrscheinlichkeiten zugeschrieben.

Die *logische Konzeption* von Wahrscheinlichkeit betrachtet diese als Maß dafür, inwieweit ein Satz von Aussagen die Wahrheit einer anderen Aussage nahelegt. Der Schluss auf den Grad der Wahrheit wird dabei als objektiv und zwingend angesehen. Aus dieser Perspektive ist Wahrscheinlichkeit eine Erweiterung der Logik. Letztere gestattet unter bestimmten Voraussetzungen, aufgrund gegebener Tatsachen mittels eines rein mechanischen Kalküls die Wahrheit oder Falschheit neuer Aussagen abzuleiten. Wahrscheinlichkeit bedeutet nun ein eingeschränktes Wahrheitsurteil. Zu den Vertretern

der logischen Konzeption zählen Carnap und John Meynard Keynes. Ein Kalkül, das die Bestimmung dieser Wahrscheinlichkeitswerte umfassend ermöglicht, so wie das logische Kalkül die Bestimmung von Wahrheitswerten erlaubt, existiert jedoch nicht, so dass das Problem der Zuordnung von Wahrscheinlichkeiten zu Ereignissen innerhalb dieser Konzeption bislang nicht gelöst werden konnte.

Anhänger einer *frequentistischen* Interpretation, hierzu gehören von Mises und Neyman, konzeptualisieren Wahrscheinlichkeit über beobachtete Häufigkeiten. Wenn eine Münze sehr oft geworfen wurde und in genau 47% der Fälle Kopf fiel, so sollten wir nicht davon ausgehen, dass die Münze fair ist. Da sich Wahrscheinlichkeiten aus Sicht der Frequentisten auf relative Häufigkeiten beziehen, würden diese ein Wahrscheinlichkeitsurteil darüber, dass im nächsten Wurf Kopf fällt ablehnen. Sie würden entweder argumentieren, dass eine solche Aussage bedeutungslos ist, da sie sich nicht auf relative Häufigkeiten bezieht, oder aber dass sie genau 1 oder 0 sein muss. Letzteres wird klar, wenn wir davon ausgehen, dass sich die Münze bereits in der Luft befindet. Jetzt ist der Ausgang des Wurfes durch die physikalischen Gesetze determiniert. Wir haben vielleicht nicht genügend Informationen, um den Ausgang vorhersagen zu können, aber dessen Ergebnis steht bereits fest. Wahrscheinlichkeiten sind gemäß der frequentistischen Auffassung Aussagen über objektive Eigenschaften der Welt und nicht über den Glauben bezüglich der Eigenschaften der Welt.

Auch aus einem weiteren Grund kann es sinnvoll erscheinen, Wahrscheinlichkeiten über Einzelaussagen abzulehnen, auch wenn relevante Informationen in Form von relativen Häufigkeiten vorliegen. Stellen wir uns Clarisse, ein sechsjähriges deutsches Mädchen, vor. Wir wollen nun wissen, wie hoch die Wahrscheinlichkeit dafür ist, dass ihre Eltern das Kind zum Elementarunterricht in der Musikschule anmelden. Wenn dies bei 12% der Kinder geschieht, sollten wir natürlich sagen, dass die Chance .12 beträgt. Es mag aber gleichzeitig sein, dass bei den Mädchen 14% angemeldet werden – dann sollten wir natürlich .14 sagen. Sicherlich sollten wir dann aber auch berücksichtigen, dass die Quote bei Stadtkindern deutlich höher ist als bei Landkindern. In München, wo Clarisse wohnt, betrage die Quote für Mädchen 19%, also beziffern wir die Chance doch lieber auf .19. Dann sollten wir jedoch auch unbedingt berücksichtigen, dass beide Eltern Berufsmusiker bei den Münchener Philharmonikern sind – sicherlich ein wichtiger Faktor. Nehmen wir an, es habe in den letzten zehn Jahren noch genau ein weiteres solches Ehepaar gegeben und diese hätten ihre Tochter *nicht* in den Elementarunterricht geschickt. Sollen wir dann sagen, dass die Chance bei Clarisse 0 ist, dass es also völlig ausgeschlossen ist, dass sie den Musikunterricht besuchen wird? Wohl kaum. Aus diesen Gründen sind aus Sicht der Frequentisten nur Aussagen über relative Häufigkeiten wie „wirf die Münze auf die gleiche Art wieder sehr oft, dann wird wieder in 47% der Fälle Kopf fallen“ sinnvoll.

Subjektivisten, zu denen Savage und de Finetti gehören, ist diese Konzeption einerseits viel zu eng. Im Gegensatz zu allen übrigen Konzeptionen betrachten sie Wahrscheinlichkeit als Grad einer persönlichen Überzeugung. D. h. diese ist nicht einem Ereignis zu eigen sondern einer Überzeugung bezüglich eines Ereignisses. Dem entsprechend gibt es innerhalb dieser Konzeption auch nicht die Einschränkung, dass sich nur bestimmten Ereignissen Wahrscheinlichkeiten zuschreiben lassen. Damit einher geht, dass verschiedene Personen bezüglich der Wahrscheinlichkeit eines Ereignisses

unterschiedlicher Meinung sein dürfen, da es hier ja nicht um eine objektive Eigenschaft geht. Wenn also einer die Wahrscheinlichkeit, dass Clarisse zum Elementarunterricht gehen wird, mit 0 beziffert und jemand anderes mit .19, so bereitet dies Subjektivisten keinerlei Schwierigkeiten. Aus dieser Perspektive würde es auch keinen Sinn machen zu sagen, dass eine der beiden Überzeugungen überlegen ist. Die Zuschreibung einer Wahrscheinlichkeit zu einem einzelnen Ereignis entzieht sich einer Bewertung.

Subjektivisten kritisieren die frequentistische Position nicht nur dahingehend, dass diese die Anwendbarkeit des Wahrscheinlichkeitskonzeptes unnötig einschränken. Sie werfen ihr auch vor, nicht widerspruchsfrei zu sein. Die frequentistische Axiomatik behandelt Wahrscheinlichkeiten nämlich nur in Bezug auf unendlich lange Beobachtungsfolgen. In dieser Form ist Wahrscheinlichkeit natürlich nicht handhabbar, denn, wie Keynes richtig erkannte, „in the long run we shall all be dead“. Dass wir nicht auf unendlich lange Folgen zurückgreifen können, brauchte zunächst kein Problem zu sein. Schließlich operiert die Geometrie auch mit einem Konzept des Punktes, wonach dieser unendlich klein ist. Dennoch kann ein Geometer, dessen Bleistift naturgemäß dicker ist, sinnvoll deren Konzepte anwenden. Ebenso kann ein Frequentist argumentieren, dass er sich mit „hinreichend langen“ Beobachtungsreihen begnügen kann. Das Problem besteht jedoch darin, dass er nicht angeben kann, *wie* lang so eine Reihe sein sollte, denn sobald er das tut, lässt sich nachweisen, dass er sich so verhält, als ob er den zur Diskussion stehenden Ereignissen *a priori* eine bestimmte Wahrscheinlichkeit zuerkennt, was ja gerade unvereinbar mit der frequentistischen Position ist (Good, 1959, Savage, 1954).

Entscheidungen unter Unsicherheit mittels der EU-Regel zu beschreiben, ist quasi nur aus subjektivistischer Sicht möglich. Denn nur aus diesem Verständnis heraus ist die Kalkulation eines Erwartungswertes für ein Einzelereignis sinnvoll und gleichzeitig problemlos möglich.

2.7 Zusammenfassung

Aktuelle Bemühungen, Entscheidungen unter Unsicherheit zu beschreiben, stehen am Ende einer Tradition, die in der im 16. Jahrhundert aufkeimenden Wahrscheinlichkeitsrechnung ihre Wurzeln findet. Am Beginn dieser Entwicklung galt dabei als vernünftig, die Option mit dem höchsten Erwartungswert zu wählen. Paradoxa, die aus dieser Regel entstehen, führten dazu, dass sie von der Konzeption abgelöst wurden, nicht den erwarteten Wert sondern den erwarteten Nutzen zu maximieren. Kritisiert wurde dieses Konzept aber dahingehend, dass sich der Nutzen von Gütern nicht unabhängig von deren Wert betrachten lasse. Diese Kritik wurde dadurch hinfällig, dass von Neumann und Morgenstern 1944 eine Möglichkeit vorstellten, Nutzenwerte aus Wahlen zwischen Optionen mit verschiedenen Wahrscheinlichkeiten abzuleiten. Nur auf der Grundlage einer subjektivistischen Wahrscheinlichkeitskonzeption ist es jedoch sinnvoll, beliebigen Einzelereignissen eine Wahrscheinlichkeit zuzuweisen und somit dem Prinzip der Maximierung des erwarteten Nutzens ein breites Anwendungsfeld zu erschließen.

3. WEITERE THEORETISCHE ENTWICKLUNGEN: VON DER SUBJECTIVE EXPECTED UTILITY THEORIE ZUM KONZEPT DER ENTSCHEIDUNGSGEWICHTE

In Kapitel 3 werden neuere theoretische Entwicklungen dargestellt. Ziel ist dabei nicht, einen vollständigen Überblick über den gegenwärtigen Stand der Theoriebildung zu geben, als vielmehr einige Theorien, Konzepte und Phänomene vorzustellen, die für die vorliegende Arbeit wesentlich sind. Zunächst wird erläutert, wie sich subjektive Wahrscheinlichkeiten messen lassen. Dass menschliches Verhalten sich nicht immer dadurch adäquat beschreiben lässt, dass man annimmt, sie maximierten ihren subjektiv erwarteten Nutzen, wird exemplarisch anhand dreier Phänomene illustriert. Das neuere Konzept der Wahrscheinlichkeitsgewichte, das Bestandteil der meisten aktuellen Theorien über Entscheidungen unter Unsicherheit ist, ist geeignet, auch diese Phänomene angemessen beschreiben zu können.

3.1 Die Messbarkeit subjektiver Wahrscheinlichkeit: Savages Subjective Expected Utility Theorie

Bezüglich der Messbarkeit subjektiver Wahrscheinlichkeiten hat Savage einen wichtigen Beitrag mit seiner *Subjective Expected Utility* Theorie (SEUT, Savage, 1954) geliefert. Er zeigte, ähnlich wie von Neumann und Morgenstern dies für Nutzen taten, dass dann, wenn die Wahlen eines Entscheiders, die dieser zwischen verschiedenen unsicheren Optionen trifft, bestimmten Axiomen folgen, nicht nur Nutzenwerte sondern auch subjektive Wahrscheinlichkeiten ableitbar sind, dergestalt dass der daraus resultierende subjektiv erwartete Nutzen maximiert wird. D. h. wenn ein Entscheider sich gemäß den Axiomen verhält, so entscheidet er so, als ob er verschiedenen Zuständen einen subjektiven Nutzen zuschriebe, als ob er dem Eintreffen dieser Zustände eine gewisse Wahrscheinlichkeit zuerkenne und als ob er den resultierenden subjektiv erwarteten Nutzen maximiere. Die Axiome selbst sieht Savage als Ausdruck von Rationalität an – rationale Entscheidungen seien solche, die den Axiomen folgen. Insofern beansprucht die SEUT normative Geltung. Savage ist dabei gedämpft optimistisch, dass sie gleichzeitig tatsächliches Wahlverhalten adäquat beschreiben kann, denn er geht davon aus, dass die Axiome allgemein als vernünftig angesehen werden und Menschen diesen folgen möchten. Vermutlich würden sie dies nicht in allen Fällen tun. Eine Verletzung ist aber aus Savages Sicht eher als Lapsus zu interpretieren – wenn wir jemanden fragen, wieviel 17 mal 11 ist, und eine falsche Antwort erhalten, so würden wir daraus auch nicht den Schluss ziehen, dass diese Person die Rechenregeln unsinnig findet. Wenn das Problem genauer überdacht wird, sollte dies, so Savage, dann also zu axiomkonformen Wahlen führen.

Die Axiome der SEUT sollen im Folgenden kurz dargestellt werden. Ich folge dabei der Darstellung von Dawes (1988), die in meinen Augen etwas einfacher ist als diejenige, die Savage selbst gibt.

A , B , und C seien Optionen im Raum aller möglichen Optionen S . \preccurlyeq bedeutet „wird nicht vorgezogen“, \succ bedeutet „wird vorgezogen“, und $=$ zeigt die Äquivalenz zweier Optionen an.

(I) Vergleichbarkeit: Es gilt entweder $A \preccurlyeq B$, $B \preccurlyeq A$ oder beides. Dieses Axiom besagt, dass alle Optionen miteinander vergleichbar sind und dass entweder eine Option gegenüber der anderen präferiert wird oder andernfalls Indifferenz vorliegt.

(II) Transitivität: Wenn $A \preccurlyeq B$ und $B \preccurlyeq C$ dann ist auch $A \preccurlyeq C$. Entscheidungen sollen also transitiv sein; wenn Ihnen im Moment eine Orange lieber ist als ein Schokoriegel und ein Schokoriegel lieber als ein Frikadelle, dann sollten Sie auch die Orange der Frikadelle vorziehen. Dieses Gebot lässt sich leicht begründen. Nehmen wir an, eine Person habe die Präferenzordnung $A > B > C$. Wenn wir ihr Option C schenken, so sollte sie gewillt sein, einen gewissen Betrag zu bezahlen, um diese gegen B einzutauschen, schließlich findet sie B ja besser. Ebenso wird sie danach dazu bereit sein, eine weitere Summe für den Tausch von B gegen A zu zahlen. Ihre Wahlen werden intransitiv, wenn sie nun, im Besitz von A , bereit ist, dies gegen einen Aufpreis in C umzuwandeln. Dann hat sie wiederum C , womit sie begonnen hat, aber weniger Geld in der Tasche. Eine solche Serie wird (aus mir nicht bekannten Gründen) auch als *Dutch Book* bezeichnet.

Schlussendlich ist aus der hier vorgestellten entscheidungstheoretischen Perspektive das Kernkriterium für rationale Entscheidungen, dass diese widerspruchsfrei sein sollen (Dawes, 1988). Somit kann eine einzelne Wahl, zumindest aus dieser Sicht, niemals irrational sein. Sie kann allenfalls ungewöhnliche Präferenzen aufdecken. Dies ist etwa aus Sicht der meisten Leute bei den Tauschaktionen von Hans im Glück der Fall, der einen Goldklumpen schrittweise gegen das nackte Nichts tauscht. Aus Sicht der SEUT ist diese Tauschfolge jedoch keineswegs irrational, da sie keine Intransitivität aufweist.

(III) Geschlossenheit: Wenn A und B Optionen in S sind, so ist auch (ApB) eine mögliche Option. Dieses Axiom erfordert, dass eine Entscheiderin dazu in der Lage ist, eine Wahrscheinlichkeitsmischung verschiedener Optionen ebenfalls als mögliche Option zu begreifen. Wenn Sie etwa vor der Wahl stehen, in welches von zwei Kinos Sie gehen sollen, wobei in beiden Fällen der Film ausverkauft sein könnte, so erfordert Axiom III, dass Sie in der Lage sind sich vorzustellen, dass Sie diese Entscheidung wahlweise von einem Münzwurf oder aber vom Rat einer Freundin abhängig machen. Dawes (1988) begründet dieses Axiom so: „If people were incapable of doing so, there would be little point in theorizing about decision making“ (S. 155).

(IV) Verteilung der Wahrscheinlichkeiten über Alternativen: $[(ApB)qB] = (ApqB)$. Dieses Axiom stellt sicher, dass man bei der Verknüpfung von Wahrscheinlichkeiten den Kolmogoroff Axiomen folgt. So sollte man etwa bei der Wahl zwischen den beiden folgenden Wetten indifferent sein, da diese ineinander überführbar und somit im Prinzip identisch sind.

(1) Sie erhalten 10,- DM mit einer Wahrscheinlichkeit von .25, andernfalls nichts.

(2) In Stufe I haben Sie eine Chance von .5, Stufe II zu erreichen. Falls Ihnen dies nicht gelingt, ist das Spiel beendet. Wenn Sie Stufe II erreichen, erhalten Sie mit einer Wahrscheinlichkeit von .5 10,- DM, andernfalls nichts.

(V) Unabhängigkeit: $A \preccurlyeq B$ dann und nur dann wenn $(ApC) \preccurlyeq (BpC)$. Dies wird auch als *Sure Thing* Prinzip bezeichnet und besagt, dass die Präferenz zwischen zwei Optionen nicht dadurch umgekehrt werden darf, dass beiden Optionen identische Konsequenzen hinzugefügt werden. Wenn

Sie also lieber eine eins zu zehn Chance auf eine Woche Urlaub in New York haben wollen als eine gleich hohe Chance auf eine einwöchige Mittelmeerkreuzfahrt, so sollte sich daran nichts ändern, wenn Sie in beiden Fällen zusätzlich tausend Mark Taschengeld gewinnen. Falls eine zusätzliche dritte Wahlmöglichkeit eingeführt wird, die für Sie weniger attraktiv ist als die beiden anderen (etwa dass Sie eine Neckermann Jogginghose geschenkt bekommen), so sollte auch dies nicht zu einer Präferenzumkehr führen.

In meinen Augen ist dieses Prinzip nur dann gerechtfertigt, wenn man die Einschränkung einführt, dass C gegenüber A und B gleichermaßen konkurrierend, kompetitiv oder neutral zu sein hat. So ist es durchaus verständlich, wenn sich bei einer Person die Präferenz zwischen beiden Reiselotterien dadurch verkehrt, dass zusammen mit der Reise ein Einkaufsgutschein für Gesellschaftskleidung zu gewinnen ist.

Ein etwas anders gelagertes Beispiel, bei dem ebenfalls die Verletzung des Sure Thing Prinzips unproblematisch erscheint, bringt Gigerenzer (1996): Mr. Smart ist für Sonntag 21.00h bei einer Kollegin eingeladen. Nachdem er im Wohnzimmer Platz genommen hat, bietet diese ihm Nüsse und Cracker an (A), von denen der Gast aber nichts nimmt (B), da er vor dem Essen seinen Appetit nicht mindern möchte. Einige Zeit später kommt der Ehemann der Kollegin hinzu und bringt Kuchen und Tee mit (C), so dass die Palette der Optionen auf drei gewachsen ist (A, B, C). C ist jedoch nicht nur einfach eine weitere Option, sie zerstört auch Mr. Smarts Hoffnung auf ein gehaltvolles Abendessen, mit der Folge, dass dieser nun lieber Cracker und Nüsse isst als nichts. Hier wird also eine dritte Option hinzugefügt, die das Wahlverhalten verändert, obwohl sie nicht die attraktivste ist und somit gemäß des Sure Thing Prinzipes eigentlich irrelevant sein sollte.

Wie später noch zu zeigen sein wird, eignet sich die SEUT vor allem deswegen nur bedingt zur Deskription von Entscheidungsverhalten, weil Menschen systematisch das Sure Thing Prinzip verletzen – auch bei solchen Wahlen bei denen, anders als bei den gerade genannten Beispielen, dessen Befolgung geboten erscheint.

(VI) Konsistenz: $A \preccurlyeq B$ dann und nur dann, wenn $A \preccurlyeq (ApB) \preccurlyeq B$. Wenn Ihnen also eine Orange lieber ist als eine Banane, so sollte Ihnen jede Lotterie, bei der Sie eine wie auch immer gearbeitete Chance auf die Orange haben, andernfalls aber die Banane erhalten, lieber sein, als die Banane unmittelbar zu wählen. Ebenso sollte Ihnen jede Chance (Ap) auf eine Orange lieber sein als eine ebenso große Chance (Bp) auf eine Banane. Dawes (1988) sieht dieses Axiom als selbstevident an. Ich teile diese Meinung nicht unbedingt. Stellen wir uns Gilberte vor, die 27 Jahre alt ist, Philosophie studiert hat und mittlerweile als Netzwerkadministratorin in einer Computerfirma arbeitet; mit ihrem Job dort ist sie sehr zufrieden. Unverhofft flattert ihr das Angebot ins Haus, an einer amerikanischen Elite-Universität zu promovieren. Eine erfolgreiche akademische Laufbahn in ihrer angestammten Disziplin wäre ihr mit diesem Abschluss in dem fremden Land fast sicher, nicht aber in Deutschland. Auf eine zweite Chance dieser Art kann sie nicht zählen, ihr erscheint es aber auch unwahrscheinlich, dass sie in Deutschland noch einmal einen so guten Job fände, falls es ihr in den USA nicht gefiele und sie wieder von dort zurückkehrte. Obwohl sie eher dazu neigt, in die USA zu gehen, fällt es ihr verständlicher Weise sehr schwer, eine Entscheidung zu treffen. Könnte es in dieser Situation für Gilberte nicht angenehmer sein, wenn sie nicht selbst die Wahl zu treffen hätte,

sondern wenn ihr künftiges Schicksal (Deutschland oder USA) statt dessen von äußeren Umständen, die sie nicht kontrollieren kann, abhänge und sie somit von der Last der Entscheidung befreit wäre?

(VII) Lösbarekeit: Wenn $A \preccurlyeq B \preccurlyeq C$, so existiert eine Wahrscheinlichkeit p , so dass gilt $B = (ApC)$. Dies bedeutet, dass es keine Alternative gibt, die einer anderen hinsichtlich ihres subjektiven Nutzens unendlich über- oder unterlegen ist. Kein Gut oder Zustand soll also so attraktiv sein, dass nicht eine derart geringe Realisierungschance denkbar ist, dass diese „Lotterie“ nur noch soviel wert ist wie ein anderes positives Gut oder ein anderer positiver Zustand. Ebenso soll es keinen Schaden geben, der so groß ist, dass wir diesbezüglich nicht das geringste Risiko eingehen würden, wenn wir uns damit einen Vorteil erkaufen. Wenn man utopische Wetten, die ewige Glückseligkeit oder ewige Verdammnis einschließen, außen vor lässt, erscheint dieses Prinzip durchaus angemessen. Obwohl Leben und Unversehrtheit für uns wichtige Werte sind, riskieren wir diese doch im Straßenverkehr, um uns eine Tiefkühlpizza zu kaufen.

3.2 Drei Phänomene, die Unzulänglichkeiten der SEUT exemplarisch aufzeigen

Wie bereits erwähnt, verstößen Menschen in ihrem Entscheidungsverhalten gegen die Axiome der SEUT. Drei entsprechende Phänomene sollen hier aufgegriffen werden, nämlich (i) Ambiguität von Wahrscheinlichkeiten, deren Einfluss das Ellsberg-Paradox demonstriert, (ii) das Allais-Paradox sowie (iii) Unterschiede, die sich zwischen dem Bieten für und der Wahl zwischen Lotterien ergeben. Ziel ist dabei nicht, zu einer umfassenden Würdigung der SEUT zu gelangen, sondern eine Reihe empirischer Phänomene und theoretischer Konzepte vorzustellen, die für die vorliegende Arbeit relevant sind. Den Stellenwert der SEUT für die Entscheidungsforschung diskutieren Kühberger (1994) und Pfister (1994).

Ehe man sich Gedanken darüber macht, inwieweit sich menschliches Entscheidungsverhalten angemessen mit der SEUT beschreiben lässt, sollte man realistischerweise die von ihr berücksichtigten Dimensionen des Nutzens und der subjektiven Wahrscheinlichkeit als nicht absolut feststehende Größen konzeptualisieren, sondern zufällige Fluktuationen erlauben. Wenn jemand sich absolut nicht entscheiden kann, ob er lieber eine 1 000 DM teure Reise nach London oder eine ebenso teure Reise nach Paris buchen soll, so erwarten wir natürlich nicht, dass sein Urteil plötzlich feststeht, wenn sich der Preis der London-Tour um eine Mark verringert. Mit der SEUT wäre aber genau dies zu erwarten (Tversky, 1972). Wenn Nutzen und Wahrscheinlichkeit sinnhafterweise als Zufallsvariablen gedacht werden, so erwarten wir selbstverständlich, dass die SEUT-Axiome gelegentlich verletzt werden. Solche Verstöße sind für die Theorie unproblematisch, solange sie unsystematisch sind.

3.2.1 Ambiguität

Die Vorstellung subjektiver Wahrscheinlichkeit als einer Zufallsvariable berührt jedoch bereits das erste der drei oben genannten Probleme, ein Problem, das Savage selbst bei einer subjektivistischen

Wahrscheinlichkeitsauffassung gegeben sieht. Es gibt nämlich, so Savage (1954), Wahrscheinlichkeiten, bei deren Bestimmung wir uns sicher fühlen. Dies kann der Fall sein, wenn wir den Ausgang „Kopf“ eines Münzwurfs beurteilen; bei der Zuweisung anderer Wahrscheinlichkeiten, etwa dem Ereignis „Banküberfall mit Geiselnahme innerhalb der nächsten vier Wochen in der BRD“, sind wir uns weniger sicher. Was aber heißt hier „sicher“ oder „weniger sicher“? Die Bedeutung dieser Begriffe ist offensichtlich vage, und dieses Konzept der Vagheit lässt sich innerhalb der Theorie subjektiver Wahrscheinlichkeiten nicht fassen. Um einen solchen Sachverhalt diffuser, nicht präziser Wahrscheinlichkeiten zu bezeichnen, hat sich der Begriff *Ambiguität* etabliert.

Zunächst erscheint es aus Sicht der SEUT möglich, Ambiguität als Wahrscheinlichkeit zweiter Ordnung zu begreifen; dies ist in Abb. 3 dargestellt. Ich könnte also für relativ wahrscheinlich halten, dass die Wahrscheinlichkeit des zu beurteilenden Ereignisses bei .02 liegt und die Wahrscheinlichkeiten .05 und .001 als vergleichsweise unwahrscheinlich beurteilen. Mathematisch gesehen macht dieses Unterfangen jedoch keinen Sinn, weil sich aus einer solchen Wahrscheinlichkeitsverteilung zweiter Ordnung wiederum ein Erwartungswert für die Wahrscheinlichkeit des zu beurteilenden Ereignisses ergibt, der dann die subjektive Wahrscheinlichkeit bilden sollte.

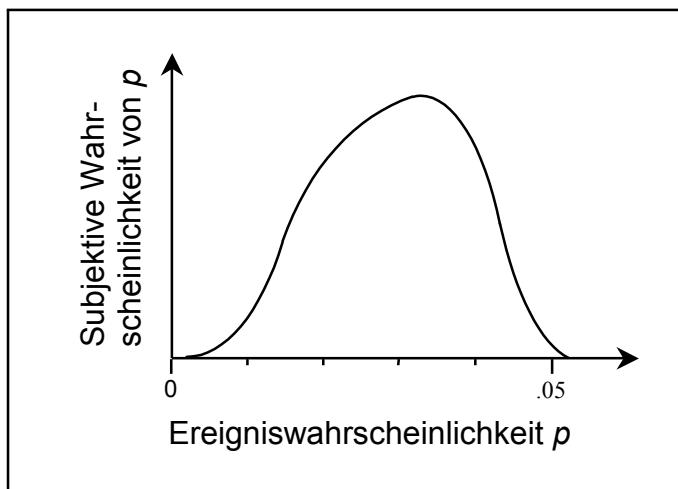


Abb. 3 Darstellung einer Wahrscheinlichkeit zweiter Ordnung.

Dem Konzept der Ambiguität kann die SEUT also nicht Rechnung tragen, wie Savage selber einräumt. Dies ist insofern problematisch, als dass Vagheit von Wahrscheinlichkeiten nicht nur eine subjektive Erfahrung ist, sondern auch einen unmittelbaren Einfluss auf Entscheidungen hat. Dies demonstriert das sogenannte *Ellsberg-Paradox* (Ellsberg, 1961): Stellen Sie sich zwei Urnen vor, in denen rote und/oder schwarze Kugeln liegen. Urne \mathcal{A} enthält 100 Kugeln in einem Ihnen unbekannten Mischungsverhältnis, Urne $N\mathcal{A}$ ³ enthält 50 rote und 50 schwarze Kugeln. Sie haben nun viermal nacheinander die Wahl zwischen zwei Wettten; diese sind unten aufgeführt. „Auf Rot_A wetten“ soll dabei heißen, dass eine Kugel aus Urne \mathcal{A} gezogen wird und Sie einen ausgesetzten Preis

³ Zum einfacheren Verständnis: \mathcal{A} und $N\mathcal{A}$ stehen für *ambige* und *nicht ambige* Gewinnwahrscheinlichkeiten.

gewinnen, wenn die Kugel rot ist – sollte schwarz gezogen werden, so gehen Sie leer aus; die Formulierungen „Schwarz_A“ etc. sind entsprechend zu verstehen.

- (1) Wollen Sie lieber auf Rot_A oder lieber auf Schwarz_A wetten, oder ist es egal?
- (2) Wollen Sie lieber auf Rot_{NA} oder lieber auf Schwarz_{NA} wetten, oder ist das egal?
- (3) Wetten Sie lieber auf Rot_A oder lieber auf Rot_{NA}, oder sind Sie indifferent?
- (4) Möchten Sie lieber auf Schwarz_A oder lieber auf Schwarz_{NA} wetten, oder ist es Ihnen egal?

Typischerweise präferieren Vpn Rot_{NA} gegenüber Rot_A, das heißt sie versuchen eine rote Kugel eher aus der Urne mit der bekannten Farbmischung zu ziehen. Savages Ansatz folgend würde man daraus den Schluss ziehen, dass die Vpn glauben, dass Urne \mathcal{A} mehr schwarze als rote Kugeln enthält. In der Regel wird aber auch anschließend Schwarz_{NA} gegenüber Schwarz_A bevorzugt. Man müsste also auch folgern, dass die Leute glauben, in Urne \mathcal{A} seien mehr rote als schwarze Kugeln. Dies steht aber im Widerspruch zur vorherigen Inferenz; die Wahlen laufen den SEUT-Axiomen zuwider⁴.

Ambiguität beeinflusst Entscheidungen unter Unsicherheit also systematisch, während die SEUT Ambiguität nicht beschreiben kann und von einer normativen Warte aus fordert, dass die mangelnde Genauigkeit von Wahrscheinlichkeitsinformationen keinen Einfluss auf Entscheidungen haben soll⁵.

3.2.2 Das Phänomen des Preference-Reversal und das Contingent Weighting Modell

Ein zweiter systematischer Verstoß gegen die SEU-Konzeption kommt im Phänomen des *Preference-Reversals* zum Ausdruck. Wenn verschiedenen Wahlmöglichkeiten ein Nutzenwert zu eigen ist, wie die SEUT annimmt, so sollte egal sein, auf welche Weise man diesen misst. Dass dies jedoch nicht der Fall ist, zeigt ein Experiment von Tversky, Sattath und Slovic (1988), die eine Gruppe von Vpn mit folgendem Problem konfrontierte: „Pro Jahr werden in Israel etwa 600 Personen in Verkehrsunfällen getötet. Das Verkehrsministerium erwägt verschiedene Programme, um die Zahl der Opfer zu senken. Bitte beurteilen Sie die folgenden zwei Programme, für die jeweils angegeben ist, was sie pro Jahr kosten und mit wieviel Todesopfern nach deren Einführung zu rechnen ist.“

Programm A: 500 Opfer, 55 Mio. \$ Kosten

Programm B: 570 Opfer, 12 Mio. \$ Kosten

Welches Programm favorisieren Sie?“

Ein solches Problem, bei dem zwischen zwei oder mehr vorgegebenen Alternativen zu entscheiden ist, wird als *Choice Problem* bezeichnet. Im vorliegenden Fall entschieden sich zwei Drittel der

⁴ Die hier gegebene Darstellung folgt unmittelbar derjenigen Ellsbergs (1961), der aus mir nicht nachvollziehbaren Gründen darauf besteht, dass die Vpn nicht Urne *und* Gewinnfarbe bestimmen dürfen und auch nicht wissen sollen, welche Wahlen ihnen im Folgenden noch präsentiert werden. Dieses Vorgehen legt den Schluss nahe, dass das geschilderte Wahlverhalten dadurch zustande kommt, dass die Vpn dem Versuchsleiter misstrauen und fürchten, von ihm aufs Glatteis geführt zu werden. Systematische Ambiguitätsvermeidung findet sich jedoch auch dann, wenn den Vpn überlassen wird, die Gewinnfarbe selbst zu bestimmen (z. B. Raiffa, 1961).

⁵ Eine in meinen Augen diesbezüglich sehr überzeugende Analyse der Ellsberg Aufgabe gibt Raiffa (1961).

Vpn für Option A und ein Drittel für Option B . Zwei andere Gruppen von Vpn hatten dieses Problem als sogenannte *Matching* Aufgabe zu bearbeiten: Es wurden wiederum beide Optionen vorgegeben, bei jeweils einer der beiden waren die Kosten jedoch nicht spezifiziert. Die Aufgabe der Vpn bestand nun darin, für diese Option einen Kostenbetrag dergestalt festzulegen, dass beide Optionen gleich attraktiv wären. Wurde einer Person beispielsweise der Preis der Option B gezeigt und sie legte darauf hin den Preis der Option A mit 30 Millionen Dollar fest, so kann man schließen, dass sie B gegenüber der oben dargestellten Ausgangsoption A bevorzugt, da A mehr als die in der Matching-aufgabe spezifizierte Summe kostet. Unter der Matching Bedingung beurteilten nun 96% der Vpn das in der Choice Bedingung klar unterlegene Programm B als attraktiver. Welche Option als besser angesehen wurde, hing also in starkem Maße von der Art der Messung ab.

Tversky et al. (1988) erklären das Phänomen des Preference-Reversals damit, dass die beiden Abfragemodi die Aufmerksamkeit auf verschiedene Aspekte der Optionen lenken. Im Choice Paradigma entscheiden Personen eher lexikographisch, d. h. anhand des wichtigsten Aspektes der zu vergleichenden Optionen – in unserem Fall also anhand der Zahl der geretteten Menschenleben. Diese Strategie erspart einem, zwei mehr oder minder inkomensurable Größen (Geld und Menschenleben) miteinander in Beziehung setzen zu müssen. Im Matching Paradigma kommt man um diesen Vergleich jedoch nicht herum, und die Aufmerksamkeit wird hier stärker auf die einzustellende Größe, Geld, gelenkt, mit der Folge, dass dieser Dimension beim Vergleich der Optionen ein stärkeres Gewicht zukommt.

Beim Vergleich zweier Lotterien ist die Gewinnwahrscheinlichkeit typischerweise die wichtigere Dimension. So fanden etwa Lichtenstein & Slovic (1971) bei den folgenden zwei Wahlmöglichkeiten mit quasi identischen positiven Erwartungswerten

- (A) Sie bekommen 4\$ mit einer Chance von 99%, andernfalls verlieren Sie 1\$.
- (B) Sie bekommen 16\$ mit einer Chance von 33%, andernfalls verlieren Sie 2\$.

eine Präferenz für A . Mussten Vpn jedoch Geld bieten, um eine Option spielen zu dürfen, so waren sie bereit, für B mehr Geld auszugeben. Da die Auszahlungsbeträge hier ähnlich hoch waren, lässt sich das Wahlverhalten nicht mit dem Prinzip des abnehmenden Grenznutzens erklären. Statt dessen ist davon auszugehen, dass der erwartete Nutzen beider Optionen gleich hoch war. Vielmehr zeigt sich hier eine generell vorhandene Tendenz der Risikoaversion im Gewinnbereich (Kahneman & Tversky, 1979). Im allgemeinen ist uns also der Spatz in der Hand tatsächlich lieber als die Taube auf dem Dach.

Eine mögliche mathematische Modellierung dieses Phänomens stellen Tversky et al. (1988) in ihrem *Contingent Weighting* Ansatz dar. Aus diesem lässt sich die Entscheidungsregel ableiten, dass Spiel i , in dem entweder ϵ_i mit der Wahrscheinlichkeit p_i gewonnen wird oder andernfalls nichts, gegenüber Spiel j , in welchem ϵ_j mit der Wahrscheinlichkeit p_j gewonnen wird oder andernfalls nichts, dann und nur dann vorgezogen wird wenn gilt, dass

$$\epsilon_i p_i^0 \geq \epsilon_j p_j^0.$$

Der Exponent θ ist dabei eine Variable, über die sich regeln lässt, ob sich Wahlen insgesamt stärker an der Wert- oder an der Wahrscheinlichkeitskomponente orientieren. Bei einem θ von eins ist die relative Wichtigkeit von Wahrscheinlichkeits- und Wertinformation gleich groß. Mit wachsendem θ vergrößert sich die Variabilität der Wahrscheinlichkeitskomponente und somit wird auch ihr relatives Gewicht bei der Entscheidung größer. Umgekehrt verhält es sich für kleiner werdende Werte. Wenn sich also eine Person zwischen den beiden erwartungswertgleichen Optionen

- (A) Sie erhalten 95,- DM sicher.
- (B) Sie erhalten 100,- DM mit einer Chance von 95%, andernfalls nichts.

für die sichere Option A entscheidet und somit der Wahrscheinlichkeit eine höhere Priorität beimisst als der Gewinnhöhe, so lässt sich dies darstellen, indem wir etwa für θ den Wert zwei annehmen. Die wahrgenommene Attraktivität der ersten Option wäre dann $1^2 \cdot 95 \text{ DM} = 95 \text{ DM}$, die der zweiten jedoch nur $.95^2 \cdot 100 \text{ DM} = 90,25 \text{ DM}$.

3.2.3 Das Allais-Paradox

Eine dritte Art systematischen Verstoßens gegen die SEUT-Axiome stellt das sogenannte Allais-Paradox dar: Stellen Sie sich vor, Sie hätten die angenehme Wahl zwischen den folgenden Optionen A und B .

- (A) Sie erhalten 2 500,- DM mit $p = .33$, andernfalls nichts.
- (B) Sie erhalten 2 400,- DM mit $p = .34$, andernfalls nichts.

Wie würden Sie sich entscheiden? Wenn Sie A wählen, entscheiden Sie wie die große Mehrheit bei diesem Problem (Kahneman & Tversky, 1979).

Überdenken Sie nun die beiden nachfolgenden Optionen C und D :

- (C) Sie erhalten entweder 2 500,- DM (mit $p = .33$) oder 2 400,- DM (mit $p = .66$) oder nichts (mit $p = .01$).
- (D) Sie erhalten 2 400,- DM mit Sicherheit.

Wie würden Sie sich nun entscheiden? Wenn Sie D wählen, entscheiden Sie wiederum wie die große Mehrheit. Das von den meisten Personen gezeigte Wahlmuster A/D verstößt jedoch gegen das Sure Thing Prinzip (ebenso die Wahlen B/C). Denn A lässt sich auf dieselbe Weise in C umwandeln wie B in D , nämlich durch hinzufügen einer Chance von $p = .66$, 2 400 DM zu gewinnen. Die Hinzufügung identischer Konsequenzen zu den Optionen A und B führt also zu einer Präferenzumkehr – und genau dies verbietet das Sure Thing Prinzip. Eine andere Analyse macht den Widerspruch, der

in den Wahlen A/D steckt vielleicht deutlicher. $u(x)$ sei der Nutzen des Geldbetrages x . Dann folgt aus der Wahl von A die folgende Ungleichung:

$$(1) .33 \cdot u(2\ 500) > .34 \cdot u(2\ 400)$$

Wenn wir $u(0) = 0$ setzen so ergibt sich aus der Wahl von D die folgende Ungleichung:

$$(2) u(2\ 400) > .33 \cdot u(2\ 500) + .66 \cdot u(2\ 400)$$

Diese lässt sich durch Subtraktion von $.66 u(2\ 400)$ umwandeln in

$$(3) .34 \cdot u(2\ 400) > .33 \cdot u(2\ 500),$$

was im Widerspruch zu (1) steht.

Ein ähnlich gelagertes Problem stammt ebenfalls von Allais. Bei der Wahl zwischen den beiden folgenden Optionen

- (A) Sie erhalten 4 000,- DM mit $p = .8$, andernfalls nichts.
- (B) Sie erhalten 3 000,- DM sicher.

entscheidet sich ein klare Mehrheit für die sichere Variante. Bei der Wahl zwischen den folgenden Optionen C und D

- (C) Sie erhalten 4 000,- DM mit $p = .2$, andernfalls nichts.
- (D) Sie erhalten 3 000,- DM mit $p = .25$, andernfalls nichts.

erweist sich für eine deutliche Mehrheit C als attraktiver (Kahneman & Tversky, 1979). Auch dieses Wahlmuster stellt eine systematische Verletzung der SEUT dar. Option C kann nämlich beschrieben werden als $(A, .25)$ und Option D als $(B, .25)$. Wenn B nun attraktiver ist als A , so gebietet Axiom VI, dass auch alle (Bp) attraktiver sein sollen als die entsprechenden (Ap) ; bei der Bevorzugung von C gegenüber D ist jedoch das Gegenteil der Fall. Offensichtlich schadet die Verteilung der Gewinnwahrscheinlichkeit der Attraktivität einer sicheren Option mehr als derjenigen einer lediglich recht wahrscheinlichen Option. Diese allgemein feststellbare überproportionale Attraktivitätsminderung, die sich aus der Umwandlung eines sicheren in ein unsicheres Ereignis ergibt, wird als *Certainty Effect* bezeichnet. Dieser zeigt sich auch im Allais Paradox.

3.3 Das Konzept der Entscheidungsgewichte

Dass die hier vorgestellten Fehler *systematisch* auftreten, ist, wie bereits erwähnt, der Grund dafür, dass sich die SEUT nur bedingt zur Deskription menschlichen Entscheidungsverhaltens eignet. Es eröffnet aber gleichzeitig die Möglichkeit einer anderen Beschreibung. Hier hat sich das Konzept des *Entscheidungsgewichtes* („decision weight“) etabliert. Wenn der sichere Gewinn einer Geldsumme x

den positiven Nutzen $u(x)$ hat, so ist natürlich jede unsichere Option (xp) auf diesen Geldbetrag weniger attraktiv. Ginge es darum, den Betrag x zu verlieren, wäre natürlich umgekehrt die unsichere Option (xp) weniger unattraktiv. In der SEUT erfolgt diese Attraktivitätsminderung durch die Multiplikation des Nutzens mit der subjektiven Wahrscheinlichkeit p_{subj} , so dass sich $u(xp) = p_{\text{subj}} \cdot U(x)$ ergibt. Die Idee der Entscheidungsgewichte (diese seien mit w bezeichnet) ist nun, die Verringerung der Attraktivität durch die Multiplikation mit einem Wert w zu beschreiben, der eine Funktion der subjektiven Wahrscheinlichkeit darstellt. Diese Idee ist relativ alt und taucht bereits 1954 bei Edwards auf. Für die Gewichte gilt „[that they] are not probabilities: they do not obey the probability axioms and they should not be interpreted as measures of degree or belief“ (Kahneman & Tversky, 1979, S. 280). Die Gewichtungsfunktion $w = f(p_{\text{subj}})$ lässt sich ebenso wie die Nutzenfunktion $U = f(x)$ empirisch ermitteln.

Das Wahlverhalten gegenüber den beiden folgenden, erwartungswertgleichen Optionen (in den eckigen Klammern steht, wieviel Prozent der Versuchspersonen sich bei Kahneman und Tversky, 1979, für die jeweilige Alternative entschieden)

- (A) Sie erhalten 6 000,- DM mit $p = .25$, andernfalls nichts [18%].
- (B) Sie erhalten entweder 4 000,- DM (mit $p = .25$) oder 2 000,- DM (mit $p = .25$) oder nichts (mit $p = .5$) [82%].

zeigt, dass – wie bereits im Prinzip des abnehmenden Grenznutzens formuliert – die Nutzenfunktion im Gewinnbereich konkav verläuft (vgl. Abb. 4). Das entgegengesetzte Wahlverhalten, das sich bei der Wahl zwischen den Verlustoptionen C und D zeigt,

- (C) Sie verlieren 6 000,- DM mit $p = .25$, andernfalls nichts [70%].
- (D) Sie verlieren entweder 4 000,- DM (mit $p = .25$) oder 2 000,- DM (mit $p = .25$) oder nichts (mit $p = .5$) [30%].

deutet auf einen konvexen Verlauf im Verlustbereich hin. Zweimal 100 DM zu verlieren sollte also schwerer wiegen als der einmalige Verlust von 200 DM. Ein gutes Gespür für den Verlauf der Nutzenfunktionen im Gewinn- und Verlustbereich zeigt Machiavelli (1513/1986), wenn er Herrschern rät „[man muss] alle Gewalttaten auf einmal begehen, damit sie weniger fühlbar werden und dadurch weniger verletzen; Wohltaten hingegen muss man nach und nach erweisen, damit sie besser wahrgenommen werden“ (S. 73).

Bei den gerade betrachteten Wahlen ist es unproblematisch, dass das Verhältnis zwischen subjektiven Wahrscheinlichkeiten und Entscheidungsgewichten bislang unbekannt ist, da nur eine einzige Wahrscheinlichkeit relevant ist und somit die Attraktivität aller Geldbeträge auf eine zwar unbekannte aber gleiche Art verändert wird.

Ist die Nutzenfunktion nach diesem Prinzip bestimmt worden, kann jetzt auch die Gewichtungsfunktion nach folgendem Schema identifiziert werden (Beispiele wiederum aus Kahneman & Tversky, 1979):

- (A) Sie erhalten 6 000,- DM mit $p = .45$, andernfalls nichts [14%].
- (B) Sie erhalten 3 000,- DM mit $p = .90$, andernfalls nichts [86%].

Somit ergibt sich:

$$w(.45) \cdot u(6\,000) < w(.90) \cdot u(3\,000)$$

Da die Nutzenfunktion im Gewinnbereich konkav ist, gilt:

$$w(.45) / w(.90) < u(3\,000) / u(6\,000) < .5$$

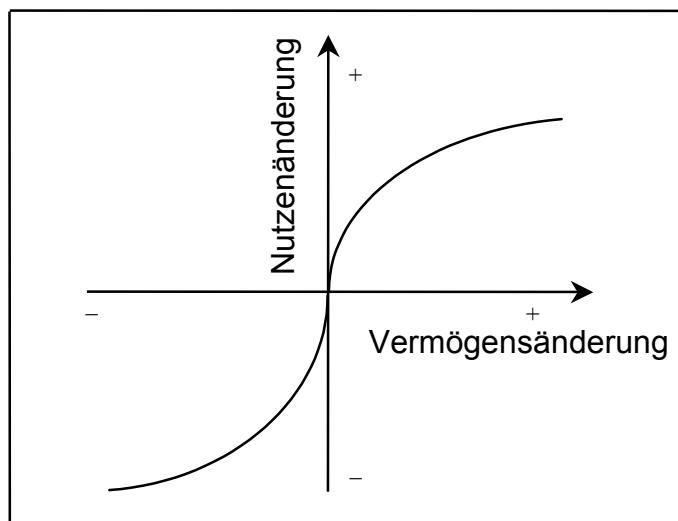


Abb. 4 Typischer Verlauf von Nutzenfunktionen im Gewinn- und Verlustbereich.

Im Bereich größerer Wahrscheinlichkeiten scheinen die Gewichte also subadditiv zu sein, $w(.45) + w(.45) < w(.90)$; auch der Certainty Effekt deutet auf diese Eigenschaft hin. Damit zwei Lotterien, bei denen die Gewinnchancen .90 und .45 betragen, als gleich attraktiv wahrgenommen werden, muss diejenige mit der schlechteren Gewinnchance im Gewinnfall also *mehr* als den doppelten Nutzen der vergleichsweise sicheren Lotterie einbringen. Anders verhält es sich bei kleinen Wahrscheinlichkeiten, wie das Wahlmuster bei den folgenden Alternativen zeigt (Kahneman & Tversky, 1979):

- (A) Sie erhalten 6 000,- DM mit $p = .001$, andernfalls nichts [73%].
- (B) Sie erhalten 3 000,- DM mit $p = .002$, andernfalls nichts [27%].

Hieraus folgt $w(.001) / w(.002) > .5$; die Chance von zwei Promille ist deutlich weniger als doppelt so attraktiv wie die Chance von eins in tausend. Kahneman und Tversky (1979) schlagen für Wahrscheinlichkeiten eine Gewichtungsfunktion vor, die annähernd s-förmig ist. S-förmige Funktionen postulieren auch Allais (1986), Hogarth und Einhorn (1990) und Lopes (1995); Abb. 5 zeigt eine empirisch ermittelte Funktion (Tversky & Fox, 1995), die den theoretischen Vorstellungen sehr gut entspricht; Tversky und Kahneman (1992) finden ebenfalls eine solche Funktion.

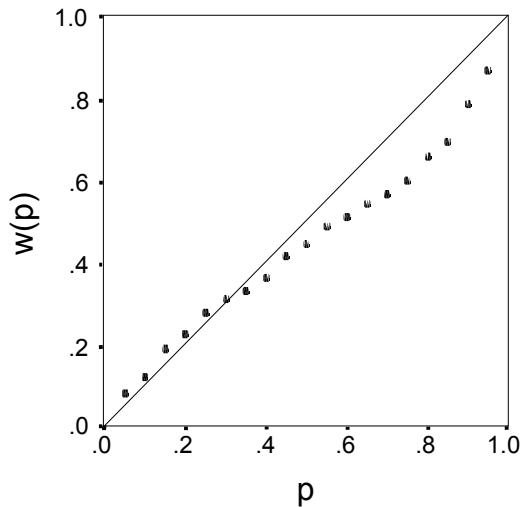


Abb. 5 Empirisch ermittelte Funktion für die Transformation von Wahrscheinlichkeiten in Entscheidungsgewichte $w(p)$. Aus Tversky & Fox, 1995.

Die vorgestellte Nutzen- sowie die Gewichtungsfunktion sind ein Kernstück von Kahneman und Tverskys *Prospect Theorie* (1979). Diese erhebt den Anspruch, solche Entscheidungen unter Unsicherheit zu beschreiben, bei denen objektive, klar umrissene Wahrscheinlichkeiten vorliegen. Der Vollständigkeit halber sei angemerkt, dass die Prospect Theorie nicht nur in Hinsicht auf das Konzept von Entscheidungsgewichten von der SEUT abweicht; die anderen Unterschiede sind im vorliegenden Kontext jedoch nicht relevant und sollen deshalb hier nicht behandelt werden.

Auch der Exponent θ aus dem Contingent Weighting Ansatz (s. 3.2.2) ist eine Gewichtung. Während aber die Entscheidungsgewichte w der Prospect Theorie die Wahrscheinlichkeiten in eine Beziehung zueinander setzen, indem diesen je nach ihrer Höhe eine unterschiedliche Bedeutung zukommt, regelt θ das Verhältnis der Wahrscheinlichkeiten insgesamt zu den Werten der Lotterien.

3.4 Die Venture Theorie

Der Geltungsbereich der Prospect Theorie beschränkt sich auf solche Entscheidungen, bei denen die beteiligten Wahrscheinlichkeiten wohldefiniert sind. Bei den meisten Alltagsproblemen ist dies aber nicht der Fall. Wenn etwa jemand überlegt, ob er die Mühe auf sich nimmt, sich einer medizinischen Vorsorgeuntersuchung zu unterziehen, so wird für diese Entscheidung mit ausschlaggebend sein, wie groß er das Erkrankungsrisiko einschätzt. Kaum ein Laie wird aber mit diesem Risiko eine

klar umrissene Wahrscheinlichkeit verbinden. Die 1990 von Hogarth und Einhorn vorgestellte *Ven-ture* Theorie stellt nun eine Erweiterung der Prospect Theorie dar, die beansprucht, nicht nur solche Entscheidungen unter Unsicherheit modellieren zu können, die auf präzisen und wohldefinierten Wahrscheinlichkeiten beruhen, sondern auch solche, bei denen die Informationen über die relevanten Wahrscheinlichkeiten unpräzise also ambige sind. Von der Prospect Theorie übernimmt sie (neben Aspekten, die für die vorliegende Arbeit nicht relevant sind und daher hier auch nicht vorgestellt werden sollen) die Idee einer abflachenden Nutzenkurve sowie die Vorstellung, dass Wahrscheinlichkeiten in Entscheidungsgewichte transformiert werden, wobei auch hier ein s-förmiger Zusammenhang angenommen und der Wertebereich auf das Intervall von null bis eins beschränkt wird.

Die Autoren gehen davon aus, dass Entscheidungsgewichte im Verlauf eines zweistufigen Prozesses entstehen. Dieser wird in Anlehnung an die von Tversky und Kahneman (1974) postulierte *Anchoring and Adjustment* Heuristik beschrieben: Am Beginn soll ein einfacher Ankerwert quasi probehalber als Entscheidungsgewicht dienen; in einer typischen experimentellen Aufgabe soll es sich dabei um die vom Experimentator vorgegebene Wahrscheinlichkeit handeln, unter alltags-näheren Bedingungen kann es sich um einen erfahrungsgelenkten Schätzwert, ein Expertenurteil oder ähnliches handeln. Im Falle ambiger Wahrscheinlichkeiten bleibt die Theorie bezüglich der Ankersetzung etwas nebulös: „In ambiguous circumstances, the anchor is assumed to be some initial value of the probability, that is typically available to the decision maker“ (Hogarth & Einhorn, 1990, S. 783). Der Anker- oder Startwert wird dann durch die mentale Simulation anderer möglicher Werte verändert⁶. Dieser Prozess soll durch vier Faktoren beeinflusst werden, die bestimmen, wie stark und in welche Richtung der Endwert, der dann dem Entscheidungsgewicht entspricht, vom Ankerwert abweicht.

Den ersten Faktor (i) bezeichnen Hogarth und Einhorn (1990) als Ergebnisunsicherheit („outcome uncertainty“). Wenn bei einer Lotterie ein Betrag ϵ mit der Wahrscheinlichkeit P gewonnen wird und andernfalls nichts, so besteht bei $p = 1$ bzw. $p = 0$ keine Ergebnisunsicherheit, bei $p = .5$ ist sie hingegen maximal. Auch die Häufigkeit, mit der eine Lotterie gespielt wird, beeinflusst die Ergebnisunsicherheit. Wenn wir davon ausgehen, dass eine Lotterie mit einer Gewinnwahrscheinlichkeit von .5 oft gespielt wird, so ist der Gesamtausgang relativ gut kalkulierbar: höchstwahrscheinlich wird er dem Erwartungswert relativ nahe kommen. In diesem Fall wird die Ergebnisunsicherheit also ähnlich gering werden wie bei $p = 1$. Wird eine solche Lotterie aber nur einmal gespielt, so kann nur gewonnen oder verloren werden; beide Endzustände sind vom Erwartungswert weit entfernt, eine fundierte Vorhersage des Ergebnisses lässt sich nicht machen, die Ergebnisunsicherheit ist somit hoch. Die Autoren gehen davon aus, dass eine größere Ergebnisunsicherheit zu einer verstärkten Vorstellung anderer Wahrscheinlichkeitswerte und somit auch zu einer stärkeren Modifikation des Ankerwertes führt.

Gleiches gilt (ii) für den Faktor Ambiguität. Hier wird angenommen, dass undeutliche Wahrscheinlichkeitsinformation eine stärkere gedankliche Beschäftigung anregt, die wieder zu einer stärkeren Abweichung des Entscheidungsgewichtes vom Ankerwert führt. Der gleiche Effekt soll auf-

treten, wenn es sich um ein Ereignis handelt, an das ein vergleichsweise hoher Nutzen bzw. ein relativ schwerer Schaden gekoppelt ist (iii). Im Vergleich zu Ereignissen, deren Konsequenzen eher belanglos sind, soll auch hier ein Mehr an gedanklicher Auseinandersetzung dazu führen, dass die Entscheidungsgewichte stärker vom Anker abweichen. Schließlich soll (iv) bedeutsam sein, ob die anstehende Entscheidung Gewinne oder Verluste betrifft. Die Autoren unterstellen eine generell konservative Haltung bezüglich Risiko. Somit sollen die Entscheidungsgewichte im Gewinnbereich eher unterhalb des Ankerwertes liegen; umgekehrt sollten im Verlustfall die Entscheidungsgewichte eher höher sein als der Anker.

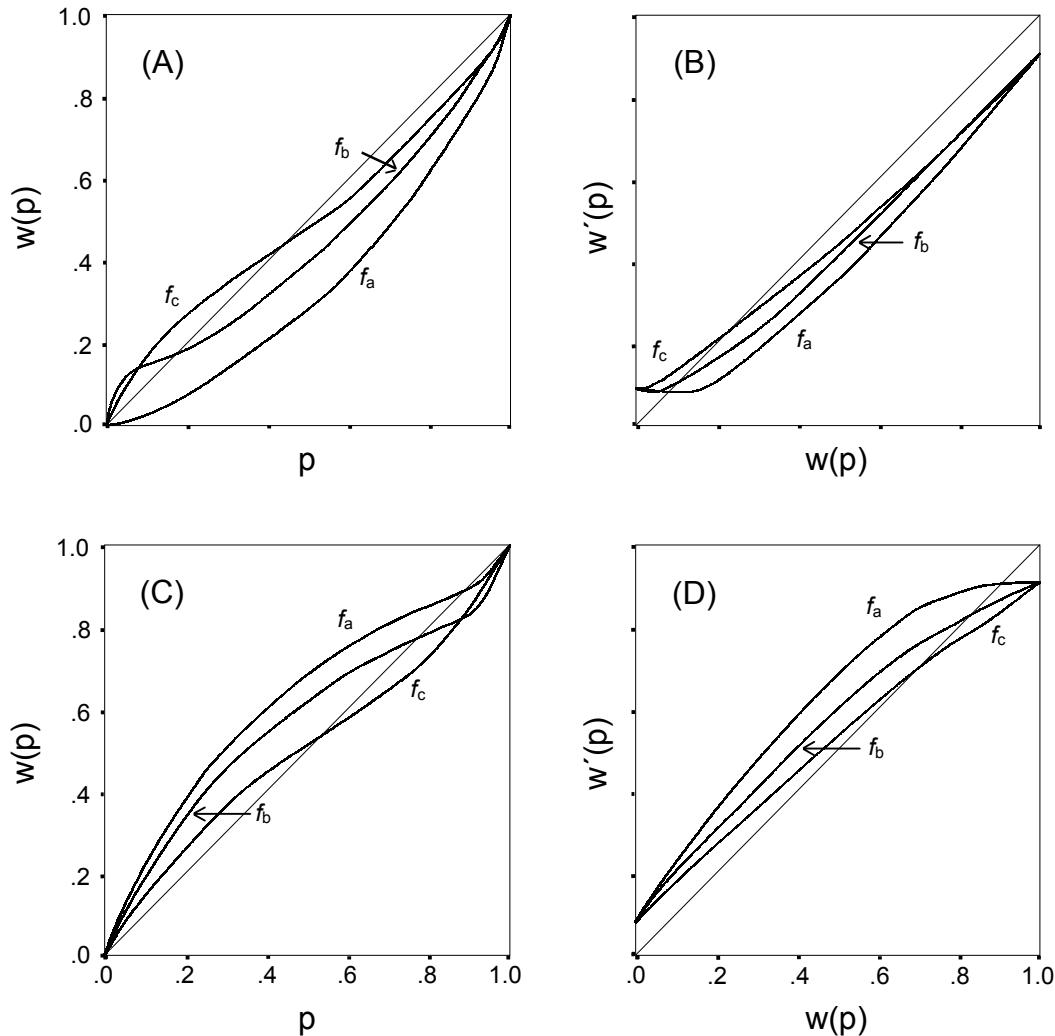


Abb. 6 Von Hogarth und Einhorn (1990) in der Venture Theorie postulierte Funktionen für die Umwandlung von Wahrscheinlichkeitsinformation in Entscheidungsgewichte. Die Quadranten A und B gelten für den Gewinn-, C und D für den Verlustbereich. A und C zeigen die Umwandlung präziser Wahrscheinlichkeiten. B und D zeigen Funktionen für ambige Wahrscheinlichkeiten; sie erhalten als Input den Output der Funktionen in den linken Quadranten; in diesem Fall bezeichnet p dann die Ankerwerte. Die Funktionen sind abhängig von der Größe des Payoffs. Je weiter der Indexbuchstabe einer Funktion im Alphabet vorne steht, desto höher sind die Payoff-Beträge, für die diese Funktion gilt.

⁶ Warum dies auch dann geschehen soll, wenn die Wahrscheinlichkeiten eindeutig sind, wird nicht erläutert.

Der linke obere Quadrant der Abb. 6 zeigt, wie nach den Vorstellungen der Theorie genau bestimmbarre Wahrscheinlichkeiten in Abhängigkeit von der Gewinnhöhe in Entscheidungsgewichte überführt werden. Dem rechten oberen Quadranten ist zu entnehmen, wie bei Vorliegen ambiger Wahrscheinlichkeitsinformation Ankerwerte (wiederum in Abhängigkeit von der Gewinnhöhe) in Entscheidungsgewichte transformiert werden. Die oberen Darstellungen beziehen sich auf den Gewinnbereich; der untere Bereich zeigt die entsprechenden Zusammenhänge im Verlustbereich.

Neben der Venture Theorie gibt es noch eine Fülle anderer Ansätze, die Entscheidungen unter Ambiguität zu beschreiben suchen; einen Überblick geben Camerer und Weber (1992). Wie später klar werden wird, ist es für die Zwecke dieser Arbeit jedoch hinreichend, allein die Venture Theorie zu betrachten.

3.5 Zusammenfassung

Drei Phänomene illustrieren, dass Menschen bei Entscheidungen unter Unsicherheit systematisch vom Prinzip der Maximierung des erwarteten Nutzens abweichen. (i) Der Grad der Präzision der Wahrscheinlichkeitsinformation (d. h. deren Ambiguität) beeinflusst das Entscheidungsverhalten. (ii) Je nach Aufgabentyp kommt der Wert- oder der Wahrscheinlichkeitskomponente der relevanten Optionen ein höheres Gewicht zu. Dies lässt sich über die Höhe des Exponenten θ in der Ungleichung von S. 23 beschreiben; diese Art von Entscheidungsgewicht beschreibt also die Wichtigkeit, die der Wert- im Vergleich zur Wahrscheinlichkeitsdimension zukommt. (iii) Als Certainty Effekt wird das Phänomen beschrieben, dass bei Gewinnoptionen eine Minderung der Gewinnwahrscheinlichkeit um einen festen Betrag schmerzlicher ist, wenn die Gewinnwahrscheinlichkeit vorher 1 betrug als wenn sie vorher geringer als 1 war. Diesem und verwandten Phänomenen kann Rechnung getragen werden, indem die subjektiven Wahrscheinlichkeiten einer entsprechenden Transformation unterzogen werden. Die daraus resultierenden Werte w werden ebenfalls als Entscheidungsgewichte bezeichnet; im Gegensatz zu θ beschreiben sie aber die relative Bedeutung von verschiedenen Wahrscheinlichkeiten zueinander und lassen die Frage, wie wichtig die Wahrscheinlichkeits- im Vergleich zur Wertdimension ist, unberührt. Auf das Konzept der Wahrscheinlichkeitsgewichte w greift auch die Venture Theorie zurück, um Entscheidungen unter Ambiguität zu beschreiben.

4. FRAGESTELLUNG

In Kapitel 4 wird die Fragestellung der vorliegenden Arbeit vorgestellt.

4.1 Die Gewichtungshypothese

Wie in Kapitel 3.2.1 geschildert wurde, übt es auf die Attraktivität von Optionen einen systematischen Einfluss aus, ob Wahrscheinlichkeitsinformationen präzise oder aber verschwommen, ambig, sind. Eine Fülle von Untersuchungen hat sich dieses Phänomens angenommen – in der Regel ist die Wirkung dergestalt, dass (bei Gewinnaussichten) ambige Optionen weniger attraktiv sind als nicht ambige – und eine kaum minder große Zahl theoretischer Ansätze sucht, diesem Befund gerecht zu werden (vgl. Camerer & Weber, 1992). Ein Einfluss von Ambiguität auf die Attraktivität von Handlungsoptionen findet sich nicht nur für die relativ alltagsferne Urnenaufgabe, anhand derer das Phänomen hier eingeführt wurde, sondern auch für alltagsnähere Szenarien. So zeigten sich etwa in einer Studie von MacCrimmon (1968) Manager eher geneigt, in solche Länder zu investieren, über die Informationen über ihre Wirtschaftsgeschichte vorlagen (niedrige Ambiguität) als in solche, über die keine solche Daten existierten (hohe Ambiguität).

Es liegt nahe, neben dem häufig untersuchten Einfluss von Ambiguität auf die Attraktivität von Optionen einen zweiten Effekt von Ambiguität auf das Entscheidungsverhalten zu vermuten, dem aber bislang nicht systematisch nachgegangen wurde: Dass nämlich dann, wenn die Wahrscheinlichkeitsinformation zunehmend ambiger wird, Entscheider ihre Strategie dahingehend wechseln, dass sie der Wertinformation ein höheres Gewicht beimessen, so dass letztere also für ihr Urteil wichtiger wird. Dies postulieren auch Wallsten, Budescu und Tsao (1997): „When trading-off among dimensions for the purpose of choosing or evaluating alternatives, the weight accorded a dimension is a positive function of its precision“ (S. 32). Diese Vermutung soll *Gewichtungshypothese* heißen.

Die Venture Theorie macht keine Aussage zu diesem Punkt; d. h. sie geht zumindest implizit davon aus, dass der Grad der Ambiguität keinen Einfluss darauf hat, welches Gewicht der Wahrscheinlichkeitsdimension zukommt. Mir ist auch sonst keine Theorie bekannt, die den möglichen Einfluss von Ambiguität auf die Gewichtung der Wahrscheinlichkeitsinformation berücksichtigt.

4.2 Ambiguität – was ist das eigentlich genau?

Obwohl das Konzept ambiger Wahrscheinlichkeit ein Kernstück der Untersuchungsfrage ist, wurde es bislang nicht definiert. Aus gutem Grund. Die Definitionen von Ambiguität sind zahlreich und verschieden. Begründen wir uns zur Anschauung mit zweien.

Einhorn und Hogarth (1985) schreiben: „Ambiguity is an increasing function of the number of distributions that can not be ruled out [] by one's knowledge of the situation“ (S. 435). Diese Beschreibung von Ambiguität hebt also auf die Zahl möglicher Wahrscheinlichkeitsverteilungen ab; interessiert mich etwa die Wahrscheinlichkeit, eine schwarze Kugel aus einer Urne mit einem mir

unbekannten Mischungsverhältnis schwarzer und roter Kugeln zu ziehen, so bärge dem zufolge eine Urne mit nur einer Kugel weniger Ambiguität als eine mit zehn Kugeln, denn im ersten Fall gibt es lediglich zwei Möglichkeiten ($p = 1$ oder $p = 0$), im zweiten Fall sind die Möglichkeiten aber manigfach. Eine Stärke dieser Definition ist, dass sie hinreichend konkret ist, eine bestimmte Art der Messung von Ambiguität nahe zu legen; in einer Versuchssituation ist diese leicht herzustellen, die Übertragung auf alltägliche Phänomene dürfte jedoch schwer fallen. Denn wieviele mögliche Wahrscheinlichkeitsverteilungen gibt es für das Ereignis „Banküberfall mit Geiselnahme in der Woche X in der BRD“, und wieviele dieser Verteilungen kann ich aufgrund der Zusatzinformation, dass die Woche X lediglich drei Werkstage hat, ausschließen?

Während die Definition von Einhorn und Hogarth auf objektive Merkmale der Situation abhebt, betonen Frisch und Baron (1988) das subjektive Erleben: „We stress the subjective nature of ambiguity because all probability judgments reflect missing information“ (S. 152).

Für die Zwecke dieser Arbeit ist es nicht notwendig, sich auf eine bestimmte Definition von Ambiguität zu versteifen. Eine operationale Definition, die später mit der Beschreibung des hier verwendeten Paradigmas gegeben wird, reicht aus. Ob die mit dieser zu findenden Effekte dann auch mit anderen Arten unpräziser Wahrscheinlichkeitsinformation erzielt werden, bleibt dann eine empirische Frage.

4.3 Simulation I: Der Zusammenhang zwischen Erwartungswerten und Wahrscheinlichkeitsinformation bei zunehmender Ambiguität

Diffuser Wahrscheinlichkeitsinformation weniger Gewicht beizumessen, erscheint unmittelbar plausibel, denn diese sollte nur sehr ungenau Aufschluss über den Wert der betreffenden Option geben können. In welchem Ausmaß der Informationsgehalt der Wahrscheinlichkeitsinformation mit zunehmender Ambiguität sinken kann, soll anhand einer einfachen Computersimulation demonstriert werden. Für eine solche Simulation ist es notwendig, spezifische Randbedingungen festzulegen, und diese müssen keineswegs so auch für alle natürlichen Situationen gelten. Ich beanspruche daher nicht, dass sich die Ergebnisse eins zu eins auf natürliche Entscheidungen unter Ambiguität übertragen lassen; sie erscheinen mir aber geeignet, die Sinnhaftigkeit des in der Gewichtungshypothese postulierten Verhaltens plausibler zu machen.

Kern der Simulation ist ein artifizieller Mechanismus, der zufällig Lotterien erzeugt und diese dann (mit einem mehr oder minder hohen Maß an Ambiguität versehen) einem Entscheider vorlegt, der die Attraktivität der Lotterien einschätzen soll; der Einfachheit halber wollen wir die Attraktivität dabei mit dem Erwartungswert der Lotterie gleichsetzen. In jeder Lotterie wird entweder mit der Wahrscheinlichkeit P der Betrag ϵ gewonnen oder andernfalls nichts. Die Gewinnbeträge werden zufällig aus dem Intervall [3, 97] gezogen, die Wahrscheinlichkeiten ebenfalls zufällig aus dem Bereich [.03, .97]. Beide Variablen sind gleichverteilt und unabhängig voneinander. Ambiguität wird dergestalt erzeugt, dass dem Entscheider nicht der Wahrscheinlichkeitswert P mitgeteilt wird sondern statt dessen zwei Werte, P_{lower} und P_{upper} . Diese bilden die Unter- und die Obergrenze eines Intervalls, für das folgendes gilt: (i) die wahre Wahrscheinlichkeit P ist in diesem Intervall enthalten;

(ii) $P_{\text{lower}} \geq 0$ und (iii) $P_{\text{upper}} \leq 1$. Wie die Abb. 7 beispielhaft zeigt, ergibt sich für die Lage des ambiguitätsformenden Intervalls $[P_{\text{lower}}, P_{\text{upper}}]$ eine Vielzahl von Möglichkeiten. Für jede Lotterie wird dabei eine der möglichen Positionen zufällig ausgewählt. Je größer das Intervall $[P_{\text{lower}}, P_{\text{upper}}]$ ist, desto höher ist der Lotterie Ambiguität; deren Stärke definiere ich im gegebenen Kontext als $P_{\text{upper}} - P_{\text{lower}}$.

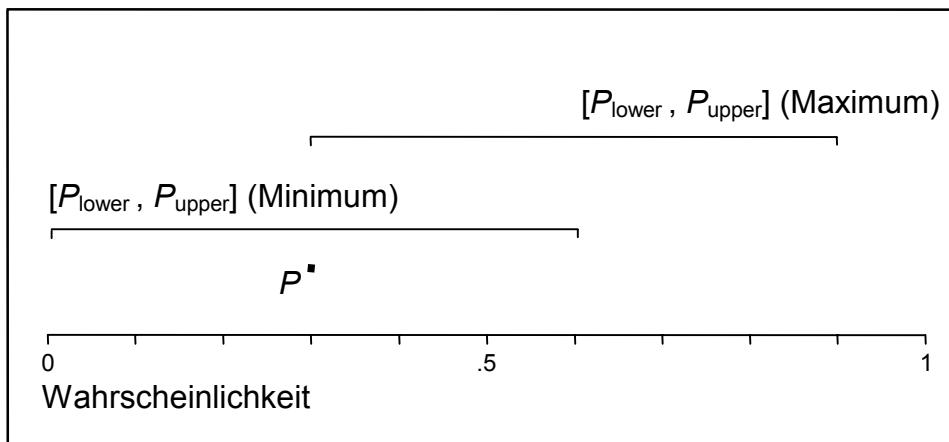


Abb. 7 Die beiden extremen Platzierungen für das ambiguitätsformende Intervall $[P_{\text{lower}}, P_{\text{upper}}]$ bei einer wahren Wahrscheinlichkeit von $P = .3$ und einer Ambiguität von 60%.

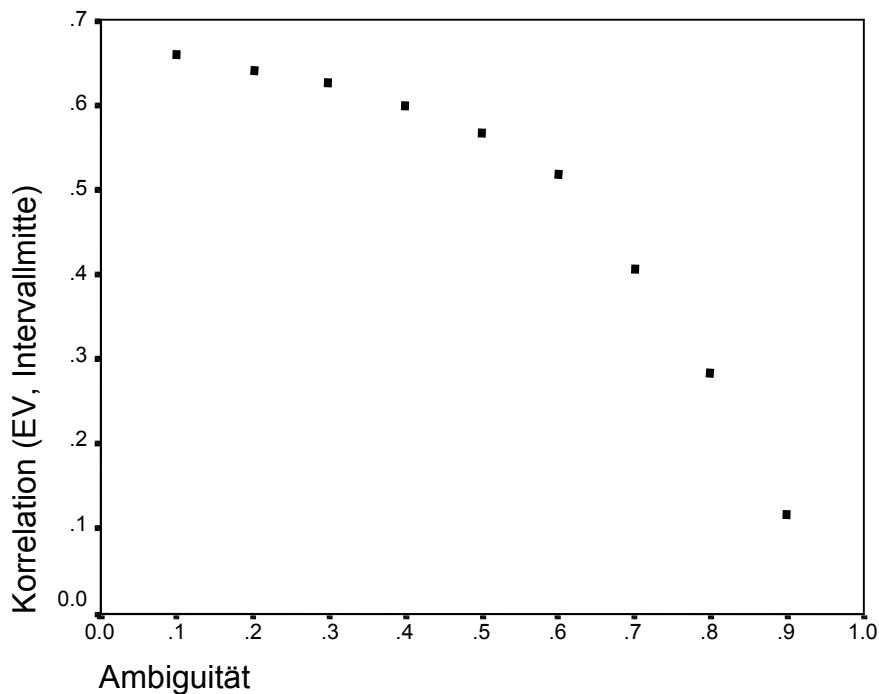


Abb. 8 Simulation I: Mit zunehmender Ambiguität verringert sich der Zusammenhang zwischen der Wahrscheinlichkeitsinformation und den Erwartungswerten der Lotterien. Jedem Datenpunkt liegen 10 000 Fälle zu Grunde.

Wie sich, in Abhängigkeit vom Grad der Ambiguität, die Produkt-Moment Korrelation zwischen den Erwartungswerten der Lotterien mit der Mitte der ambiguitätsformenden Intervalle [P_{lower} , P_{upper}] entwickelt, zeigt Abb. 8. Mit zunehmender Ambiguität fällt der Zusammenhang zwischen den Erwartungswerten der Lotterien und den Intervallmittelpunkten immer stärker ab. Der Zusammenhang zwischen ϵ und den Erwartungswerten wird von steigender Ambiguität nicht berührt. Somit erscheint es geboten, der Wertkomponente mit zunehmendem Rauschen in der Wahrscheinlichkeitsinformation ein immer stärkeres Gewicht bei der Bewertung der Attraktivität der Lotterien zukommen zu lassen.

4.4 Die Relevanz der Gewichtungshypothese

Im Bereich des Urteilens und Entscheidens unter Unsicherheit gibt es eine reiche Untersuchungstradition, die sich damit befasst, was sinnvolle normative Standards sind (z. B. Birnbaum, 1983), und untersucht, inwieweit Menschen diesen nachkommen. Beispielhaft seien hier nur zwei Gebiete erwähnt – die Frage, ob Menschen den Grad der Zuverlässigkeit ihres unsicheren Wissens adäquat einschätzen können (z. B. McClelland & Bolger, 1994), und die Frage, inwieweit sie mit bedingten Wahrscheinlichkeiten angemessen umgehen (z. B. Kahneman & Tversky, 1973). An diese Tradition, geeignete normative Maßstäbe für menschliches Verhalten unter Unsicherheit zu finden und dieses an ihnen zu messen, schließt die Erforschung der Gewichtungshypothese an.

Die Gewichtungshypothese ist dabei nicht nur von theoretischem Interesse, denn wenn diese, was plausibel erscheint, tatsächlich zuträfe, könnte dies auch weitreichende praktische Konsequenzen haben. Denn wenn Menschen Entscheidungen treffen, so beziehen sie in vielen Fällen das Wissen von Experten mit ein; insbesondere dann, wenn es sich um schwerwiegende medizinische, politische, ökonomische, technische oder juristische Entscheidungen handelt. Ein wichtiger Teil des Expertenwissens besteht darin, die Wahrscheinlichkeiten verschiedener Ausgänge beurteilen zu können. Dieses Wissen gilt es, den Entscheidern weiterzugeben. Menschen bevorzugen dabei im Allgemeinen Informationen über die Wahrscheinlichkeit entscheidungsrelevanter Ereignisse in numerischer Form zu erhalten, de facto fassen sie aber Grade subjektiver Überzeugung eher in Worte, benutzen also Formulierungen wie „höchst wahrscheinlich“, „eher nicht damit zu rechnen, dass ...“, etc. (Wallsten, Budescu, Zwick & Kemp, 1993) – was natürlich zur Folge hat, dass wir als Empfänger einer Botschaft diese in der Regel im ungeliebten Format erhalten. Dabei dürfte es sich auf Seiten des Senders im Allgemeinen nicht um ein reflektiertes Verhalten handeln. Wenngleich sich Ausnahmen finden (wie z. B. „fifty-fifty“), ist solchen verbalen Umschreibungen in der Regel ein hohes Maß an Ambiguität zu eigen (z. B. Mosteller & Youtz, 1990), und diese wird auf Seiten des Empfängers als stärker empfunden als auf Seiten des Senders (Fillenbaum, Wallsten, Cohen & Cox, 1991). Wenn Experten nun dadurch, dass sie Wahrscheinlichkeiten gewohnheitsgemäß in Worten ausdrücken, Ambiguität inflationieren und darüber, so die Gewichtungshypothese zutreffend ist, die Entscheidungen der Berateten unbewusst und systematisch beeinflussen, so ist mehr als fraglich, ob dies einen wünschenswerten Zustand darstellt.

4.5 Zusammenfassung

Mit zunehmender Ambiguität lässt sich der Erwartungswert einer Option nur noch ungenau schätzen. Je stärker die Ambiguität ausgeprägt ist, desto schwächer wird der Zusammenhang zwischen der verfügbaren Wahrscheinlichkeitsinformation und dem Erwartungswert. Daher erscheint es vernünftig, die Wahrscheinlichkeitsinformation mit steigender Ambiguität schwächer zu gewichten. Die Gewichtungshypothese postuliert, dass Menschen genau dies tun. Die vorliegende Arbeit geht den Fragen nach, ob die Gewichtungshypothese zutrifft (Kapitel 9 bis 12) und inwieweit ein solches Gewichtungsverhalten angemessen ist (Kapitel 13).

5. WAS BISHER GESCHAH, ...

In Kapitel 5 werden zwei Studien vorgestellt, die scheinbar geeignet sind, Licht auf die Gewichtungshypothese zu werfen.

Auch wenn die Gewichtungshypothese bislang nicht systematisch untersucht worden ist, gibt es doch zwei Arbeiten, die scheinbar in der Lage sind, diese zu stützen. Und wenn sich Wallsten et al. (1997) hinter den der Gewichtungshypothese zu Grunde liegenden Gedanken stellen (s. 4.1), so nicht allein aus Plausibilitätsgründen sondern auch unter Bezugnahme auf diese beiden Arbeiten, die ich im Folgenden vorstellen möchte: Die Studie von Gonzalez-Vallejo, Erev und Wallsten (1994) sowie die Arbeit von Gonzalez-Vallejo und Wallsten (1992).

5.1 Die Studie von Gonzalez-Vallejo, Erev und Wallsten (1994)

Die Untersuchung von Gonzalez-Vallejo, Erev und Wallsten (1994) wurde durch drei aufeinander aufbauende Arbeiten motiviert (Budescu, Weinberg & Wallsten, 1988, Budescu & Wallsten, 1990, Erev & Cohen, 1990). Die erste dieser Untersuchungen wurde durch eine Lücke im Bereich der Forschung zu Entscheidungen unter Unsicherheit angestoßen. So mussten die Autoren 1988 feststellen: „... there is virtually no literature dealing with decisions based on verbally expressed beliefs“ (S. 281). Bis zu diesem Zeitpunkt gab es also praktisch keine Arbeiten, die Entscheidungen, die auf verbalen Wahrscheinlichkeitsinformationen basieren, untersucht haben. Der Grund hierfür dürfte darin liegen, dass die SEUT gemeinsam mit moderneren Ansätzen, die sich aus dieser entwickelt haben, ein Konzept von subjektiver Wahrscheinlichkeit vertritt, nach dem selbige einen Grad subjektiver Überzeugung darstellt, der auf die reellen Zahlen abgebildet werden kann. Um diesen Grad subjektiver Überzeugung kontrolliert variieren zu lassen, bietet sich die Arbeit mit Lotterien an, deren Gewinnwahrscheinlichkeiten numerisch dargestellt werden, wie dies auch in den bisher gegebenen Beispielen der Fall war.

Eine solche Darstellungsweise ist aber insofern wenig realitätsangemessen, als dass Menschen, wie bereits erwähnt, im Allgemeinen zwar bevorzugen, Informationen über die Wahrscheinlichkeit entscheidungsrelevanter Ereignisse in numerischer Form zu erhalten, de facto aber Grade subjektiver Überzeugung eher in Worte fassen (Wallsten, Budescu, Zwick & Kemp, 1993). Während die alltägliche Informationsvermittlung also eher auf Worten basiert – eine Ärztin wir Ihnen vielleicht sagen, dass Ihr Risiko für ein bestimmtes Leiden aufgrund dieser und jener Faktoren „erhöht“ ist, sie wird dieses Risiko aber selten numerisch spezifizieren („die Gefahr beträgt bei Ihnen 15%“) – stellt die Laborforschung Chancen und Risiken von Wahlmöglichkeiten fast ausschließlich in Form von Zahlen dar (typische Beispiele sind etwa Kahneman & Tversky, 1979, Wu, 1994).

Von den vier genannten Untersuchungen, die sich mit Entscheidungen bei verbaler Wahrscheinlichkeitsinformation befassten, soll die letzte von Gonzalez-Vallejo, Erev und Wallsten (1994) ausführlich vorgestellt werden. Sie zieht die Quintessenz aus den vorangegangenen Studien, stärkt scheinbar die Gewichtungshypothese und dient meinen eigenen Untersuchungen insofern als Vorbild, als dass diese deren Paradigma in einem wesentlichen Teil übernehmen.

Die teilnehmenden Vpn wurden in zwei Gruppen eingeteilt, in Entscheider und Experten. Die Entscheider bekamen Glücksspiele in Gruppen zu sechs Lotterien präsentiert. Bei jeder dieser Lotterien wurde entweder ein festgelegter Betrag € gewonnen, was mit einer bestimmten Chance P geschah, oder es passierte nichts. Alle Spiele in einer Gruppe unterschieden sich sowohl hinsichtlich P als auch in Bezug auf €. Die Aufgabe der Entscheider bestand nun darin, alle sechs Spiele in eine Rangordnung hinsichtlich ihrer Attraktivität zu bringen; die subjektiv beste Lotterie sollte dabei den Rang sechs erhalten, die schwächste Rang eins. Zwölf Bündel von Glücksspielen waren solcherart zu bearbeiten. Zum Schluss des Experiments wurden verabredungsgemäß zwei dieser zwölf zufällig ausgewählt, und alle darin enthaltenen Lotterien wurden ausgespielt. Wenn eine Lotterie gewann, so wurde deren Wert € mal dem vergebenen Rang ausbezahlt, worüber die Entscheider vor dem Zeitpunkt der Rangvergabe informiert worden waren. Somit sollten die Entscheider motiviert gewesen sein, die Rangordnung tatsächlich nach der wahrgenommenen Attraktivität der Lotterien vorzunehmen, da sie auf diese Weise ihren erwarteten Nutzen optimierten. Während die Auszahlungsbeträge der einzelnen Lotterien für die Entscheider unmittelbar ersichtlich waren, hatten sie nur bedingte Informationen über die Gewinnchancen eines jeden Spiels.

Die Aufgabe der Experten bestand darin, genau diese Information bereitzustellen. Die Gewinnchance einer jeden Lotterie war über ein einfaches Computerspiel operationalisiert worden. Dabei bewegte sich ein kleines Objekt auf ein Ziel zu; wenn es dieses erreichte, war das Spiel gewonnen. Falls es jedoch von einer der auf dem Weg befindlichen, sich öffnenden und schließenden Barrieren berührt wurde, war das Spiel verloren. Die Experten hatten die Gelegenheit, das zu einer jeden Lotterie zugehörige Computerspiel mehrfach zu beobachten und somit zu einer Einschätzung der Gewinnchance dieser Lotterie zu kommen. Jeder Experte gab dabei für jede Lotterie eine Einschätzung der Gewinnchance in numerischer Form ab (d. h. als Prozentangabe) und eine in Form einer verbalen Beschreibung (z. B. „gute Chance“). Die Experten wurden dadurch zu möglichst akkuraten Beschreibungen motiviert, dass ihre Bezahlung von der Qualität der Entscheidungen der von ihnen informierten Entscheider abhing. Die zwölf Lotteriebündel, die jeder Entscheider zu bearbeiten

hatte, bestanden aus zwei identischen Blöcken à sechs Aufgaben. Im einen Block waren die Gewinnchancen jedoch numerisch und im anderen verbal beschrieben.

Vorhergehende Studien mit ähnlichem Aufbau (Budescu, Weinberg & Wallsten, 1988, Budescu & Wallsten, 1990, Erev & Cohen, 1990) hatten keinen einheitlichen Einfluss des Informationsformates (verbal vs. numerisch) auf die Qualität von Entscheidungen gefunden. Dies ist insofern überraschend, als dass das Verständnis von Formulierungen, mit denen Unsicherheit beschrieben wird, von Person zu Person deutlich variiert (Mosteller & Youtz, 1990, geben einen guten Überblick über die entsprechende Forschung). Dem entsprechend sollte die verbal gegebene Information also ein gewisses Maß an Rauschen mit sich bringen, dass zu einer Minderung der durchschnittlichen Entscheidungsgüte führen sollte. Warum ist dies nicht der Fall? Eine Vorstudie von Gonzalez-Vallejo, Erev & Wallsten (1994) hatte folgendes gezeigt: Wenn die Entscheidungen auf der Grundlage numerischer Wahrscheinlichkeitsinformationen gefällt wurden, so orientierte sich die Rangordnung, in die die Lotterien gebracht worden waren, vornehmlich an der Gewinnwahrscheinlichkeit der einzelnen Spiele; war die Information über die Gewinnchancen hingegen verbaler Natur, zeigte sich ein stärkerer Zusammenhang zwischen der Präferenzordnung und den Auszahlungsbeträgen der Spiele (vgl. Abb. 9). Dies führte dazu, dass die Leistung im numerischen Format nur dann derjenigen im verbalen überlegen war, wenn die Erwartungswerte der Lotterien stark mit der Wahrscheinlichkeitskomponente korrelierten. Umgekehrt fand sich eine Überlegenheit des verbalen Formats für die Spielesbündel, bei denen Erwartungswert und Wertkomponente eng zusammen hingen.

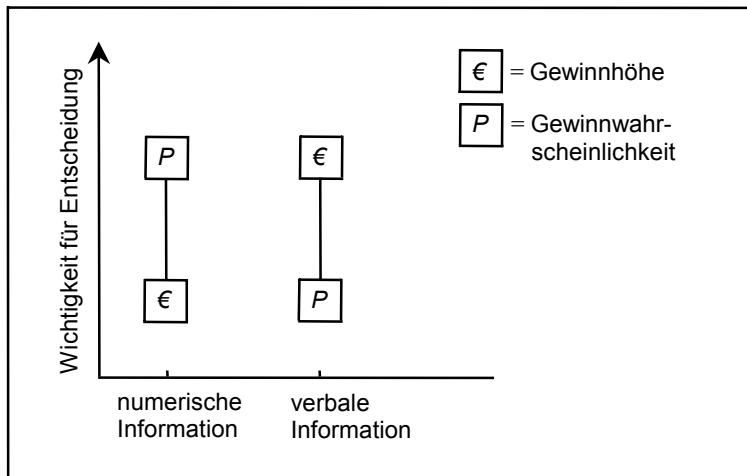


Abb. 9 Schematische Darstellung der Ergebnisse von Gonzalez-Vallejo et al. (1994). Wichtigkeit wurde dabei gemessen als $V(C, \epsilon)$.

Dies veranlasste Gonzalez-Vallejo et al. (1994) dazu, den Zusammenhang zwischen dem Erwartungswert auf der einen Seite sowie der Wert- bzw. Wahrscheinlichkeitskomponente auf der anderen Seite systematisch zu variieren, um den vermuteten Interaktionseffekt von Informationsformat und Korrelationsstruktur zu überprüfen.

Als Zusammenhangsmaß hat sich dabei die V -Korrelation (Nelson, 1984) etabliert. Wenn alle Mitglieder Z einer Menge einmal in die Rangreihe X und dann in die Rangreihe Y gebracht werden, so bestimmt sich die Übereinstimmung dieser beiden Rangreihen als $V(X, Y) = G / (G + UG)$. Dabei werden alle möglichen Kombinationen zweier Z gebildet; G ist die Zahl aller durch X und Y gleich geordneten Paare und UG die Zahl aller ungleich geordneten Paare. Z -Paare, die in X oder Y eine Rangbindung aufweisen, werden nicht berücksichtigt. Ein Beispiel mag die Berechnung der V -Korrelation anhand der folgenden drei Lotterien verdeutlichen:

	ϵ in DM	P	EV in DM
(A)	5	.5	2,5
(B)	8	.2	1,6
(C)	10	.4	4

Werden die Lotterien nun jeweils nach ϵ , P und EV geordnet, so ergeben sich die Rangordnungen $C/B/A$, $A/C/B$ und $C/A/B$. Insgesamt gibt es drei Paare; nachfolgend ist aufgelistet, inwieweit EV und ϵ sowie EV und P diese gleichsinnig ordnen; ein „+“ kennzeichnet dabei eine gleichsinnige, ein „–“ eine gegensinnige Ordnung.

Paar	EV/ϵ	EV/P
A/B	–	+
A/C	+	–
B/C	+	+

Somit ergibt sich für diese drei Lotterien ein gleich starker Zusammenhang zwischen Erwartungswert und Wertkomponente auf der einen Seite und Erwartungswert und Wahrscheinlichkeitskomponente auf der anderen Seite; $V(EV, \epsilon) = V(EV, P) = 2 / (2+1) \approx .67$. Nach dem selben Schema lässt sich auch bestimmen, inwieweit sich die Rangordnung, die eine Vp hergestellt hat, an ϵ und P orientiert.

Der V -Koeffizient kann Werte von null bis eins annehmen. Eins kennzeichnet einen perfekten positiven Zusammenhang, .5 das Fehlen eines Zusammenhangs und null einen perfekten negativen Zusammenhang.

Die Lotterien bei Gonzalez-Vallejo et al. (1994) waren so konstruiert, dass für jedes der sechs Spielesbündel galt $V(EV, \epsilon) + V(EV, P) = 1$; die einzelnen Korrelationen betrugen .80, .67, .53, .47, .33 und .20.

In Übereinstimmung mit den Ergebnissen der vorangegangenen Studien fand sich keine generelle Überlegenheit numerischer Wahrscheinlichkeitsinformation: Der Unterschied zwischen der Summe der erwarteten Gewinne aller Lotterien, so wurde Entscheidungsqualität operationalisiert, betrug weniger als 0,3% und erwies sich als nicht signifikant. In Übereinstimmung mit der Untersuchungshypothese fand sich jedoch eine Überlegenheit numerischer Information in Spielesbündeln, bei denen der Erwartungswert stark mit P korrelierte, während Entscheidungen auf Grundlage verbaler

Information zu besseren Ergebnissen führten, wenn der Erwartungswert hoch mit € korrelierte. Abb. 10 zeigt die Leistungsunterschiede in Abhängigkeit von $V(EV, \epsilon)$.

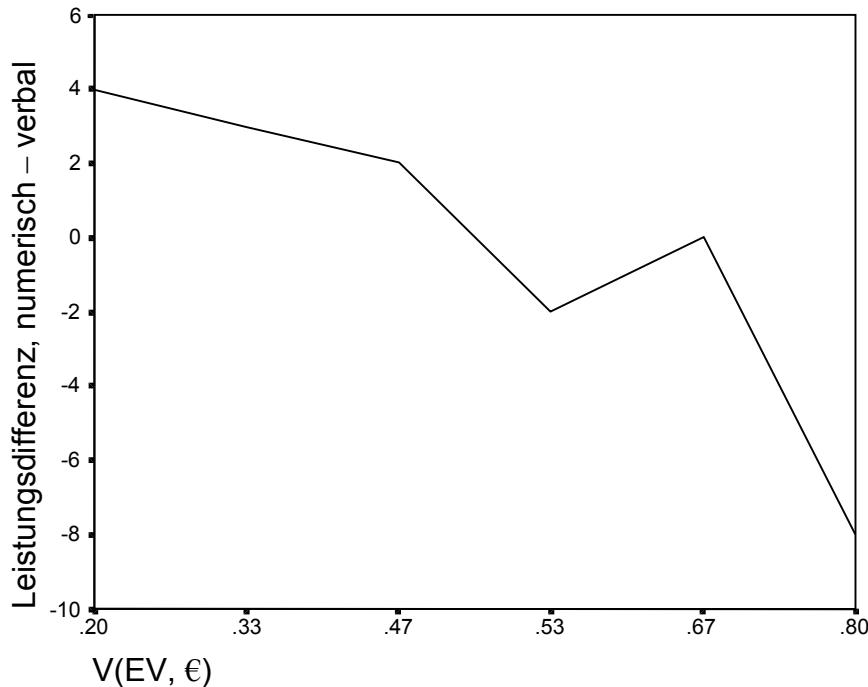


Abb. 10 Ergebnis von Gonzalez-Vallejo et al. (1994): Je höher die Erwartungswerte der Lotterien mit deren Werten korrelierten, desto günstiger wirkte sich verbale gegenüber numerischer Wahrscheinlichkeitsinformation aus. Beachte, $V(EV, P)$ war stets $1-V(EV, \epsilon)$.

Ebenfalls im Übereinstimmung mit der Untersuchungshypothese orientierten sich die von den Vpn vorgenommenen Rangordnungen (abgekürzt mit C für „choice“) bei verbaler Information stärker an der Wertkomponente der Lotterien als bei numerischer Information. Im ersten Fall betrug $V(C, \epsilon)$ im Mittel .40 und im zweiten .35, dieser Unterschied erwies sich als signifikant.

Die Befunde von Gonzalez-Vallejo et al. (1994) lassen sich so interpretieren, dass die Vpn in Abhängigkeit davon, ob sie Wahrscheinlichkeitsinformation in Worten oder Zahlen erhielten, ihre Entscheidungsstrategie wechselten: Im ersten Fall orientierten sie sich stärker an der Wertinformation als im zweiten. Scheinbar wurde dieser Dimension hier also eine höhere Bedeutung zugemessen, als dies bei numerischer Wahrscheinlichkeitsinformation der Fall war. Wenn man nun plausibler Weise annimmt, dass das ausschlaggebende Moment hierfür die Ambiguität war, die verbalen Wahrscheinlichkeitsbeschreibungen generell zu eigen ist, so scheint dieser Befund die Gewichtungshypothese zu stützen.

5.2 Die Studie von Gonzalez-Vallejo und Wallsten (1992)

Eine zweite Studie, die sich der Frage widmet, ob verbale Wahrscheinlichkeitsinformation dazu führt, dass sich Entscheider stärker auf die Wertdimension stützen, ist die von Gonzalez-Vallejo und Wallsten (1992). Um dieser Frage auf den Grund zu gehen, machten sie sich das Phänomen des

Preference Reversals zu Nutze. Wie bereits geschildert wird dieser Effekt damit erklärt, dass die relative Bedeutung, die der Wahrscheinlichkeitsdimension zukommt, größer ist, wenn Vpn zwischen Lotterien wählen als wenn sie für Lotterien bieten (vgl. 3.2.2). Abb. 11 illustriert dies noch einmal anhand zweier erwartungswertgleicher Lotterien. Diejenige mit der höheren Gewinnchance soll dabei P-Lotterie heißen und die, welche den größeren Gewinn verspricht, €-Lotterie. Die im jeweiligen Paradigma dominierende Dimension ist in der Abbildung optisch hervorgehoben. Typischer Weise würde bei einer Wahl zwischen beiden Möglichkeiten die P-Lotterie bevorzugt. Sollen Vpn jedoch angeben, wieviel sie zu zahlen bereit wären, um die jeweilige Lotterie spielen zu dürfen, so erzielt typischer Weise die €-Lotterie den höheren Wert.

Numerische Darbietung der Wahrscheinlichkeit				
Paradigma	Lotterietyp	ϵ	P	bevorzugt wird
Wählen	P-Lotterie	20,- DM	<u>.90</u>	***
	€-Lotterie	90,- DM	<u>.20</u>	
Bieten	P-Lotterie	<u>20,- DM</u>	.90	
	€-Lotterie	<u>90,- DM</u>	.20	***

Abb. 11 Schematische Darstellung des Preference Reversals. Die Lotteriedimension, die in dem jeweiligen Paradigma besondere Wichtigkeit besitzt, ist optisch hervorgehoben.

Der Wechsel des Paradigmas führt also zu einer Präferenzumkehr. Von der Gewichtungshypothese ausgehend ist zu erwarten, dass eine solche paradigmabedingte Präferenzumkehr nachlässt, wenn die Gewinnwahrscheinlichkeiten verbal ausgedrückt werden; denn dann ist ϵ auch im Wahl-Paradigma die wichtigere Dimension, so die Argumentation von Gonzalez-Vallejo und Wallsten (1992). Abb. 12 zeichnet diesen Gedankengang wiederum schematisch nach, die Übersetzung der Gewinnwahrscheinlichkeiten in Worte erfolgte dabei ad hoc durch mich. In beiden Paradigmen kommt jetzt der Wertdimension besonderes Gewicht zu, also wird in beiden Paradigmen einheitlich die €-Lotterie bevorzugt, das Phänomen der Präferenzumkehr bleibt aus, bzw. wird geringer.

Wie in der zuvor dargestellten Studie gliederten sich die Vpn auch bei der Untersuchung von Gonzalez-Vallejo und ihrem Koautor (1992) in Experten und Entscheider. Die Aufgabe der Experten bestand wiederum darin, die Chance für den positiven Ausgang eines kleinen, selbständig am Computer ablaufenden Videospiels einzuschätzen; sie taten dies sowohl durch eine Prozentangabe als auch durch die Auswahl eines Begriffes aus einer vorgegebenen Liste. Diese Angaben flossen in die Konstruktion von Lotterien ein. Paare von Lotterien mit gleichen oder ähnlichen Erwartungswerten wurden den Entscheidern vorgelegt, diese gaben an, welche sie lieber spielen wollten. Eine zweite Aufgabe bestand darin, für alle Lotterien den sogenannten *Minimum Selling Price* (MSP) festzu-

legen, das heißt die Vpn sollten angeben, was man ihnen mindestens bieten müsste, damit sie bereit wären, die jeweilige Lotterie zu verkaufen. Beide Prozeduren, Wählen und den MSP festlegen, wurden sowohl mit den Lotterien im numerischen wie auch mit denen im verbalen Format durchgeführt.

Verbale Darbietung der Wahrscheinlichkeit				
Paradigma	Lotterietyp	ϵ	P	bevorzugt wird
Wählen	P-Lotterie	<u>20,- DM</u>	„hohe Chance“	
	ϵ -Lotterie	<u>90,- DM</u>	„mäßige Chance“	***
Bieten	P-Lotterie	<u>20,- DM</u>	„hohe Chance“	
	ϵ -Lotterie	<u>90,- DM</u>	„mäßige Chance“	***

Abb. 12 Schematische Darstellung, warum, der Logik von Gonzalez-Vallejo et al. (1992) folgend, verbale Wahrscheinlichkeitsinformation zu einer Verminderung des Preference Reversals führen sollte; anders als bei Lotterien mit numerischer Wahrscheinlichkeitsinformation soll die Wertdimension hier auch im Wahlparadigma wichtiger als die Wahrscheinlichkeitsinformation sein.

Tatsächlich sank die Quote der Präferenzwechsel von knapp 25% im numerischen Modus auf knapp 20% im verbalen, der Unterschied erwies sich als statistisch signifikant. Gonzalez-Vallejo und Wallsten (1992) interpretierten dies als Bestätigung ihrer Vermutung, dass verbale Entscheidungsinformation Entscheider sich stärker an der Wertdimension der Optionen orientieren lässt. Auch dieser Befund lässt sich scheinbar im Sinne der Gewichtungshypothese interpretieren, wenn man nämlich die entscheidende Kraft für diese Umorientierung wiederum in der Ambiguität der verbalen Information vermutet.

5.3 Zusammenfassung

Zwei Studien sind scheinbar geeignet, die Gewichtungshypothese zu stützen. Beide vergleichen zwei verschiedene Modi der Information über Wahrscheinlichkeiten, verbale und numerische Angaben. In der Untersuchung von Gonzalez-Vallejo et al. (1994) zeigte sich dabei, dass im verbalen Modus die Attraktivitätsbewertungen von Lotterien höher mit deren Werten korrelierten als im numerischen. In der Studie von Gonzalez-Vallejo et al. (1992) hatte sich gezeigt, dass die Zahl der Preference Reversals nachließ, wenn Wahrscheinlichkeiten nicht numerisch sondern verbal dargeboten wurden. Beide Phänomene sind zu erwarten, wenn verbale Information Vpn veranlasst, die Attraktivität von Optionen stärker anhand ihrer Auszahlungsbeträge zu bewerten.

6. ... WARUM DIES UNZUREICHEND IST, ...

In Kapitel 6 wird dargelegt, warum die beiden zuvor vorgestellten Studien keine Rückschlüsse auf die Gültigkeit der Gewichtungshypothese bzw. des enger gefassten Postulats, dass sich Entscheider im Falle verbaler Wahrscheinlichkeitsinformation stärker an den Werten der gegebenen Optionen orientieren, zulassen.

Dass ich im vorangegangenen Kapitel oft davon gesprochen habe, dass die dort behandelten Untersuchungen die Gewichtungshypothese „scheinbar“ stützen, ist kein Zufall – sie können es nicht. Ehe ich dies für beide Studien zeigen kann, müssen wir uns der Frage zuwenden, wie sich die Bedeutung verbaler Ausdrücke von Wahrscheinlichkeit fassen lässt, und auf welche Weise eine solche Information in Entscheidungen einfließen kann. Dies soll im Folgenden geschehen.

Wie sich die Bedeutung verbaler Ausdrücke von Wahrscheinlichkeit fassen lässt

Das elaborierteste Modell, das die Bedeutung verbaler Wahrscheinlichkeitsausdrücke beschreibt, greift auf das Konzept der Zugehörigkeitsfunktion („membership function“) zurück, das aus der Fuzzy-Logic kommt (als Einführung z. B. Drösser, 1993). Die Bedeutung, die ein Wahrscheinlichkeitsausdruck für eine gegebene Person hat, wird dabei dadurch beschrieben, dass zu jeder Wahr-

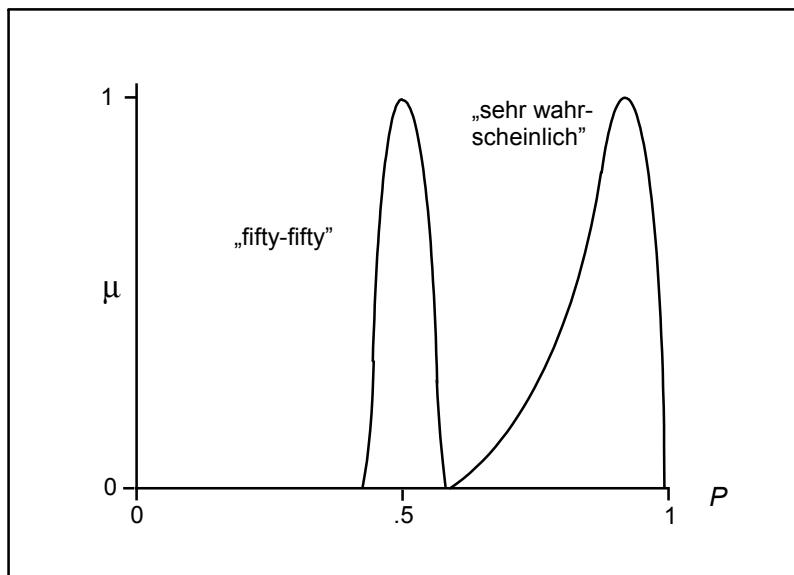


Abb. 13 Hypothetische Zugehörigkeitsfunktionen für zwei verbale Wahrscheinlichkeitsbeschreibungen.

scheinlichkeit angegeben wird, inwieweit diese mit der Bedeutung des Begriffs in Einklang steht. Dies geschieht in der Regel über die Zuordnung eines Wertes μ , der von null bis eins reichen kann. Null bedeutet dabei, dass die gegebene Wahrscheinlichkeit gar nicht durch den Begriff abgedeckt ist, eins zeigt eine maximale Deckung an. Abb. 13 zeigt zwei hypothetische Funktionen. Wallsten, Budescu, Rapoport, Zwick und Forsyth (1986) zeigten, dass das Konzept der Zugehörigkeitsfunktionen sinnvoll auf verbale Wahrscheinlichkeitsausdrücke anwendbar ist.

Sie haben solche Funktionen mittels einer Paarvergleichsmethode erhoben; dabei wurden jeweils zwei Wahrscheinlichkeiten in Form eines Glücksrades dargestellt und dazu ein Begriff, der eine Wahrscheinlichkeit beschreibt, präsentiert. Die Vpn sollten angeben, welche der beiden Wahrscheinlichkeiten der Begriff besser beschreibt und darüber hinaus skalieren, wieviel besser er dies tut. Die so gewonnenen Zugehörigkeitsfunktionen erwiesen sich als reliabel und bedeutsam, und die Ergebnisse ließen eine Interpretation der Funktionen auf Intervallskalenniveau zu.

Wie verbale Wahrscheinlichkeitsinformation in Entscheidungen einfließen kann

Wallsten, Budescu und Erev (1988) entwarfen ein von ihnen so getauftes $\nu\text{-}\mu$ Modell um zu erklären, wie eine verbal gegebene Wahrscheinlichkeitsinformation in eine Entscheidung unter Unsicherheit einfließen kann. Ausgangspunkt ist wiederum die Idee, dass sich die Bedeutung des Begriffs in Form einer Zugehörigkeitsfunktion repräsentieren lässt. Für die Entscheidungsfindung wird nun zufällig eine Wahrscheinlichkeit gezogen; die Ziehungswahrscheinlichkeit ist dabei proportional zu μ . Gezogen wird jedoch nicht aus dem gesamten Wahrscheinlichkeitsspektrum sondern lediglich aus dem Bereich, bei dem μ einen bestimmten Schwellenwert ν , der individuell verschieden sein kann, überschreitet (vgl. Abb. 14). Wallsten et al. (1988) überprüften dieses Modell, indem sie Vpn zunächst Wahlen zwischen jeweils zwei Lotterien treffen ließen. Bei einer war die Gewinnwahrscheinlichkeit jeweils graphisch dargestellt und bei der anderen in Worten. Im Anschluss daran wurden die Zugehörigkeitsfunktionen der verwendeten Wahrscheinlichkeitsausdrücke erhoben. Die gefundenen Beziehungen zwischen den Wahlen und den Zugehörigkeitsfunktionen standen in Einklang mit dem $\nu\text{-}\mu$ Modell.

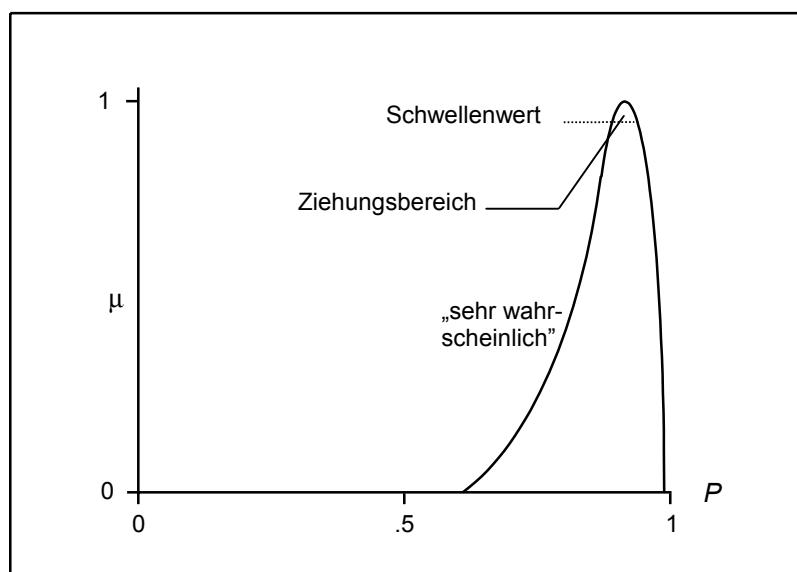


Abb. 14 Illustration des $\nu\text{-}\mu$ Modells: Der verbalen Umschreibung „sehr wahrscheinlich“ wird dergestalt ein numerischer Wert zugeordnet, dass aus der Verteilung oberhalb des Kriteriums ν ein Wert gezogen wird.

Bedeutungsunterschiede verbaler Wahrscheinlichkeitsinformation zwischen Sendern und Empfängern

Fillenbaum, Wallsten, Cohen und Cox (1991) untersuchten, inwieweit Aufgabencharakteristika die Form solcher Zugehörigkeitsfunktionen beeinflussen. Zu diesem Zweck hatten deren Vpn unter anderem eine Selektions- und eine Evaluationssaufgabe zu absolvieren. Bei ersterer wurden die Vpn in die Lage eines Senders versetzt, der einer anderen Person eine gegebene Wahrscheinlichkeit in Worten übermitteln soll. Dies geschah, indem für jeden der insgesamt 25 zu kommunizierenden Wahrscheinlichkeitswerte aus einer Liste mit vorgegebenen Formulierungen die am besten passende herauszusuchen war. Anschließend gaben die Vpn auf einer Skala an, wie gut der ausgewählte Begriff die Wahrscheinlichkeit beschreibt. Danach konnten sie eine zweitbeste Formulierung auswählen, deren Güte dann ebenfalls einzuschätzen war; diesen Zyklus durchliefen die Vpn so lange, bis sie für die gegebene Wahrscheinlichkeit keine weitere Formulierung mehr auswählen wollten. Die Evaluationsaufgabe versetzte die Teilnehmer in die Rolle eines Empfängers. Sie sollten sich vorstellen, dass ein Freund ihnen die Gewinnwahrscheinlichkeit verschiedener Lotterien in Worten beschrieben hätte. Sie könnten nun diese Formulierung zusammen mit der tatsächlichen Wahrscheinlichkeit sehen und sollten die Angemessenheit dieser Formulierung einschätzen. Dabei wurden dieselben Wahrscheinlichkeitsbeschreibungen und Wahrscheinlichkeiten wie in der Selektionsaufgabe benutzt. Auf diese Weise ließen sich für die selben Formulierungen Zugehörigkeitsfunktionen aus der Perspektive des Senders und des Empfängers erstellen. Dabei zeigte sich, dass diese aus der Empfängerperspektive breiter und vor allem zur Mitte des Wahrscheinlichkeitspektrums hin verschoben waren.

Da sich somit bei den Empfängern die Gipfel der μ -Verteilungen zur Mitte hin verschoben haben, kann man erwarten, dass die Übersetzung von Wahrscheinlichkeiten in Sprache und die anschließende Rückübersetzung dieser Formulierungen in Zahlen durch einen Empfänger zu einer Regression dieser rückübersetzten Werte zur Mitte führt. Die Verteilung der von den Sendern wahrgenommenen Wahrscheinlichkeitswerte wird also im Allgemeinen schmäler und spitzer sein als die Verteilung der gesendeten Werte. Hiervon ausgehend wird schnell klar werden, warum die Studien von Gonzalez-Vallejo, Erev und Wallsten (1994) sowie Gonzalez-Vallejo und Wallsten (1992) keine Schlüsse auf die Gültigkeit der Gewichtungshypothese zulassen.

6.1 Warum die Studie von Gonzalez-Vallejo, Erev und Wallsten (1994) nichts über die Gewichtungshypothese aussagt

In der Studie von Gonzalez-Vallejo et al. (1994) nahmen die Entscheider die Wertinformation in beiden Versuchsbedingungen (verbale vs. numerische Wahrscheinlichkeitspräsentation) gleich wahr. Die Gewinnhöhen der einzelnen Lotterien wurden unter beiden Bedingungen in gleicher Weise unverfälscht wahrgenommen. Die Wahrnehmung der Gewinnwahrscheinlichkeiten war jedoch abhängig von der Versuchsbedingung. Während die Entscheider das in numerischer Form gesendete Signal ohne Informationsverlust empfingen, führte die verbale Übermittlung aus den soeben genannten Gründen wohl zu einem Signal, das in Richtung .5 verschoben war. Während die Variabi-

lität des wahrgenommenen Signals ϵ also unter beiden Bedingungen gleich groß war, war die Streuung der von den Entscheidern wahrgenommenen Wahrscheinlichkeit P unter der Bedingung numerischer Information größer, weil hier das Verständnis des Signals nicht zu einer Regression zur Mitte führte. Die Variabilitätseinschränkung des P -Signals in der verbalen Bedingung führte dann dazu, dass die Erwartungswerte, die die Entscheider aus den ihnen verfügbaren Informationen für die einzelnen Lotterien abschätzen konnten, bei numerischer Information weniger stark von ϵ abhingen als bei verbaler Information.

Dies lässt sich anhand eines Beispiels leicht veranschaulichen. Betrachten wir dazu zwei Spielesbündel, eines mit den Lotterien A, B, C und das zweite mit den Spielen A', B', C' , wobei in jedem dieser sechs Spiele der Betrag ϵ mit der Wahrscheinlichkeit P gewonnen wird oder andernfalls nichts:

Spiel	P	ϵ	EV	Korrelationen
A	.2	50	10	$V(EV, P) = 2/3$
B	.5	80	40	$V(EV, \epsilon) = 2/3$
C	.8	20	16	
A'	.4	50	20	$V(EV, P) = 1/3$
B'	.5	80	40	$V(EV, \epsilon) = 1$
C'	.6	20	12	

Im ersten Bündel ist die Variabilität von P und ϵ gleich groß, $V(EV, P)$ wie auch $V(EV, \epsilon)$ betragen .67. Das zweite Spielesbündel ist mit dem ersten identisch, bis auf die Tatsache dass hier P einer Regression zur Mitte unterliegt und die Variabilität der Wahrscheinlichkeitsdimension somit verringert ist. In Folge dessen steigt der Zusammenhang der Erwartungswerte mit ϵ an, während $V(EV, P)$ auf .33 absinkt. Wenn Entscheidungsfinder innerhalb eines jeden Spielesbündels eine Attraktivitätsrangfolge C der Lotterien herstellen, so ist natürlich damit zu rechnen, dass diese im zweiten Bündel stärker mit ϵ korrelieren als dies im ersten der Fall ist. Ein solcher Unterschied in $V(C, \epsilon)$ gibt dann aber keinerlei Hinweis auf unterschiedliche Entscheidungsstrategien in beiden Bündeln, sondern ergibt sich ist allein als Konsequenz von deren unterschiedlicher mathematischer Struktur.

Somit war in der Untersuchung von Gonzalez-Vallejo et al. (1994) aus rein mathematischen Gründen *auf jeden Fall* zu erwarten, dass die Rangordnung C , in die die Vpn eine Serie von Lotterien brachten, bei verbaler Information stärker mit der Rangordnung, die sich aus dem Wert der Lotterien ergab, zusammenhing, als dies bei numerischer Information der Fall war. Abb. 15 zeichnet den Gedankengang, der zu diesem Resultat führt, noch einmal schematisch nach. Wenn also Gonzalez-Vallejo und Mitarbeiter, wie geschehen, unter der Bedingung verbaler Information eine höhere Korrelation $V(C, \epsilon)$ fanden als unter der numerischen Bedingung, so ist völlig unklar, ob dies lediglich ein Effekt der Regression zur Mitte war, oder ob sich hier *zusätzlich* auch ein Strategiewchsel, wie ihn die Gewichtungshypothese postuliert, gezeigt hat.

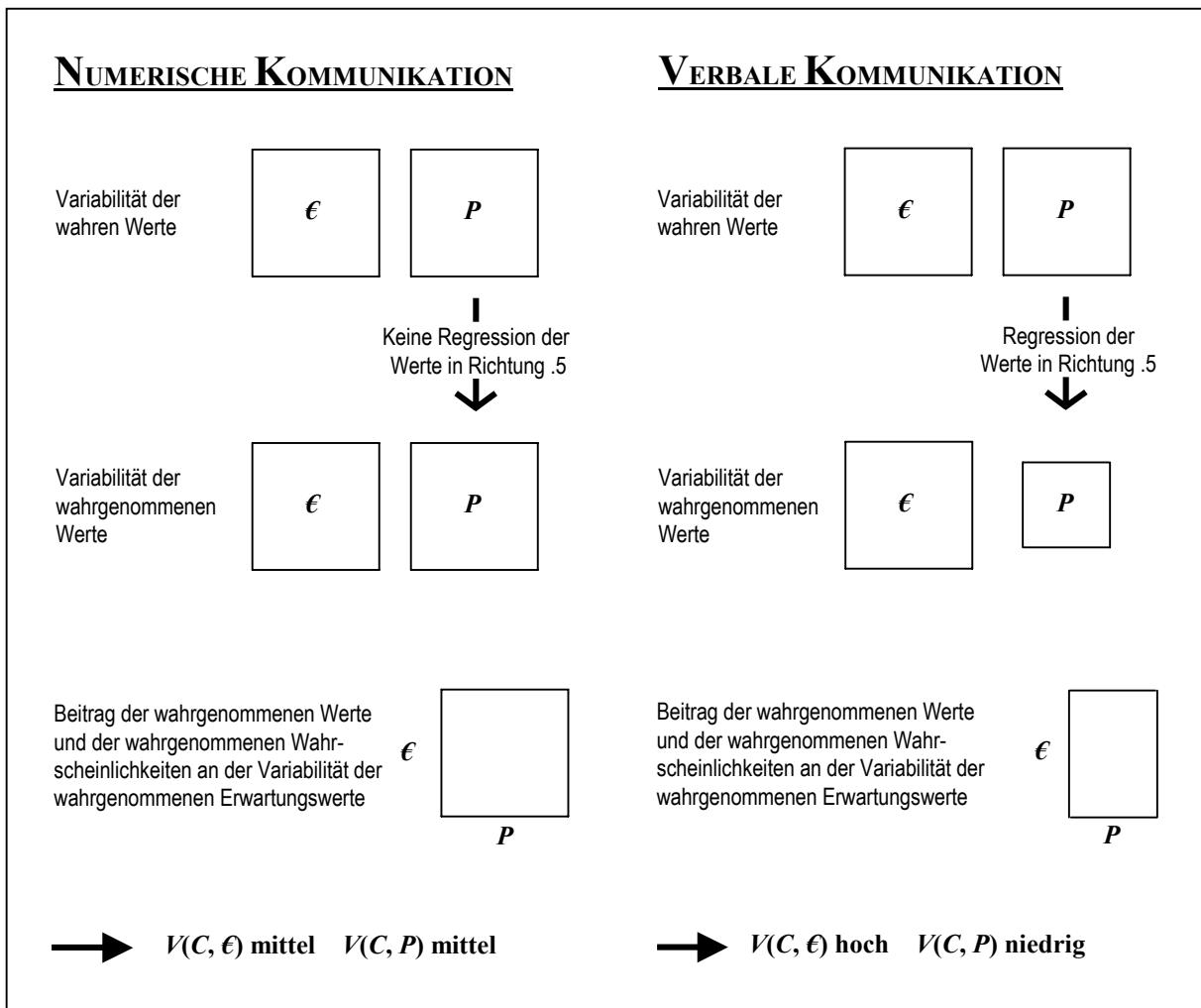


Abb. 15 Paradigma von Gonzalez-Vallejo et al. (1994): Bei verbaler Wahrscheinlichkeitsinformation haben die vom Entscheider wahrgenommenen Wahrscheinlichkeiten eine geringere Variabilität als im numerischen Modus. Im verbalen Modus sind daher die wahrgenommenen Erwartungswerte der Lotterien stärker von deren Werten abhängig. Dadurch erhöht sich hier $V(C, V)$ auch dann, wenn kein Strategiewechsel im Sinne der Gewichtungshypothese vollzogen wird.

Auch der zweite Befund aus dieser Studie erlaubt keinen Rückschluss auf die Gewichtungshypothese. Dieser bestand ja in einer Überlegenheit numerischer Information in Spielesäcken, bei denen der Erwartungswert stark mit P korrelierte, während Entscheidungen auf Grundlage verbaler Information zu besseren Ergebnissen führten, wenn der Erwartungswert hoch mit ϵ korrelierte (vgl. Abb. 10). Ein Absinken des Graphen in Abb. 10 ist aber auch dann zu erwarten, wenn bei verbaler Wahrscheinlichkeitsinformation genau so entschieden wird wie bei numerischer: Wie bereits erläutert, ist das verbal gegebene Signal mit einem Fehler behaftet, der dem numerischen fehlt, und der daher führt, dass verbale Beschreibungen von Wahrscheinlichkeiten von verschiedenen Menschen unterschiedlich verstanden werden. Dass dieser Fehler stärker zum Tragen kommt, wenn die Erwartungswerte der Lotterien vor allem von der Wahrscheinlichkeitsinformation abhängen, ist selbstverständlich. Somit muss die numerische Information insbesondere dann überlegen sein, wenn die Erwartungswerte stark von P abhängen.

6.2 Warum die Studie von Gonzalez-Vallejo und Wallsten (1992) nichts über die Gewichtungshypothese aussagt

Auch das Ergebnis von Gonzalez-Vallejo et al. (1992), dass bei verbaler Wahrscheinlichkeitsinformation weniger Präferenzwechsel zu beobachten waren als bei numerischer, ist völlig unabhängig davon, ob die Vpn ihre Strategie wechselten oder nicht, zu erwarten. Die Ursache hierfür liegt wiederum darin, dass die Wahrnehmung einer in Worten ausgedrückten Gewinnchance der Regression zur Mitte unterliegt (vgl. S. 45). Als Folge davon muss man davon ausgehen, dass in dem Versuch bei verbaler Wahrscheinlichkeitsinformation die wahrgenommene Wahrscheinlichkeit der P-Lotterie geringer ist als die wahre Wahrscheinlichkeit; umgekehrt ist davon auszugehen, dass unter dieser Bedingung die €-Lotterie attraktiver wurde, weil hier die wahrgenommene Gewinnchance höher war als die tatsächliche. Die Regression zur Mitte führte also dazu, dass der wahrgenommene Erwartungswert der P-Lotterie sank, während derjenige der €-Lotterie stieg. Da Erwartungswert und Attraktivität solcher Lotterien eng miteinander verknüpft sind, stieg somit natürlich auch die Chance, dass die €-Lotterie auch im Wahl-Paradigma bevorzugt wurde, völlig unabhängig davon, ob der Wertdimension der Lotterien nun ein höheres Gewicht zugemessen wurde oder nicht. Abb. 16 stellt diesen Gedankengang noch einmal dar, wobei die Dekodierung der Wahrscheinlichkeitsbeschreibungen ad hoc durch mich erfolgte.

6.3 Zusammenfassung

Verbale Wahrscheinlichkeitsbeschreibungen werden von Empfängern systematisch anders verstanden als von Sendern. Und zwar in einer Weise, die erwarten lässt, dass eine Rückübersetzung der verbalen Ausdrücke in Zahlen dazu führt, dass diese, verglichen mit den Ausgangswerten, einer Regression zur Mitte unterliegen. Aufgrund dieser Tatsache sind die Ergebnisse, die Gonzalez-Vallejo et al. (1994) und Gonzalez-Vallejo et al. (1992) erzielt haben, völlig unabhängig von einem Strategiewechsel, wie ihn die Gewichtungshypothese postuliert, zu erwarten. Ob die Ergebnisse ausschließlich durch die Regression zur Mitte zustande gekommen sind, oder ob auch zusätzlich eine Höhergewichtung der Wertinformation bei verbaler Wahrscheinlichkeitsinformation dazu beigetragen hat, ist nicht zu ermitteln.

Verbale Darbietung der Wahrscheinlichkeit					
Paradigma	Lotterietyp	ϵ	P	subjektiver Erwartungswert	bevorzugt wird
Wählen	P-Lotterie	Objektiver Wert: 20,- DM	Objektiver Wert: .90	16,- DM	???
		Gesagt wird: 20,- DM	Gesagt wird: „hohe Chance“		
		Verstanden wird: 20,- DM	Verstanden wird: .80		
	€-Lotterie	Objektiver Wert: 90,- DM	Objektiver Wert: .20	27,- DM	???
		Gesagt wird: 90,- DM	Gesagt wird: „mäßige Chance“		
		Verstanden wird: 90,- DM	Verstanden wird: .30		
Bieten	P-Lotterie	Objektiver Wert: 20,- DM	Objektiver Wert: .90	16,- DM	
		Gesagt wird: 20,- DM	Gesagt wird: „hohe Chance“		
		Verstanden wird: 20,- DM	Verstanden wird: .80		
	€-Lotterie	Objektiver Wert: 90,- DM	Objektiver Wert: .20	27,- DM	***
		Gesagt wird: 90,- DM	Gesagt wird: „mäßige Chance“		
		Verstanden wird: 90,- DM	Verstanden wird: .30		

Abb. 16 Untersuchung von Gonzalez-Vallejo et al. (1992): Wird die Wahrscheinlichkeitsinformation, wie hier dargestellt, verbal gegeben, so führt dies dazu, dass der wahrgenommene Erwartungswert der €-Lotterie steigt. Daher sollte im Wahl-Paradigma die P-Lotterie nicht mehr unbedingt bevorzugt werden. Somit verringert sich die Quote der Präferenzumkehrungen auch dann, wenn verbale Wahrscheinlichkeitsinformation nicht dazu führt, dass der Wertdimension bei der Entscheidung ein höheres Gewicht zukommt.

7. ... UND WIE MAN ES BESSER MACHEN KANN, ...

In Kapitel 7 wird das grundlegende Paradigma dieser Arbeit vorgestellt. Aufgrund der gerade beschriebenen Schwächen bisheriger Studien ist es notwendig, für die Überprüfung der Gewichtungshypothese ein neue Auswertungsstrategie zu entwickeln. Die Analyse des Exponenten θ aus dem Contingent Weighting Modell erweist sich dabei als geeignete Alternative. Wie dieser aus den Antworten von Vpn geschätzt werden kann wird exemplarisch aufgezeigt.

Ziel der vorliegenden Arbeit ist zu überprüfen, ob zunehmende Ambiguität dazu führt, dass Entscheidungen sich, im Sinne eines Strategiewechsels, zunehmend auf die gegebene Wertinformation stützen (Gewichtungshypothese). Um den Grad der Ambiguität kontrolliert variieren zu können, erschien es mir zunächst nicht ratsam, diese dadurch herzustellen, dass wie in den gerade beschriebenen Studien verbale Beschreibungen der Wahrscheinlichkeiten zum Einsatz kommen. Wie Ambiguität statt dessen hergestellt wurde, wird im nächsten Teil beschrieben, in dem das grundlegende Paradigma dieser Arbeit dargestellt wird. Im darauf folgenden Teil wird dann erläutert, wie sich die Gültigkeit der Gewichtungshypothese mittels der so zu gewinnenden Daten überprüfen lässt.

7.1 Das grundlegende Paradigma der vorliegenden Arbeit

Das zur Erhellung dieser Fragestellung verwendete Paradigma ist dem der oben beschriebenen Studie von Gonzalez-Vallejo et al. (1994) sehr ähnlich: Die von den Vpn zu bearbeitende computergesteuerte experimentelle Aufgabe bestand wiederum darin, Gruppen von jeweils sechs Lotterien gemäß ihrer Attraktivität in eine Rangreihe zu bringen. Jede Lotterie bestand aus einem Punktewert ϵ und einer Gewinnwahrscheinlichkeit P . Ziel war, möglichst viele Punkte zu erreichen. Die Auszahlungsregel war dabei die bereits beschriebene: Im Gewinnfall erbrachte eine Lotterie $\epsilon \cdot \text{Rangplatz}$ Punkte, andernfalls nichts. Im Sinne der Maximierung des Erwartungswertes oder einer verwandten Größe war es also notwendig, der attraktivsten Lotterie Rang sechs zu geben, der zweitbesten Rang fünf usw.

Die Darstellung der Gewinnwahrscheinlichkeit erfolgte über ein Glücksrad, das aus einem gelben und einem weißen Sektor bestand, wobei die Gewinnchance dem Größenanteil der gelben Fläche entsprach. Eine solche Darstellung von Wahrscheinlichkeit mittels eines Glücksrades ist nicht ungewöhnlich. Befunde von Budescu und Weiss (1987) zeigen, dass eine solche Darstellungsform einer numerischen im Wesentlichen äquivalent ist.

Ambiguität wurde im hier verwendeten Paradigma dadurch hergestellt, dass über einen Teil des Glücksrades eine Verdeckung gelegt wurde. Deren Platzierung erfolgte zufällig, aber stets so, dass genau eine der beiden Sektorengrenzen abgedeckt war. Dadurch konnte die Größe der Gewinnfläche nicht mehr zuverlässig abgeschätzt werden. Die Stärke der Ambiguität ließ sich nun stufenlos über die Größe der Verdeckung festlegen, und sie konnte präzise in Prozentwerten angegeben werden. Abb. 17 zeigt exemplarisch eine Lotterie mit einer Ambiguität von 25%.

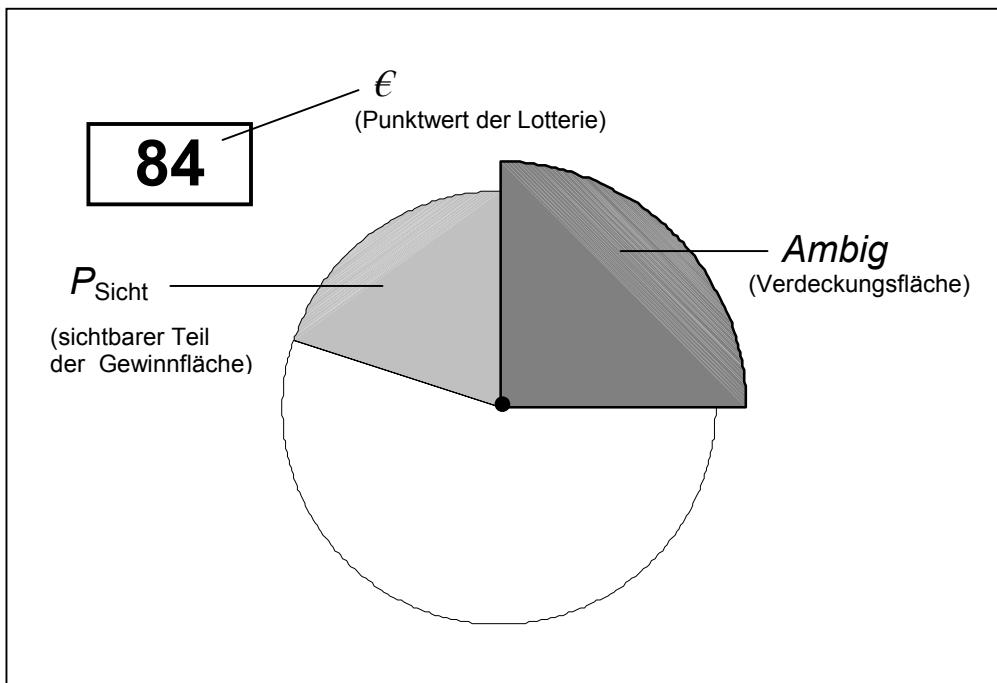


Abb. 17 Lotterie mit 25% Ambiguität. Im Gewinnfall erbringt diese Lotterie 84 Punkte mal dem ihr zugewiesenen Rangplatz. Die Gewinnwahrscheinlichkeit ist gleich der Größe der hellgrauen Fläche, von der jedoch nur ein Teil (P_{Sicht}) zu sehen ist; der Rest unbekannter Größe wird von der dunkelgrauen Verdeckungsfläche überlagert.

Der Spielraum für die Platzierung der Verdeckungsfläche ist in Abhängigkeit von P und dem Grad der Ambiguität unterschiedlich groß: Beträgt letztere beispielsweise 25% und $P = .5$, so kann P_{Sicht} zwischen .26 und .49 liegen, der Spielraum für die Platzierung der Verdeckungsfläche ist also hoch. Beträgt P jedoch nur .03, so kann P_{Sicht} nur die Werte .01 oder .02 annehmen, es besteht also nur ein minimaler Spielraum für die Platzierung der Verdeckungsfläche. Restriktionen dieser Art führen dazu, dass bei der Betrachtung einer ambigen Lotterie der Erwartungswert für P nicht gleich $P_{\text{Sicht}} + \text{Ambig}/2$ ist, wie man intuitiv leicht vermutet. Der Schätzwert $P_{\text{Sicht}} + \text{Ambig}/2$ unterliegt nämlich einer Regression in Richtung .5. Darauf, wie sich die Erwartungswerte für P verhalten, wird zu einem späteren Zeitpunkt ausführlich eingegangen (13.2.1 bzw. Abb. A1 im Appendix).

Die Randbedingungen der Simulation I (Kap. 4.3) sind identisch mit denen des hier vorgestellten Paradigmas. Die Verdeckungsfläche ist dabei das Äquivalent zu dem dort verwendeten ambiguitätsformenden Intervall $[P_{\text{lower}}, P_{\text{upper}}]$.

7.2 Simulation II: Ist in diesem Paradigma $V(C, \epsilon)$ eine geeignete abhängige Variable?

Mittels einer zweiten Simulation soll überprüft werden, ob das von Gonzalez-Vallejo, Erev & Wallsten (1994) verwendete Korrelationsmaß $V(C, \epsilon)$ in meinem Paradigma, in dem Ambiguität nicht verbal sondern auf die beschriebene Weise erzeugt wird, möglicher Weise doch geeignet ist, Licht auf die Gewichtungshypothese zu werfen. In der Simulation II werden dazu die Aufgaben auf dieselbe Weise generiert und dargeboten, wie im experimentellen Paradigma, bzw. wie in Simula-

tion I. Diesmal schließt die Simulation auch zwei virtuelle Entscheider ein, die die Lotterien nach deren wahrgenommener Attraktivität in eine Rangordnung bringen. Der erste Entscheider verfolgt dabei konsequent eine Strategie, die *Normal* heißen soll. Die Attraktivität einer Lotterie ist dabei identisch mit dem vermuteten Erwartungswert der Lotterie und ergibt sich als

$$\text{Attraktivität} = \epsilon \cdot (P_{\text{Sicht}} + Ambig/2).$$

Dabei entspricht P_{Sicht} dem sichtbaren Teil der Gewinnfläche und $Ambig$ der Größe der Verdeckung, d. h. dem Grad der Ambiguität (vgl. Abb. 17). Es wird hier also davon ausgegangen, dass genau unter der Hälfte der Verdeckungsfläche Gewinnfläche liegt.

Ein zweiter Entscheider beurteilt die Attraktivität durchgehend anhand einer anderen Strategie, die *Konservativ* heißen soll. Für sie gilt

$$\text{Attraktivität} = \epsilon \cdot P_{\text{Sicht}}.$$

Eine Überlegung, die zu einer solchen Strategie führt, wäre: „Sicher bin ich mir nur der Gewinnfläche, die ich sehen kann. Ich verlasse mich lieber nicht darauf, dass unter der Verdeckungsfläche auch noch etwas ist und stütze mich nur auf das, was ich sehen kann. Falls noch etwas dazu kommt – schön. Aber darüber könnte ich eh nur spekulieren, also beachte ich es gar nicht erst.“.

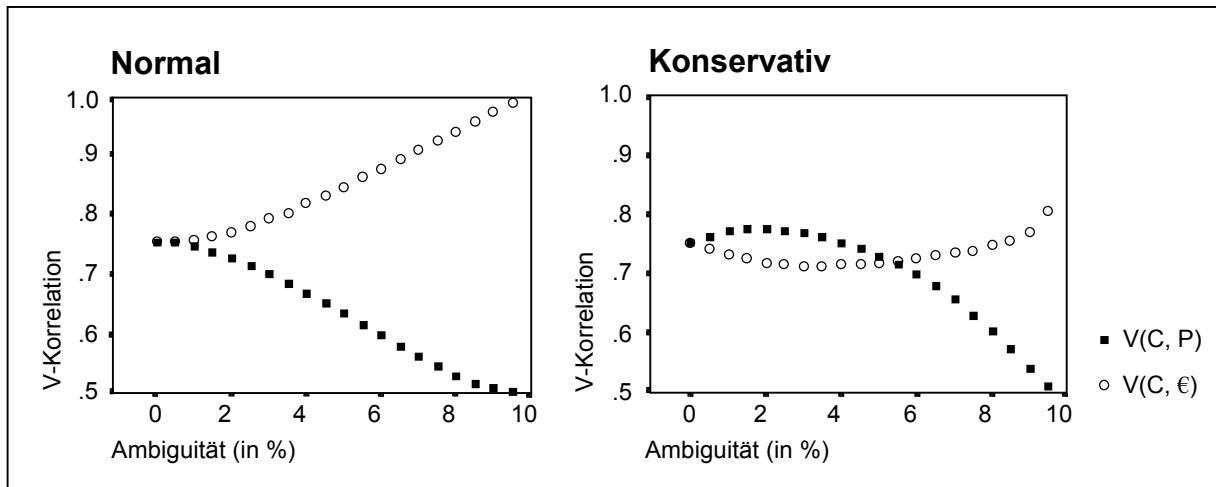


Abb. 18 Analyse der Strategien *Normal* und *Konservativ* im Paradigma der vorliegenden Arbeit: Bei zunehmender Ambiguität ändern sich die Korrelationsstrukturen, auch ohne dass die Entscheidungsstrategie sich wandelt. Daher kann $V(C, \epsilon)$ auch hier nicht verwendet werden, um die Gewichtungshypothese zu überprüfen. – Die Daten resultieren aus einer Monte-Carlo Studie, jedem Datenpunkt liegen 50 000 Aufgaben zu Grunde.

Abbildung 18 zeigt für beide Strategien getrennt, wie sich mit zunehmender Ambiguität $V(C, \epsilon)$ und $V(C, P)$ verändern. Obwohl in der Simulation zunehmende Ambiguität keinen Einfluss auf die Strategie der Entscheider ausübt (sie verfahren ja immer nach der selben Regel), ändern sich beide Maße erheblich. Zudem führen zwei Strategien, die a priori beide als gleichermaßen plausibel angesehen werden können, zu völlig unterschiedlichen Verläufen der abhängigen Variablen. Auch im von mir verwendeten Paradigma erlaubt also die Betrachtung, wie stark die Rangordnung, in die eine

Vp die Lotterien bringt, mit den Werten oder Wahrscheinlichkeiten dieser Lotterien zusammenhängt, keinen Aufschluss darüber, wieviel Gewicht dieser Entscheider dabei den beiden Dimensionen zugemessen hat.

7.2.1 Analyse zweier Strategien

Allen Interessierten soll im Folgenden nahe gebracht werden, warum die doch sehr ähnlichen Strategien *Normal* und *Konservativ* zu so unterschiedlichen Korrelationsstrukturen führen. Dem Verständnis der weiteren Arbeit schadet es jedoch nicht, diesen Abschnitt zu überspringen.

Dass bei *Konservativ* die Korrelationen der Wahlen mit P stets größer sind als bei *Normal*, ist relativ einfach zu erklären. Die Erwartungswertschätzung des zweiten Algorithmus, $EV = \epsilon \cdot (P_{\text{Sicht}} + Ambig/2)$, lässt sich umschreiben als $EV = \epsilon \cdot P_{\text{Sicht}} + \epsilon \cdot Ambig/2$. Durch den zweiten Summanden, der der ersten Strategie fehlt, geht die Variabilität von ϵ verstärkt in die Schätzung des Erwartungswertes ein. Daraus folgt, dass die Wahlen sich insgesamt stärker an ϵ orientieren und somit automatisch auch weniger stark an P .

Überraschend ist, dass die konservative Strategie bei steigender Ambiguität zunächst dazu führt, dass die Wahlen *stärker* mit P korrelieren, obwohl der Wertebereich von P_{Sicht} mit zunehmender Ambiguität immer geringer wird und somit die Varianz von P_{Sicht} immer stärker schrumpft. Bei fehlender Ambiguität beträgt die Standardabweichung von P_{Sicht} 27,2, bei 20% Ambiguität – hier ist die Korrelation der Wahlen mit P am stärksten (vgl. Abb. 18) – 25,7 und bei 40% Ambiguität – hier ist $V(C, P)$ wieder auf das Ausgangslevel gefallen – nur noch 19,7. Eine umfassende Einsicht in die Mathematik dieses Phänomens ist mir als einfacher Psychologen nicht zuteil geworden, eine ansatzweise Erklärung ist jedoch die folgende: Während P_{Sicht} bei fehlender Ambiguität gleichverteilt ist, zeigt das Histogramm einen „badewannenförmigen“ Verlauf, sobald diese ins Spiel kommt. D. h. in zwei kleinen Bereichen nahe den Enden der Verteilung häufen sich die Werte stark, um ansonsten gleichmäßig verteilt zu sein. Dies führt dazu, dass die Chance steigt, dass sich in einem Spielebündel Lotterien befinden, die sich hinsichtlich P_{Sicht} in extremer Weise unterscheiden. In diesem Falle gehen die Unterschiede in den Erwartungswertschätzungen in hohem Maße auf die Unterschiede in P_{Sicht} zurück. Da P_{Sicht} und P bei mäßiger Ambiguität hoch korreliert sind, fällt $V(C, P)$ hoch aus.

Die Standardabweichung ist somit nur äußerst bedingt geeignet, die Variabilität in ϵ oder P_{Sicht} zu messen, da sie die hier entscheidende Verhältnisskaleninformation nicht berücksichtigt. Dies wird klar, wenn wir in einem Gedankenexperiment die Werte für ϵ einmal aus dem Bereich 3 bis 97 ziehen und das andere Mal aus dem Bereich 100 003 bis 100 097. Die Standardabweichungen beider Verteilungen sind gleich. Im ersten Fall können sich die Werte zweier Lotterien jedoch um das bis zu 32,3-fache unterscheiden, im zweiten Fall lediglich um den Faktor 1,001. Trotz gleicher Standardabweichungen der Verteilungen werden sich im ersten Fall die ϵ -Werte verschiedener Lotterien oft deutlich unterscheiden, im zweiten Fall hingegen nie. Dem entsprechend werden die Wahlen im ersten Fall hoch mit ϵ korrelieren, im zweiten so gut wie gar nicht. Ein besseres Maß als die Standardabweichung sollte der Erwartungswert für den Quotienten zweier zufällig gezogener Werte sein (größerer Wert / kleinerer Wert).

Abb. 19 zeigt, wie sich die Variabilität der subjektiven Wahrscheinlichkeiten bei steigender Ambiguität für *Normal* und *Konservativ* entwickelt (für *Normal* ergibt diese sich als $P_{\text{Sicht}} + \text{Ambig}/2$, für *Konservativ* als P_{Sicht}); gemessen wird die Variabilität dabei als der gerade beschriebene Quotient. Während dessen Verlauf den Verlauf von $V(C, P)$ bei der normalen Strategie gut abbildet, ist dies in Bezug auf die konservative Strategie nur bedingt der Fall (vgl. Abb. 19 mit Abb. 18). Im Falle von *Konservativ* steigt die Kurve von $V(C, P)$ zwar ebenso wie diejenige des Erwartungswertes zunächst an, bis sie bei 20% Ambiguität ein Maximum erreicht, danach ist die „Sinkgeschwindigkeit“ jedoch sehr verschieden. Während $V(C, P)$ den Ausgangswert bereits bei 40% Ambiguität wieder erreicht, sinkt der Erwartungswert für den Quotienten zweier zufällig erzeugter Wahrscheinlichkeiten P_{Sicht} erst bei einer Ambiguität von 80% wieder auf das Ausgangsniveau zurück.

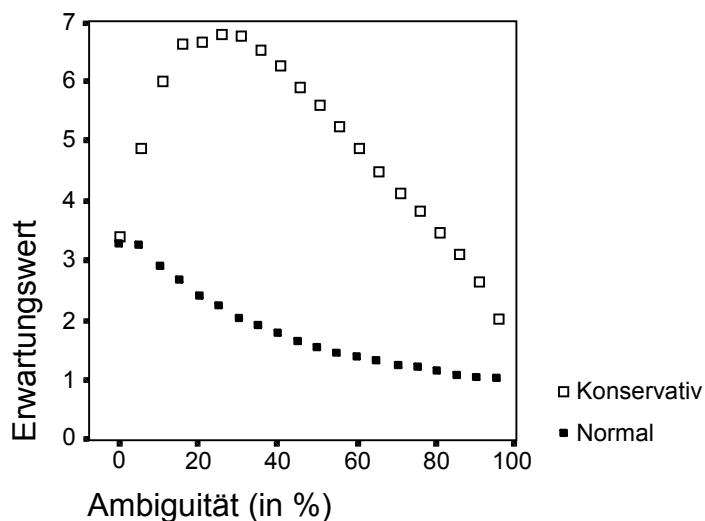


Abb. 19 Die Variabilität der subjektiven Wahrscheinlichkeiten, die sich für *Normal* und *Konservativ* ergeben, gemessen als Quotient zweier zufällig gezogener Werte.
– Ergebnis einer Monte-Carlo Simulation; jeder Datenpunkt entspricht 50 000 Quotienten.

7.3 Eine geeignete abhängige Variable

Aus den vorangegangenen Abschnitten ergibt sich, dass sich aus den Korrelationsmaßen $V(C, P)$ und $V(C, \epsilon)$ nur schwer eine Aussage darüber ableiten lässt, ob sich die Art der Beurteilungen der Lotterien durch Ambiguität verändert oder nicht. Sinnvoller ist eine Analyse auf Grund des Contingent Weighting Modells (s. 3.2.2), da der Exponent θ genau die gesuchte Information enthält – nämlich die Antwort auf die Frage „Wie stark stützt sich die Bewertung der Lotterien auf die Dimensionen ϵ und P ?“. Prinzipiell ist es möglich, θ aus den Antworten der Vpn zu schätzen. Nehmen wir aus Gründen der Vereinfachung zunächst an, dass Vpn P mittels der *normalen* Strategie schätzen, also als $P_{\text{Sicht}} + \text{Ambig}/2$, und dass sie die Lotterien in der Reihenfolge ordnen, die sich aufgrund von $\epsilon \cdot (P_{\text{Sicht}} + \text{Ambig}/2)^\theta$ ergibt. Da die Rangfolge, in die eine Vp die Lotterien gebracht hat, ebenso bekannt ist wie ϵ , P_{Sicht} und $\text{Ambig}/2$, lässt sich θ nun bestimmen.

Ein einfaches Rechenbeispiel macht klar, wie dies geschieht. Betrachten wir dazu ein Bündel von drei Lotterien X , Y , Z ohne Ambiguität; in jeder Lotterie wird entweder ϵ mit der Chance P gewonnen, andernfalls geschieht nichts. Unter C ist angegeben, wie eine Entscheiderin, für die wir θ schätzen möchten, die drei Lotterien ordnet – Option Y wird als attraktivste empfunden, Spiel X als am wenigsten günstig:

	ϵ	P	C	$\epsilon \cdot P_{\text{Sicht}}^1$	$\epsilon \cdot P_{\text{Sicht}}^{1,1}$	$\epsilon \cdot P_{\text{Sicht}}^{1,2}$
X	40	.2	1	$40 \cdot .2 = 8$	$40 \cdot .17 = 6,8$	$40 \cdot .14 = 5,6$
Y	20	.4	3	$20 \cdot .4 = 8$	$20 \cdot .36 = 7,2$	$20 \cdot .33 = 6,6$
Z	8	.9	2	$8 \cdot .9 = 7,2$	$8 \cdot .89 = 7,1$	$8 \cdot .88 = 7,0$

Wir gehen davon aus, dass die Entscheiderin die drei Optionen gemäß $\epsilon \cdot (P_{\text{Sicht}} + \text{Ambig}/2)^\theta$ ordnet hat; da hier keine Ambiguität vorliegt verkürzt sich dieser Term zu $\epsilon \cdot P_{\text{Sicht}}^\theta$, außerdem ist P mit P_{Sicht} identisch. Wenn wir für θ probhalber den Wert 1 vermuten, so ergibt sich als wahrgenommene Attraktivität für die drei Spiele deren Erwartungswert; somit sollten X und Y attraktiver als Z sein (s. o.). Setzen wir für θ 1,1 ein, so ordnet sich die Attraktivität der drei Spiele so, wie es unsere Vp auch getan hat. Für θ -Werte von 1,2 und mehr würde sich die Attraktivitätsrangfolge in Z , Y , X ändern und somit das Verhalten unserer Entscheiderin nicht mehr abbilden. Der geeignete Schätzer für θ wäre hier somit 1,1.⁷

7.4 Zusammenfassung

Die vorliegende Arbeit griff auf das Paradigma von Gonzalez-Vallejo et al. (1994) zurück, in dem Lotterien gemäß ihrer Attraktivität zu ordnen sind. Dabei wurden die Gewinnwahrscheinlichkeiten graphisch, in Form von Glücksräder, angezeigt. Ambiguität wurde darüber hergestellt, dass ein Teil des Glücksrades abgedeckt war – infolge dessen war die Gewinnchance nicht mehr eindeutig bestimmbar. Über die Größe der Abdeckung ließ sich Ambiguität stufenlos variieren. Ob diese Art der Ambiguität mit anderen Arten vergleichbar ist, etwa derjenigen, die der verbalen Beschreibung von Wahrscheinlichkeiten entspringt, kann nur empirisch geklärt werden. Simulation II zeigt, dass auch im Paradigma der vorliegenden Arbeit das von Gonzalez-Vallejo und Mitautoren benutzte Korrelationsmaß $V(C, \epsilon)$ keine geeignete abhängige Variable ist. Statt dessen bietet sich der Exponent θ aus dem Contingent Weighting Modell an, denn dieser ist ein unmittelbares Maß dafür, wie stark sich die Attraktivitätsbewertung einer Lotterie auf deren Wert oder auf deren wahrgenommene Gewinnchance stützt.

⁷ Die in dieser Arbeit gewählte Modellierung bezieht sich allein auf den Vergleich von Alternativen mit jeweils zwei Ausgängen, wobei ein Ausgang stets darin besteht, dass nichts passiert. Sollen Wahlen zwischen Alternativen, die sich in der Zahl möglicher Ausgänge unterscheiden, beschrieben werden, so ist dies in der hier gewählten Modellierung zunächst nicht möglich. Im Appendix sind unter Punkt A1 die Gründe dafür dargelegt.

8. ... WOFÜR ALLERDINGS ZWEI PROBLEME ZU LÖSEN SIND.

In Kapitel 8 wird gezeigt, dass sich θ doch nicht so einfach aus den Antworten der Vpn schätzen lässt, wie gerade behauptet wurde. Zuvor gilt es nämlich noch zwei Probleme zu überwinden, was aufs Anmutigste geschieht.

θ zu schätzen ist jedoch nicht so einfach, wie ich es in der Beispielrechnung habe erscheinen lassen. Es tun sich nämlich zwei Probleme auf, die bisher ignoriert wurden. Zum einen ist zunächst nicht klar, wie die Vpn aufgrund der ambigen Wahrscheinlichkeitsinformation zu einer Schätzung für P kommen, die $P_{\text{Schätz}}$ heißen soll. Zwei plausible Modelle, *Konservativ* und *Normal*, waren bereits beispielhaft angeführt worden; auch andere Modelle sind denkbar. Des Weiteren mag es sinnvoll erscheinen, im Geiste der Prospect Theorie anzunehmen, dass die Vpn nicht den Erwartungswert $\epsilon \cdot P_{\text{Schätz}}$ maximieren sondern eine Größe $u(\epsilon) \cdot w(P_{\text{Schätz}})$, wobei die erste Größe der Nutzen von ϵ ist und die zweite ein Entscheidungsgewicht w , in das die geschätzte Gewinnwahrscheinlichkeit $P_{\text{Schätz}}$ überführt wird (vgl. 3.3). Dies würde es nötig machen, zusätzlich eine bestimmte Nutzenfunktion sowie eine Funktion zur Transformation von Gewinnwahrscheinlichkeiten in Entscheidungsgewichte zu spezifizieren. Zunächst fehlt also ein „Basismodell“, auf dessen Grundlage θ überhaupt bestimmt werden könnte.

Aber auch wenn ein solches Modell vorhanden ist, muss noch ein zweites Problem gelöst werden: θ soll natürlich so geschätzt werden, dass es in Verbindung mit einem konkreten Basismodell solche Rangfolgen produziert, die denen der Vpn möglichst ähnlich sind. Was ist jedoch das „richtige“ Kriterium für eine gute Replikation? Mindestens drei erscheinen hier plausibel. Zum einen könnte dies sein, dass die künstlich erzeugten Antworten zu Rangreihen führen, die in gleichem Maße mit der Wahrscheinlichkeits- oder der Wertkomponente der Lotterien korrelieren wie die von den Vpn gebildeten Rangreihen. Wahlweise sollte also $V(C, P)$ oder $V(C, \epsilon)$ bei simulierten und empirischen Daten identisch sein. Ein drittes Kriterium könnte sein, dass die Replikation der von der Vp vergebenen Rangplätze möglichst genau ist. In der Beispielrechnung am Ende von Kapitel 7 war dieses Problem nicht aufgetaucht, da sich das Verhalten der Entscheiderin hier perfekt wiedergeben ließ. In einem solchen Fall führt die Anwendung aller drei Kriterien zum selben Schätzergebnis für θ .

Welches der drei Maße das geeignete ist, wird im nächsten Abschnitt anhand einer weiteren Simulation geklärt. Im darauf folgenden Teil 8.2 wird dann noch einmal ausführlich auf das Problem der Basismodelle eingegangen.

8.1 Simulation III: Welches Kriterium ist am geeignetsten, um θ zu schätzen?

Eine Simulation kann Aufschluss darüber geben, ob die drei oben genannten Kriterien zur Schätzung von θ zu unterschiedlichen Resultaten führen, und, falls ja, welches dieser drei Kriterien am sinnvollsten ist. Simulation III geht dabei so vor, dass künstliche, vom Computer erzeugte „Vpn“ Aufgaben mit Hilfe eines bestimmten Basismodells bearbeiten, wobei sie bestimmte, festgelegte θ -Werte verwenden. Dabei arbeiten sie, wie normale Menschen auch, nicht vollkommen exakt

sondern mit gewissen Fehlern (sonst bestünde auch das Problem der Kriterienwahl nicht, alle drei würden zu identischen Ergebnissen führen). Auf der Grundlage der so produzierten Daten lässt sich dann überprüfen, mit Hilfe welches der drei Kriterien die von den „Vpn“ verwendeten θ -Werte am genauesten geschätzt werden können.

30 künstliche „Vpn“ bearbeiten auf vier Ambiguitätsstufen (0%, 25%, 55% und 85%) jeweils fünfzehn Aufgaben, die dem beschriebenen Paradigma entsprechen. Die Rangordnung der Lotterien erfolgt dabei gemäß der Größe des Terms

$$(\epsilon + Fehler\epsilon) \cdot (P_{Sicht} + Ambig / 2 + Fehler_p)^0.$$

$Fehler\epsilon$ und $Fehler_p$ sind dabei gleichverteilte Zufallsvariablen im Wertebereich -10 bis 10 ($Fehler\epsilon$) bzw. -1 bis $.1$ ($Fehler_p$); die Fehlerterme unterliegen der Einschränkung, dass sie nicht zu einem negativen Wert innerhalb der Klammern führen dürfen. Für jede „Vp“ und jedes Ambiguitätslevel wird zufällig ein Wert für θ bestimmt, nämlich aus den Bereichen 3,5 bis 4,5 (0% und 55% Ambiguität) und 0,3 bis 1,3 (25% und 85% Ambiguität). Die Festlegung der Wertebereiche von θ sowie deren Zuordnung zu den Ambiguitätsniveaus ist dabei willkürlich; es geht zunächst nur darum, ob sich (i) ein in den „empirischen“ Daten vorhandenes Muster von θ in den θ -Schätzwerten widerspiegelt, und (ii) darum, anhand welches der drei Schätzkriterien das Originalmuster am besten wiedergegeben wird.

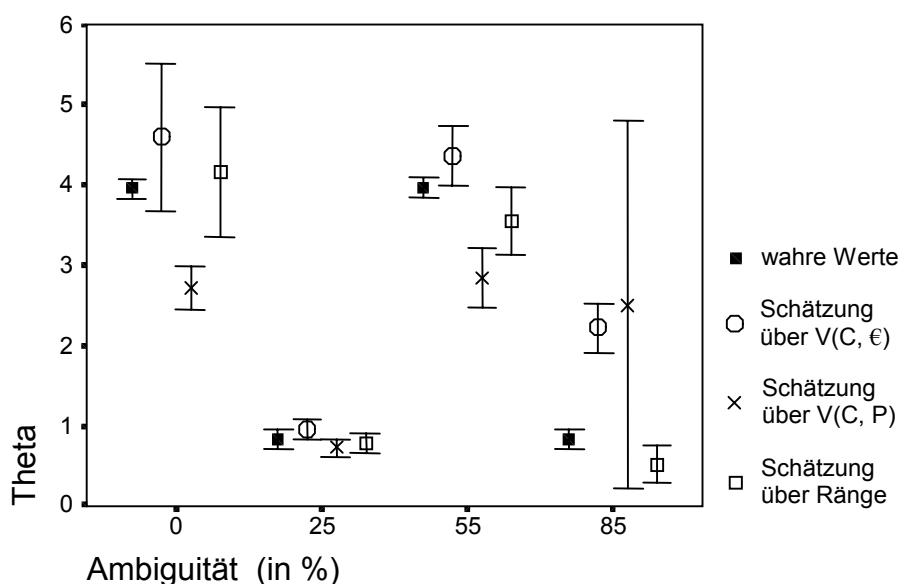


Abb. 20 Ergebnis der Simulation III: Die Schätzung von θ ist dann am genauesten, wenn die von den „Vpn“ vergebenen *Ränge* möglichst genau wiedergegeben werden. Die beiden anderen Schätzkriterien erwiesen sich als weniger tauglich. Die Winker markieren ± 2 Standardfehler.

Ein Vorteil dieser Simulation ist, dass sich das Problem des richtigen Basismodells nicht stellt. Wir wissen, dass die „Vpn“ ihre Bewertungen (bei Einfluss eines Zufallsfehlers) auf der Grundlage von *Normal* machen. Die Schätzungen für θ erfolgen daher ebenfalls auf Grundlage von *Normal*. Für

jede „Vp“ und jedes Ambiguitätsniveau wird θ dreimal geschätzt, wobei jeweils eines der drei oben beschriebenen Kriterien Verwendung findet. Der Fehler bei der Vorhersage eines Rangplatzes wird dabei als absolute Differenz zwischen prognostiziertem und tatsächlichem Rang definiert. Der Suchraum für den Schätzer reicht dabei stets von 0,1 bis 50 und wird in Schritten von 0,1 durchmessen. Oftmals findet sich, dass nicht nur ein einzelner sondern mehrere benachbarte Werte den Fehler zwischen dem vorgegebenen und dem nachvollzogenen Antwortverhalten minimieren. In diesem Fall wurden die Eckwerte dieses Intervalls sowie dessen Mitte einer weiteren Betrachtung zugeführt. Dabei erwies sich, dass der unterste Wert insgesamt den zuverlässigsten Schätzer darstellt. In Abb. 20, die die Ergebnisse zeigt, fließen daher nur diese Intervalluntergrenzen ein. Es zeigt sich klar, dass die Schätzung über Rangplätze die Ursprungswerte von θ am besten replizieren kann. Veränderungen in θ werden dabei trotz eines relativ hohen Zufallsfehlers, der in das Verhalten der „Vpn“ einfließt, recht genau abgebildet. Weitere Simulationen gleichen Typs mit anderen Fehlergrößen, θ -Wertebereichen und Antwortmodellen führen im Wesentlichen zu identischen Ergebnissen; aus Platzgründen sollen sie aber hier nicht näher dargestellt werden.

Das wesentliche Ergebnis von Simulation III lässt sich also so zusammenfassen: θ sollte so geschätzt werden, dass der Fehler, der sich bei der Vorhersage des Rangplatzes einer beliebigen Lotterie ergibt, minimiert wird. Falls nicht nur ein einzelner Wert dieses Kriterium erfüllt, ist es sinnvoll, den niedrigsten als Schätzer für θ zu verwenden.

Als Nebenresultat der vorgestellten Simulation ergibt sich, dass das Programm, welches θ schätzt, offensichtlich funktionstüchtig ist. Dies auf einem solchen Wege zu überprüfen ist insofern wünschenswert, als dass zum einen die entsprechende Software relativ komplex ist und somit das Risiko von „Webfehlern“ nicht ganz gering, und zum anderen eine Gegenprüfung durch eine Schätzung von θ per Hand praktisch ausgeschlossen ist. Die Gefahr, dass ein Fehler im Schätzprogramm durch einen parallelen Fehler in der Software, die das Verhalten der künstlichen Vpn generiert, verdeckt wird, besteht nicht, da letztere von ihrer Struktur her einfach und gut überprüfbar ist.

8.2 Die Spezifikation von Basismodellen

Das Problem, ein Basismodell spezifizieren zu müssen, bleibt jedoch bestehen. Um θ schätzen zu können, brauchen wir also eine Vermutung, besser noch genaues Wissen, wie Entscheider ambige Wahrscheinlichkeiten auflösen. Zusätzlich ist von Belang ob und wenn ja in welcher Weise diese vermuteten Wahrscheinlichkeiten in Entscheidungsgewichte w überführt werden und wie die Nutzenfunktion für die Lotteriewerte aussieht (vgl. Anfang dieses Kapitels). Eine Möglichkeit zur Spezifikation eines Basismodells besteht darin, ein solches Modell aus Theorien, die Entscheidungen unter Ambiguität beschreiben, abzuleiten. Eine andere könnte darin bestehen empirisch vorzugehen. Um herauszufinden, wie ambige Wahrscheinlichkeiten aufgelöst werden, könnten also die Vpn befragt werden, wie sie mit der ambigen Wahrscheinlichkeitsinformation verfahren, oder man könnte sie alternativ bei einer Reihe von Spielesbündeln einschätzen lassen, wie hoch die Gewinnwahrscheinlichkeit der einzelnen Lotterien ist.

Eine empirische Herangehensweise ist aus zwei Gründen problematisch. Zum einen bleibt unklar, ob Vpn dazu in der Lage sind, ihre eigene Strategie zu verbalisieren. Im Falle eines spezifischeren Nachfragens besteht somit die Gefahr, dass nur eine möglichst plausible Regel ad hoc erfunden wird (zur Problematik introspektiver Daten siehe auch Nisbett und Wilson, 1977). Für die andere Vorgehensweise, Vpn konkrete Gewinnwahrscheinlichkeiten schätzen zu lassen, ergibt sich ein ähnliches Problem. Hier bliebe unklar, ob bei der Rangordnung der Lotterien, wie sie mein experimentelles Paradigma erfordert, Ambiguität auf dieselbe Weise aufgelöst wird, wie dies nach einer expliziten Anweisung dies zu tun der Fall ist. Doch selbst wenn man so zuverlässige Informationen darüber erhält, wie Vpn ausgehend von der ambigen Wahrscheinlichkeitsinformation zu einer Schätzung der Gewinnwahrscheinlichkeit gelangen, so ergibt sich daraus allein noch nicht unbedingt ein gutes Basismodell. Schließlich zeigen etwa die hier bei der Darstellung der Prospect Theorie dargestellten Ergebnisse, dass davon auszugehen ist, dass ein solcher Schätzwert noch durch eine Gewichtungsfunktion modifiziert wird (vgl. 3.3). Trotz der damit einhergehenden Schwierigkeiten habe ich nicht nur theoriebasierte Basismodelle verwendet sondern auch ein empirisch begründetes, das auf einem entsprechenden Vorversuch fußt. Dieses wird im nächsten Abschnitt 8.2.1 vorgestellt. Im darauf folgenden Teil 8.2.2 werden dann die theoriebasierten Basismodelle vorgestellt.

Jedes der verwendeten Modelle hat θ als freien Parameter, der nach der oben beschriebenen Methode geschätzt werden soll (vgl. 8.1). Wenn dies für alle Modelle geschehen ist, kann überprüft werden, ob sie sich im Fit der empirischen Daten unterscheiden. Gibt es ein „Siegermodell“, so werden dessen Schätzwerte für θ ausgewertet. Ist dies nicht der Fall, so bleibt zu hoffen, dass die verschiedenen Modelle zu qualitativ ähnlichen θ -Schätzern kommen.

8.2.1 Ein empirisch begründetes Basismodell

Die empirische Annäherung an das Basismodellproblem sah so aus, dass 14 Vpn für jede der drei Verdeckungsgrößen, die im nachfolgenden Experiment I zum Einsatz kommen sollten (nämlich 25%, 55% und 85% Ambiguität), einen Sechserblock Lotterien präsentiert bekamen. Jede Lotterie bestand dabei aus deren Punktewert (ϵ) sowie dem zu einem Teil verdeckten Glücksrad (vgl. auch Abb. 17). Für jede Vp war jede Lotterie neu und zufällig, gemäß dem oben beschriebenen Paradigma, erstellt worden. Jeder Sechserblock war separat auf einem Blatt Papier ausgedruckt. Die Vpn bekamen die drei Aufgabenblätter zusammengeheftet ausgehändigt, wobei die Heftungsreihenfolge der drei Blätter jeweils zufällig bestimmt worden war.

Den Vpn wurde die Bedeutung der Lotterien und die Art ihrer Generierung erklärt. Ihre Aufgabe bestand darin, bei jedem Glücksrad durch Einzeichnen eines Strichs zu markieren, wo sie glaubten, dass sich das Ende der Gewinnfläche befände.

Die Einschätzaufgabe nahm nur wenige Minuten Zeit in Anspruch. Die Vpn waren zum Teil durch einen Aushang an der Uni und zum Teil durch Ansprechen von Bekannten gewonnen worden.

Als abhängige Variable diente der prozentuale Anteil der Verdeckungsfläche, der dem Gewinnbereich zugeschlagen wurde. Als mögliche Wirkgrößen, die diesen Prozess beeinflussen könnten, wurden untersucht (i) der Punktwert der Lotterie (ϵ), (ii) wieviel Prozent der sichtbaren Fläche des Glücksrades Gewinnfläche war und (iii) die Position der Lotterie auf dem Blatt. Zusätzlich wurde (iv) berücksichtigt, welche Schätzung die Vp bei der vorangegangenen Lotterie abgegeben hatte.

Dies hatte folgenden Grund: Menschen vermuten oftmals Abhängigkeiten zwischen zufälligen Ereignissen, die in Wirklichkeit voneinander unabhängig sind (z. B. Tune, 1964). Fällt etwa beim Roulette die Kugel mehrfach in Folge auf Rot, so meinen viele Menschen, dass die Chance, dass das nächste mal Schwarz kommt, jetzt erhöht ist. Dieses Phänomen wird auch als *Gambler's Fallacy* bezeichnet. Eng damit verwandt ist die Tatsache, dass Menschen oftmals an ein „Gesetz der kleinen Zahl“ glauben (z. B. Tversky & Kahneman, 1971); d. h. sie gehen davon aus, dass bereits kurze Folgen zufälliger Ereignisse die wesentlichen Charakteristika des zugrunde liegenden Zufallsprozesses widerspiegeln; sie glauben also etwa, dass auch in einer kurzen Serie von Münzwürfen das Verhältnis von Kopf und Zahl etwa ausgeglichen sein sollte. Aus diesem Grund kann man vermuten, dass sich bei der Schätzung der Größe der Gewinnfläche möglicherweise Abhängigkeiten zwischen Lotterien zeigen, etwa nach dem Motto „Wenn ich bei der vorangegangenen Lotterie bereits Pech hatte, und wie von mir geschätzt nur ein kleiner Teil der Verdeckung Gewinnfläche unter sich birgt, so sollte ich jetzt Glück haben, und unter der Verdeckung dieser Lotterie sollte sich ein relativ großes Stück Gewinnfläche verbergen“.

Bei der Prüfung, ob eine solche Abhängigkeit existiert, bin ich davon ausgegangen, dass die Lotterien in Leserichtung bearbeitet werden. Die Analyse beschränkte sich in jedem Sechserblock auf die Lotterien zwei bis sechs, da für die Lotterie eins keine Vorgängerschätzung existierte, zumindest keine für eine Lotterie mit gleicher Ambiguität.

Da es wie gesagt unsicher ist, inwieweit sich die Ergebnisse eines solchen Vorversuchs auf das Vorgehen im eigentlichen Experiment übertragen lassen, erschien es mir nicht sinnvoll, ein detailliertes, übergenaues Basismodell zu erstellen. Daher erfolgte die Auswertung nicht für jedes Ambiguitätsniveau getrennt sondern über diese gemittelt (und dies ist auch der Grund dafür, dass ich mich mit einer relative kleinen Anzahl Vpn begnügt habe). Die Schätzungen der Vpn zeigten intra-individuell deutliche Unterschiede; ermittelt man für jede Vp den Interquartilrange ihrer Schätzungen und mittelt dann diese Interquartilranges, so ergab sich ein Wert von 44,2. Lediglich eine Teilnehmerin wandte bei allen 18 Lotterien konsequent dieselbe Strategie an und schätzte, dass die Gewinnfläche stets bis zur Hälfte der Verdeckungsfläche reichte.

Die deutliche intraindividuelle Streuung der Schätzwerte deutet darauf hin, dass die Vpn versuchten, in ihren Schätzwerten typische Charakteristiken des Zufallsprozesses, wie die Variabilität und die Verteilung der Ergebnisse, abzubilden; ein solches Vorgehen wird *Probability Matching* genannt und ist bei der Vorhersage zufälliger Ereignisse weit verbreitet (Estes, 1964). In meinen Augen liegt es nahe anzunehmen, dass die Vpn davon ausgingen, dass alle Positionen unterhalb der Verdeckung mit gleicher Wahrscheinlichkeit das Ende der Gewinnfläche markierten. In diesem Falle hätte sich für die Verteilung der wahren prozentualen Anteile ein Interquartilrange von 50 ergeben (alle Ergebnisse zwischen 0% und 100% wären ja gleich wahrscheinlich gewesen). Die Interquartilranges der

Schätzungen der Vpn lagen mit einem Mittelwert von 44,2 nicht allzu weit von diesem Wert entfernt; der Unterschied zu 50 erwies sich als nicht signifikant verschieden von diesem Wert ($T = 1,1$; $df = 13$; $p = .31$).

Was den Mittelwert ihrer Schätzungen anbelangt, wichen die Vpn jedoch klar vom Erwartungswert von 50% ab – ihre eigenen Schätzungen mittelten sich zu einem Gewinnflächenanteil von 42,5%, lediglich zwei Teilnehmer gaben mittlere Schätzungen ab, die über 50% lagen. Erwartungsgemäß erwies sich der Unterschied zu 50% auch als statistisch signifikant ($T = 3,6$; $df = 13$; $p = .003$). Die Schätzungen der Gewinnwahrscheinlichkeiten fielen insgesamt also eher konservativ aus.

Die Schätzungen des Gewinnflächenanteils erwiesen sich als unbeeinflusst von der Position der Lotterie auf dem Blatt, dem Punktwert und dem Schätzwert der vorangegangenen Lotterie. Einen moderaten Einfluss zeigte der Anteil der Gewinnfläche an der sichtbaren Fläche ($P_{Sicht} / (1 - Ambig)$). Je kleiner dieser war, desto größer wurde der Anteil der Gewinnfläche unter der Verdeckung eingeschätzt. Eine solche negative Korrelation ergab sich bei 11 der 14 Vpn; die Zahl der negativen Korrelationen war somit signifikant höher als die der positiven ($\chi^2 = 4,6$; $df = 1$; $p = .03$). Der Median lag bei $r = -.39$.

Das Regressionsmodell, das sich für die Vorhersage der Schätzwerte der Vpn aus dem Anteil der Gewinnfläche an der sichtbaren Fläche ergab, wies letzterer ein β -Gewicht von $-.28$ zu; die Konstante betrug .55.

Das empirisch motivierte Basismodell sollte die Attraktivität einer Lotterie beschreiben als das Produkt aus ihrem Punktwert und ihrer subjektiven Gewinnwahrscheinlichkeit. Die subjektive Gewinnwahrscheinlichkeit wurde dabei gemäß der Parameter des gerade vorgestellten Regressionsmodells geschätzt. Die oben vorgestellte Regressionsgleichung gibt den *relativen* Anteil der Verdeckungsfläche, der als Gewinnfläche eingeschätzt wurde, wieder. Wenn dieser in eine absolute Größe umgerechnet wird, so ergibt sich das folgende Basismodell, das *Empirisch* heißen soll:

$$\text{Empirisch: Attraktivität} = \epsilon \cdot (P_{Sicht} + Ambig (.55 - .28 P_{Sicht}) / (1 - Ambig)).$$

8.2.2 Theoriegeleitete und minimalistische Basismodelle

Da ich aus den genannten Gründen (vgl. 8.2) nicht garantieren kann, dass das empirisch ermittelte Basismodell das Verhalten von Vpn in den tatsächlichen experimentellen Entscheidungssituationen adäquat beschreibt, habe ich mich dazu entschlossen, neben *Empirisch* mehrere theoriebasierte Basismodelle aufzustellen und gemeinsam mit den zwei bereits vorgestellten minimalistischen Modellen, *Normal* und *Konservativ*, zu erproben.

Die theoriegeleiteten Modelle beruhen auf der Venture Theorie. Neben dieser gibt es noch eine Vielzahl anderer Modelle zur Erklärung von Entscheidungen unter Ambiguität (vgl. Camerer & Weber, 1992). Hauptaugenmerk dieser Theorien ist es in der Regel, den wichtigsten Forschungsbefund im Bereich von Entscheidungen unter Ambiguität, der im Ellsberg Paradox zu Tage tritt, zu erklären, dass nämlich die Attraktivität von Lotterien systematisch mit der Präzision der Wahrscheinlichkeitsinformation variiert. Der Fokus dieser Arbeit ist ein anderer; ich vergleiche nicht Lot-

terien mit verschiedenen Ambiguitäten unmittelbar miteinander sondern untersucht, ob die Art der Bewertung von Lotterien vom Grad der Ambiguität abhängt, dergestalt dass bei hoher Ambiguität den Lotteriewerten ein stärkeres Augenmerk zukommt. Nicht nur die Venture Theorie schweigt zu diesem Punkt, mir ist auch keine andere Theorie bekannt, die diese Frage aufgreift; die Modelle gehen also zumindest implizit davon aus, dass Ambiguität zu keinem Strategiewechsel führt. Unter den Modellen, die überhaupt eine präzise Modellierung zulassen, nimmt die Venture Theorie eine gewisse Sonderstellung ein, insofern dass sie eine Abhängigkeit zwischen der Bildung des Entscheidungsgewichtes und dem Lotteriewert zulässt, aber nicht erzwingt. Varianten, die diese Abhängigkeit nicht vorsehen, sind dann, zumindest was die Punkte angeht, die für die Spezifikation eines Basismodells notwendig sind, im Ergebnis den anderen Modellen ähnlich. Daher erscheint es mir vertretbar auch weiterhin auf die Venture Theorie als einzigm theoretischen Modell Bezug zu nehmen und andere Theorien außer Acht zu lassen.

Obwohl die Venture Theorie Entscheidungen unter Ambiguität bereits vergleichsweise spezifisch beschreibt, bleiben dennoch, wenn man ein konkretes Basismodell für Entscheidungen im hier verwendeten Paradigma aufstellen will, eine Reihe von Freiheitsgraden bestehen. Zunächst ist die Frage zu lösen, welcher Wert als Ankerwert angenommen werden soll. Ich habe mich für $P_{\text{Sicht}} + \text{Ambig}/2$ entschieden. Dieser ist der mittlere aller möglichen Werte von P und erscheint somit als erster, von allen weiteren Überlegungen unberührter Schätzwert für die wahre Wahrscheinlichkeit hochplausibel.

Zweitens ist zu überlegen, welche der verschiedenen Funktionen aus der Abb. 6, bzw. welche Kombination von Funktionen hier angemessen ist. Die Tatsache, dass eine Vielzahl von Lotterien gespielt wird und die Ergebnisunsicherheit somit eher gering ist, spricht eher für die Anwendung der Funktionen c aus den beiden oberen Quadranten der Abb. 6. Denn wie erwähnt ist bei niedriger Ergebnisunsicherheit damit zu rechnen, dass die Ankerwerte nur wenig verändert werden, wie es bei diesen Funktionen der Fall ist. Unklar ist jedoch, ob die Unterschiede in den Punktzahlen der einzelnen Lotterien hoch genug sind, als dass zwischen verschiedenen Funktionen innerhalb eines Quadranten gewechselt werden sollte. Da mir eine Festlegung vorab schwer möglich erschien, habe ich aus der Venture Theorie verschiedene Modelle abgeleitet. Sie sollen im Folgenden dargestellt werden. P_a sei dabei stets der Ankerwert $P_{\text{Sicht}} + \text{Ambig}/2$. Die Funktionen $f_a(P_a)$, $f_b(P_a)$ und $f_c(P_a)$ bezeichnen dabei die in Abb. 6 dargestellten.

$$\text{Venture I: Attraktivität} = \epsilon \cdot f_c(P_a)$$

$$\begin{aligned}\text{Venture II: Attraktivität} &= \epsilon \cdot f_c(P_a), \quad \text{für } \epsilon \leq 50 \\ &\text{Attraktivität} = \epsilon \cdot f_b(P_a), \quad \text{für } \epsilon > 50\end{aligned}$$

$$\text{Venture III: Attraktivität} = \epsilon \cdot f_b(P_a)$$

$$\begin{aligned}\text{Venture IV: Attraktivität} &= \epsilon \cdot f_c(P_a), \quad \text{für } \epsilon < 33 \\ &\text{Attraktivität} = \epsilon \cdot f_b(P_a), \quad \text{für } 33 \leq \epsilon \leq 65 \\ &\text{Attraktivität} = \epsilon \cdot f_a(P_a), \quad \text{für } \epsilon > 65\end{aligned}$$

Mit aufsteigendem Index postulieren die Modelle eine verstärkte Verschiebung des Ankerwerts, stehen also für stärkere subjektive Unsicherheit.

Zusätzlich zu diesen vier Venture-Modellen wurden noch drei einfache Basismodelle aufgestellt; die beiden ersten sind bereits als *Normal* und *Konservativ* bekannt. Das Modell *Venture_{einfach}* setzt auf möglichst einfache Idee die grundlegende Idee der Venture-Theorie um, dass ein höherer in Aussicht gestellter Gewinn zu einer niedrigeren geschätzten Wahrscheinlichkeit führt⁸. In Abhängigkeit von der Gewinnhöhe wird also ein unterschiedlich großer Teil des verdeckenden Kuchenstücks als Gewinnfläche angenommen. Die spezifischen Parameter wurden dabei willkürlich gewählt, andere hätten diese Idee ebenfalls umgesetzt.

$$\text{Normal:} \quad \text{Attraktivität} = \epsilon \cdot (P_{\text{Sicht}} + Ambig/2)$$

$$\text{Konservativ:} \quad \text{Attraktivität} = \epsilon \cdot P_{\text{Sicht}}$$

$$Venture_{\text{einfach}}: \text{Attraktivität} = \epsilon \cdot [P_{\text{Sicht}} + (60 - \epsilon / 2) \cdot Ambig / 100]$$

Üblicherweise steht der Nutzen eines Gutes nicht in einer linearen Beziehung zu dessen Menge, sondern er kann statt dessen durch eine Sättigungsfunktion beschrieben werden (s. 2.3). Auch in

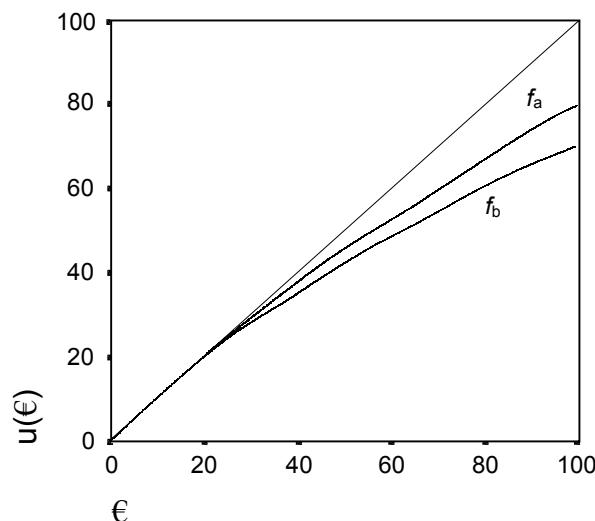


Abb. 21 Zwei nichtlineare Nutzenfunktionen, $f_a(\epsilon)$ und $f_b(\epsilon)$, die bei der Erstellung der Basismodelle Verwendung fanden.

dem von mir verwendeten Paradigma könnte ein solcher Zusammenhang bestehen, so dass etwa eine Gewinnaussicht von 80 Punkten weniger als doppelt so attraktiv wie eine von 40 Punkten ist. Diese Möglichkeit erscheint mir eher unwahrscheinlich, da die Gewinne einer Lotterie keinen Wert

⁸ Mit der Venture Theorie müsste man korrekter Weise natürlich von „Entscheidungsgewichten“ und nicht von geschätzten Wahrscheinlichkeiten sprechen. Die Unterschiede zwischen den beiden Begriffen sind jedoch im Kontext der Basismodelle irrelevant, so dass ich am einheitlichen Begriff der geschätzten Wahrscheinlichkeit festhalte.

an sich haben, sondern nur insofern, dass sie den prämienrelevanten Endpunktstand erhöhen. Zudem wurde in allen Experimenten über viele Runden gespielt (zwischen sechzig und achtzig). Somit war die relevante Punktsumme am Schluss des Versuchs mehrere Potenzen größer als der einfache Wert einer einzelnen Lotterie. Gemessen an diesem Schlussbetrag war die Bandbreite der Spielerträge also relativ gering, was einen linearen Zusammenhang zwischen Spielwert und -nutzen begünstigen sollte.

Um der Möglichkeit, dass sich der Nutzen der Lotteriewerte nicht linear entwickelt, dennoch Rechnung zu tragen, habe ich für jedes Basismodell zusätzlich zwei Varianten benutzt. Bei diesen wurde ϵ jeweils durch die in Abb. 21 dargestellten Funktionen $f_a(\epsilon)$ bzw. $f_b(\epsilon)$ ersetzt. Beide Kurven folgen der Idee eines Sättigungszusammenhangs – die spezifische Gestaltung erfolgte willkürlich.

8.3 Zusammenfassung

Zu Anfang dieses Kapitels wurden zwei Fragen aufgeworfen: Welches Kriterium ist bei der Schätzung von θ am besten geeignet, um die Passung von replizierten und empirischen Daten zu messen? Und welches Basismodell sollte zur Schätzung von θ herangezogen werden? Die Simulation III gibt Antwort auf die erste Frage und zeigt, dass θ derart geschätzt werden sollte, dass die von den Vp vorgenommenen Rangordnungen der Lotterien möglichst genau repliziert werden. Dem Basismodellproblem wird derart begegnet, dass eine ganze Reihe von Modellen ins Rennen geschickt wird. Ein Modell (*Empirisch*) beruht dabei auf einem Vorversuch, in dem die Vpn ausgehend von ambiger Wahrscheinlichkeitsinformation die wahren Wahrscheinlichkeiten schätzen sollten. Andere Modelle wurden aus der Venture Theorie abgeleitet. Die θ -Schätzungen desjenigen Modells, das den besten Fit des Versuchspersonenverhaltens erlaubt, sollen ausgewertet werden.

9. EXPERIMENT I

Experiment I überprüft die Gewichtungshypothese, also die Frage, ob ambiger Wahrscheinlichkeitsinformation ein geringeres Gewicht zukommt, anhand von Gewinnaufgaben im beschriebenen Paradigma; dabei werden vier Ambiguitätsniveaus verglichen (0%, 25%, 55% und 85%). Wenn die Gewichtungshypothese zuträfe, müsste θ mit steigender Ambiguität fallen. Das zentrale Ergebnis ist der Abb. 28 zu entnehmen.

9.1 Materialien und Aufbau

Das Paradigma von Experiment I wurde bereits unter Punkt 7.1 beschrieben. Es wurde ein einfaktorielles Within-Subjects Design verwirklicht; die Manipulation der abhängigen Variable „Grad der Ambiguität“ erfolgte dabei in den vier Abstufungen 0%, 25%, 55% und 85%⁹. Auf jeder Stufe waren 15 Aufgaben zu lösen. Alle Lotterien einer Aufgabe wurden mit demselben Grad an Ambiguität präsentiert; eine jede erforderte, die sechs Lotterien in eine Rangordnung zu bringen. Die Aufgaben wurden geblockt dargeboten, die Reihenfolge der Blöcke war dabei randomisiert. Für jede Vp wurden alle Lotterien neu und zufällig nach erzeugt, dabei galten die unter 4.2 beschriebenen Randbedingungen.

Abb. 22 gibt die Bildschirmgestaltung wieder. Die Vergabe der Wertungsfaktoren erfolgte per Mausklick in die grauen Kästchen unterhalb der Lotterien. Die Wertungsfaktoren waren in absteigender Reihenfolge zu vergeben, zuerst also Faktor sechs, dann Faktor fünf usw. Die gewählten Faktoren konnten jederzeit korrigiert werden. Durch Drücken der OK-Taste wurde ein Lotterieblock abgeschlossen und die nächsten sechs Glücksspiele wurden präsentiert. Die OK-Taste wurde erst dann freigegeben, wenn alle Wertungsfaktoren vergeben worden waren. Die Vpn erhielten keine Rückmeldung über ihren Erfolg, sie bekamen also weder Informationen über ihren aktuellen Punktestand noch darüber, welche Lotterien gewonnen hatten und welche nicht. Sie konnten lediglich am Schluss des Versuchs sehen, wieviele Punkte sie insgesamt erzielt hatten.

Entgegen der Information, die in der Versuchsinstruktion gegeben worden war, wurden die einzelnen Lotterien dabei nicht ausgespielt, sondern es wurde der Erwartungswert einer jeden Lotterie gutgeschrieben. Auch wenn im Folgenden von Punktzahlen die Rede ist, bezieht sich dies immer auf Erwartungswerte und nie auf Ergebnisse tatsächlich gespielter Lotterien, die mit einem störenden Zufallsfehler behaftet gewesen wären.

Über einen im Programm integrierten rückwärts laufenden Aufgabenzähler waren die Vpn stets darüber informiert, wieviele Aufgaben sie bereits geschafft hatten bzw. wieviele noch vor ihnen lagen.

⁹ Diese Werte waren ursprünglich einmal gewählt worden, weil sie anhand der Korrelationsmaße $V(C, V)$ und $V(C, P)$ erlauben, zwischen der *normalen* und der *konservativen* Disambiguierungsstrategie zu diskriminieren. Nachfolgende Änderungen in der Auswertungsstrategie haben dies jedoch im Wesentlichen überflüssig gemacht.

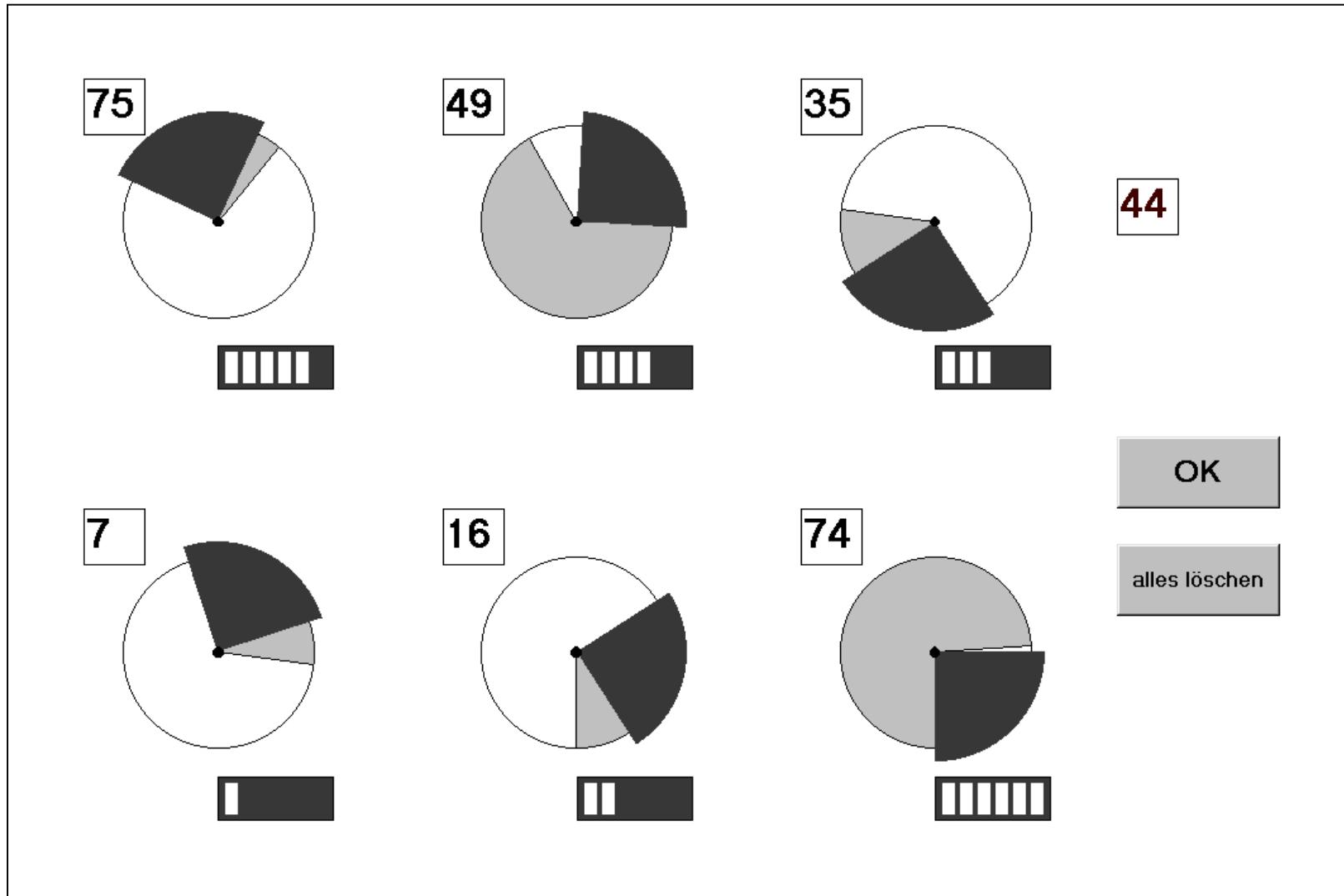


Abb. 22 Exemplarische Darstellung eines Bildschirmaufbaus bei einer Ambiguität von 25%. Im Beispiel sind bereits alle Wertungsfaktoren gesetzt – für die Lotterie oben links etwa 5. Die Box mit der 44 stellt den Rundenzähler da. Im Versuch waren die Gewinnflächen und die Wertungsfaktoren gelb dargestellt, die Verdeckungsfläche rotbraun.

9.2 Versuchsteilnehmer und Versuchsablauf

Für die Teilnahme an Experiment I konnten 26 Frauen und 32 Männer über eine Anzeige in einem örtlichen Veranstaltungskalender sowie über Aushänge an Menschen gewonnen werden. Die große Mehrheit der Vpn studierte oder machte eine Ausbildung, der Altersmedian lag bei 23 Jahren. Die Teilnahme wurde mit fünf Mark vergütet; zusätzlich konkurrierten die Teilnehmer um Prämien in Höhe von achtzig, sechzig und vierzig Mark, die für die drei Bestplatzierten ausgelobt waren.

Die Vpn lasen eine schriftliche Instruktion. Im Anhang ist unter A2 exemplarisch die Anleitung von Experiment III wiedergegeben; die Instruktion von Experiment I bestand aus entsprechenden Teilstücken. Nachdem die Vpn die Versuchsbeschreibung gelesen hatten, wurde ihnen die Bedienungsweise des Programms am Rechner gezeigt. Dabei wurde die Aufgabe noch einmal mündlich erläutert. Dies geschah in Gruppen von bis zu vier Personen. Bei der Bearbeitung der Aufgaben waren die Vpn allein oder maximal zu zweit in einem Raum.

Die Vpn hatten die Möglichkeit, vor Bearbeitung der experimentellen Aufgaben beliebig viele Übungsaufgaben zu bearbeiten. Diese enthielten nur Lotterien ohne Ambiguität. Auch hier erhielten sie keine Rückmeldung über die Glücksspielausgänge.

In der Wahl des Bearbeitungstemos waren die Teilnehmer frei. Für die sechzig Aufgaben benötigten sie im Mittel etwa 30 Minuten.

9.3 Ergebnisse

9.3.1 Ausschluss von Daten von der weiteren Analyse

Die Tatsache, dass einige der Vpn den Versuch mit einem deutlich niedrigeren Punktestand abgeschlossen haben als das Gros der übrigen, legt den Verdacht nahe, dass sie die Aufgabe nicht verstanden haben. Um solche Teilnehmer identifizieren und von der weiteren Analyse ausschließen zu können, wurden für jede Vp über alle Aufgaben hinweg $V(C, \epsilon)$ und $V(C, P)$ bestimmt und als Leistungsmaß der Mittelwert dieser beiden Werte gebildet. Eine zufällige Zuweisung der Rangwerte zu den einzelnen Lotterien führt dazu, dass sowohl $V(C, \epsilon)$ als auch $V(C, P)$ etwa den Wert .5 annehmen — zur Erinnerung: bei der V -Korrelation ist dies Ausdruck eines fehlenden Zusammenhangs. Ein solcher Wert kann also als Fehlen jeglicher Leistung im Sinne der Aufgabenstellung interpretiert werden. Der Vorteil des hier gewählten Leistungsmaßes liegt also darin, dass eine Null-Leistung eindeutig definiert ist. Eine Selektion wäre natürlich auch über die erzielten Punktwerte möglich gewesen; diese ist aber insofern schwieriger, als dass die einzelnen Lotterien für jede Vp neu erzeugt worden waren. Somit schwankten auch die Punktzahlen, die jede Vp erzielt hätte, wenn sie die optimale oder eine Zufallsstrategie verfolgt hätte.

Ein Scatterplot des Leistungsmaßes wies zwei deutliche Ausreißer mit den Werten .47 und .53 auf; die übrigen Werte zeigten eine rechtsschiefe Verteilung mit einem Median von .72. Die Daten der beiden Vpn, die die beschriebenen Ausreißer produziert hatten, wurden von der weiteren Analyse ausgeschlossen.

9.3.2 Die Leistung der Vpn

In die Fragestellung, inwieweit sich Entscheider bei verschiedenen Graden von Ambiguität in unterschiedlicher Weise auf die Wahrscheinlichkeits- bzw. die Wertkomponente der einzelnen Lotterien stützen, geht implizit die Vorstellung ein, dass Vpn einer Maximierungsstrategie folgen, dass sie ihre Entscheidungen also so treffen, dass der Erwartungswert oder eine mit diesem zusammenhängende Größe maximiert wird. Ob diese Vorstellung realistisch ist, lässt sich abschätzen, wenn man die tatsächlichen Leistungen der Vpn mit derjenigen vergleicht, die sich ergibt, wenn eine gute Maximierungsstrategie konsequent befolgt wird.

Die Ergebnisse, die *Normal* erzielt, geben einen tauglichen Vergleichsmaßstab, denn diese Strategie ist eine Maximierungsstrategie, und sie ist in verschiedenerlei Hinsicht eine „gute“ Strategie. Zum einen ist sie „nur“ eine gute Strategie und nicht die optimale. Denn es ist zu erwarten, dass sich die optimale Strategie auf zweierlei Weise von *Normal* unterscheidet. Einerseits sollte sie berücksichtigen, dass $(P_{\text{Sicht}} + \text{Ambig}/2)$ als Schätzung für P Werte produziert, die in Richtung .5 verschoben sind (vgl. 7.1). Diese Schätzung müsste also korrigiert werden, indem diese Werte noch einmal von der Mitte weggespreizt werden. Des weiteren muss die optimale Strategie berücksichtigen, dass die ambige Wahrscheinlichkeitsinformation weniger valide ist als die Wertinformation. Somit sollte also mit zunehmender Ambiguität der Wertinformation ein höheres relatives Gewicht zugemessen werden, 0 sollte also vermutlich kleiner als 1 sein. Eine Simulationsstudie, die später ausführlich dargestellt wird, zeigt jedoch, dass *Normal* Ergebnisse produziert, die sich fast nicht von denen der optimalen unterscheiden (der Unterschied beträgt lediglich .1%). Zum anderen ist *Normal* auch deswegen als Vergleichsmaßstab besonders geeignet, da es nur eines vergleichsweise geringen kognitiven Aufwandes bedarf, diese Strategie anzuwenden. Die optimale Strategie ist hingegen äußerst schwer zu erkennen und nur mit hohem rechnerischen Aufwand zu erschliessen ist, so dass sie als Maßstab für das Verhalten von Vpn auch dann untauglich wäre, wenn sie *Normal* deutlich überlegen wäre.

Für jede Vp wurde auf jeder Stufe des Faktors Ambiguität errechnet, wieviele Punkte sie mit der *normalen* Strategie gemacht hätte. Diese Werte wurden jeweils als 100% gesetzt und die real erzielten Punktewerte entsprechend in Prozentwerte umgerechnet. Abb. 23 gibt die prozentualen Differenzen zum Ergebnis von *Normal* wieder. Dieses Leistungsmaß zeigt also, um wieviel schlechter die Vpn abschnitten als eine gute Maximierungsstrategie. Da die empirischen Werte stark linksschief verteilt waren, sind hier die Mediane wiedergegeben. Um die Leistung der Vpn besser bewerten zu können, sind jeweils zwei weitere Benchmarks mit angegeben. Zum einen die Leistung, die resultiert, wenn die Lotterien nur anhand der ϵ -Werte in eine Rangordnung gebracht werden, wenn die Wahrscheinlichkeitsinformation also komplett ignoriert wird – diese Strategie soll künftig ϵ -Strategie heißen. Und zum anderen die Leistung, die sich aus einer einzigen, von mir nach dem Zufallsprinzip durchgeföhrten Rangordnung der Lotterien ergeben hat (*Zufallswahl*). Beide Vergleichsmaße sind wiederum als prozentuale Differenz zur Leistung von *Normal* angegeben.

Wie zu sehen ist, verläuft die Kennlinie der ϵ -Leistung nicht parallel zur X-Achse, welche die Leistung von *Normal* markiert. Dies liegt daran, dass der „Mehrwert“ von *Normal* daher röhrt, dass

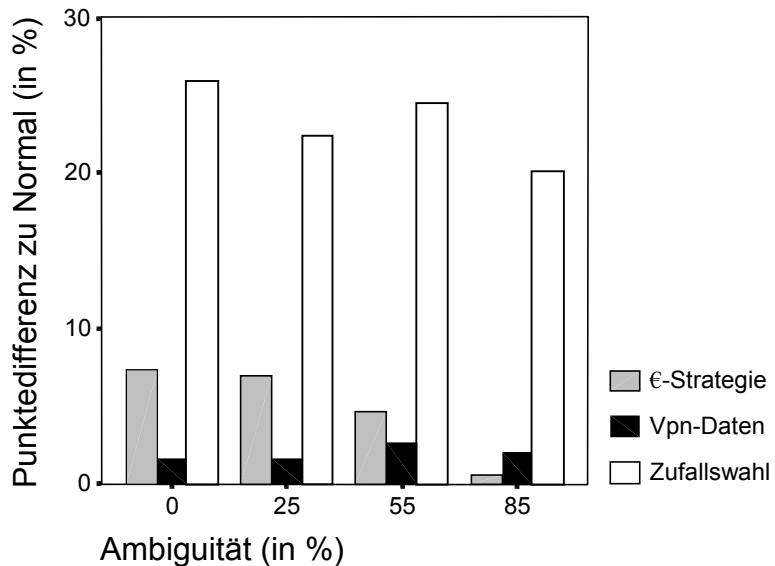


Abb. 23 Experiment I: Angegeben ist, um wieviel Prozent die Punkteleistungen der ϵ -Strategie, der Vpn und die einer Zufallswahl unter der Leistung von *Normal* lag [Mediane].

diese Strategie auch die gegebene Wahrscheinlichkeitsinformation berücksichtigt. Je stärker diese Information verrauscht ist, desto geringer fällt dann natürlich die Überlegenheit dieser besseren Strategie aus. Gleiches gilt prinzipiell auch für die Ergebnisse, die wahlloses Verhalten produziert; dieser Trend ist hier jedoch aufgrund des recht hohen Zufallsfehlers nicht eindeutig.

Die prozentuale Differenz in der Leistung zu *Normal* ist zwar ein einfaches und anschauliches Maß. Es erlaubt aber doch nicht, alle interessierenden Vergleiche in sinnvoller Weise zu ziehen. Dies hat mehrere Gründe: Mit steigender Ambiguität müssen sich natürlich die erzielten Punktwerte verringern. Damit verändert sich aber auch die Basis der Prozentberechnung. Somit erscheint es fraglich, beispielsweise einen einprozentigen Leistungsabfall bei fehlender Ambiguität mit einem ebenso großen Leistungsunterschied auf dem höchsten Ambiguitätsniveau gleichzusetzen. Dieses Problem ließe sich leicht umgehen, wenn man unmittelbar die Unterschiede in den erzielten Punktwerten betrachtete. Dieses Maß könnte aber ein zweites bestehendes Problem nicht lösen: Bei fehlender Ambiguität liefert *Normal* das perfekte Ergebnis, die Lotterien werden nach ihren Erwartungswerten geordnet, ein (nach dem hier zugrunde gelegten Auswertungsmaßstab) besseres Ergebnis lässt sich prinzipiell nicht erzielen. Jede Abweichung von *Normal* muss somit zu einem schlechteren Ergebnis führen. Bei steigender Ambiguität ist dies jedoch nicht mehr der Fall. Ein Abweichen von *Normal* wird in der Regel zwar unklug sein, denn die Strategie ist sehr gut, sie wird aber bisweilen zu einer Verbesserung des Ergebnisses führen; denn je stärker die vorhandene Information über die Lotterien verrauscht ist, desto stärker weicht die Rangordnung der Lotterien, die *Normal* produziert, von der perfekten, den Erwartungswerten der Lotterien folgenden Rangordnung ab. Somit erhöht sich die Chance, dass ein Abweichen von der *normalen* Strategie zu einem besseren Ergebnis führt.

Ich habe ein Leistungsmaß konstruiert, das die beschriebenen Probleme nicht aufweist. Es soll *Standard-Fehlerseinheit (SFE)* heißen, und es beruht auf der Idee, die Leistungsmessung auf den Punkt-

verlust zu gründen, der sich aus einer definierten Abweichung von *Normal* ergibt. Somit lassen sich dann die Leistungen auf verschiedenen Ambiguitätsniveaus sinnhaft miteinander vergleichen.

Ausgehend von der Rangreihe, die die *normale* Strategie für die sechs Lotterien einer Aufgabe herstellt, sieht diese Abweichung so aus, dass zwei zufällig ausgewählte, benachbarte Ränge vertauscht werden. Die durchschnittlich zu erwartende Punktminderung, zu der eine solche Vertauschung pro Sechserblock führt, soll dabei einer *SFE* entsprechen. Um diesen Erwartungswert zu ermitteln, habe ich mittels einer Computersimulation einen entsprechenden Durchlauf mit 2 000 000 Aufgaben (500 000 pro Ambiguitätslevel) gemacht, die nach demselben Prinzip erstellt wurden wie die Lotterien des Experimentes. Dabei ergaben sich für einen Sechserblock Lotterien gegenüber *Normal* mittlere Punktdifferenzen von 10,05 (fehlende Ambiguität), 9,84 (25% Ambiguität), 9,01 (55% Ambiguität) und 7,15 (85% Ambiguität).

Um zu überprüfen, ob *SFE* auf allen Stufen von Ambiguität gleichartig misst, habe ich eine zweite Aufgabenreihe mit einem stärkeren Fehler analysiert. Da dieser Fehler auf allen Ambiguitätsstufen gleicher Art war, sollte er auch gleiche *SFE*-Werte produzieren, andernfalls wäre das *SFE*-Maß nicht brauchbar. Um dies zu prüfen, wurde bei 2 000 000 neuen Aufgaben dergestalt von der Rangfolge, die die *normale* Strategie produziert, abgewichen, dass zwei zufällig ausgewählte Ränge vertauscht wurden, die *zwei* Rangplätze voneinander entfernt waren. Bei den Ambiguitätsstufen 0%, 25% und 85% verringerte sich die Leistung dadurch um 3,9 *SFE*, bei 55% Ambiguität ergab sich ein Abfall von 3,8 *SFE*. Um zu testen, ob hier eine systematische Abweichung vorliegt, wurden für weitere 4 000 Aufgaben die Punktdifferenzen in *SFE* ermittelt, und diese 4 000 *SFE*-Werte (1 000 pro Ambiguitätsniveau) dann einer ANOVA unterzogen. Dabei fand sich kein signifikanter Effekt des Ambiguitätsniveaus auf die *SFE*-Messung ($F = .77; df = 3; p = .51$). *SFE* misst also auf allen Ambiguitätsstufen gleichartig und ist somit ein geeignetes Maß für die Leistungsmessung.

Da die Leistungsdifferenzen zur *normalen* Strategie, gemessen in *SFE*, in allen vier Zellen extrem linksschief verteilt waren, zeigt Abb. 24 deren Mediane. Höhere Werte stehen hier wiederum für schlechtere Leistungen. Zusätzlich zu den empirischen Werten ist nochmals die Leistungseinbuße der ϵ -Strategie, ebenfalls in *SFE*, angegeben. Die zufällige Zuordnung der Ränge führte zu einem Leistungsabfall von 16,8 *SFE* auf allen Stufen von Ambiguität.

Um zu überprüfen, ob es einen generellen Einfluss der Ambiguität auf die relative Leistung der Vpn gab, bin ich wegen der durchweg starken Linksschiefe der *SFE*-Werte auf den verallgemeinerten Vorzeichentest von Friedman (Bortz, Lienert & Boehnke, 1990) ausgewichen. Pro Vp gingen dabei vier Werte ein, nämlich der Leistungsmittelwert jeder Ambiguitätsstufe, gemessen in *SFE*. Die Prüfgröße beruhte dann auf der Rangordnung dieser vier Werte, die sich für jede Vp ergab. Der per Augenschein zu vermutende Effekt wurde dabei klar bestätigt ($\chi^2 = 13,1; df = 3; p = .004$). Einzelvergleiche wurden nach dem von Conover vorgeschlagenen Verfahren (Bortz, Lienert & Boehnke, 1990, Gleichung 6.83) durchgeführt. Dabei erwiesen sich die Unterschiede zwischen den Bedingungen 0%-55% und 0%-85% sowie zwischen 25%-55% und 25%-85% als signifikant (jeweils $p < .05$; zweiseitige Testung).

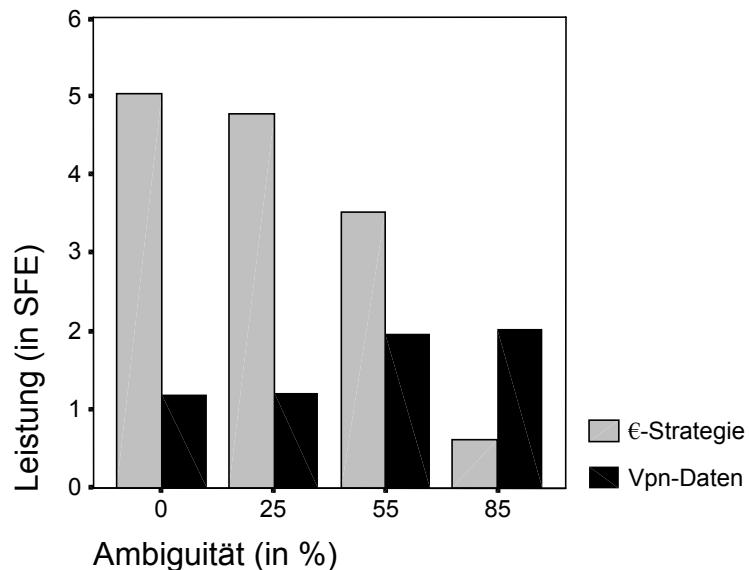


Abb. 24 Experiment I: Leistungsdifferenz der Vpn und der ϵ -Strategie zu *Normal*, gemessen in *SFE* [Mediane].

Relativiert man die Leistung der Vpn an der ϵ -Strategie, zeigte sich mit zunehmender Ambiguität eine kontinuierliche und deutliche Verschlechterung; während bei fehlender Ambiguität die mittlere Leistung, die sich aus der ϵ -konformen Rangordnung der Lotterien ergibt, von 91% der Vpn überboten wurde, schafften dies auf der höchsten Ambiguitätsstufe nur noch 21% der Teilnehmer. Wenn wir davon ausgehen, dass die Vpn insgesamt einer relativ guten Maximierungsstrategie folgten, so ist ein kontinuierlicher Leistungsabfall, gemessen an der ϵ -Leistung, zu erwarten: wie bereits erläutert, schmilzt die Überlegenheit einer guten gegenüber der ϵ -Strategie bei zunehmender Ambiguität. Das Ergebnis, dass die Vpn bei einer Ambiguität von 85% aber schlechter abschnitten als die simple ϵ -Strategie, bleibt jedoch bemerkenswert; die Differenz erweist sich dabei als statistisch hochsignifikant ($p < .001$, Rangsummentest).

Zusammenfassung

Zusammenfassend lässt sich also sagen, dass die Vpn bei fehlender und geringer Ambiguität (25%) einer Maximierungsstrategie erfolgreich nachgingen. Ihre Leistungen waren in der Regel weniger als zwei Prozent schlechter als die einer sehr guten Vergleichsstrategie und deutlich besser als diejenige einer simplen Strategie („ordne die Lotterien gemäß ihrer Punktwerte“). Die relative Leistung der Vpn ließ bei stärkerer Ambiguität (55% u. 85%) jedoch nach. Für die mittlere Vp war der Leistungsabfall gegenüber der *normalen* Strategie (gemessen in *SFE*) hier nahezu doppelt so hoch wie bei fehlender und geringer Ambiguität. Dies hatte zur Folge, dass die Leistung auf der Stufe größter Ambiguität sogar unter die Marke sank, die die simple ϵ -Strategie erreichte.

Dies kann entweder bedeuten, dass die Vpn mit zunehmender Ambiguität zu einer schlechteren Strategie wechselten, oder dass höhere Ambiguität die Teilnehmer veranlasste, ihre Strategie weniger konsequent zu verfolgen, so dass die Antworten hier stärker willkürlich gegeben wurden. Eine dritte

Erklärungsmöglichkeit wäre, dass die Probanden generell eine Strategie verfolgten, die lediglich bei geringer Ambiguität zu gutem Erfolg führt. Welche dieser drei Möglichkeiten zutrifft, soll später noch näher beleuchtet werden.

9.3.3 Die Identifikation eines geeigneten Basismodells

Zur Replikation des Versuchspersonenverhaltens und zur Schätzung von θ wurden die unter Punkt 8.2 beschriebenen acht Basismodelle ins Rennen geschickt. Wie bereits erläutert, wurde dabei jedes Modell in drei Varianten erprobt, die sich daraus ergaben, dass jeweils unterschiedliche Zusammenhänge zwischen ϵ und $u(\epsilon)$ angenommen wurden. Somit ergab sich eine Gesamtzahl von 24 Modellen.

Für jede Vp und jede Ambiguitätsstufe wurde θ auf Grundlage eines jeden Modells so geschätzt, dass das durch das Modell replizierte und das tatsächliche Verhalten der Vp einander möglichst ähnlich waren. Kriterium für Ähnlichkeit war dabei, dass die absolute Differenz zwischen den LotterieRangreihen, die die Vp gebildet hatte und denjenigen, die durch das Modell gebildet worden waren, minimal war (vgl. 8.1). Der Suchraum für θ umfasste jeweils das Intervall [.1, 50]; dieses wurde mit einer Schrittweite von .1 komplett nach dem besten Wert für θ durchsucht. Falls mehrere θ -Werte zu einer gleich guten Replikation des Versuchspersonenverhaltens führten, so wurde der unterste Wert als Schätzwert für θ herangezogen (vgl. 8.1).

Wenn auf diese Weise für jede Vp und jedes Ambiguitätsniveau θ vierundzwanzigmal auf Grundlage eben so vieler Modelle geschätzt worden war, so ergab sich für jedes der auf diese Weise optimal angepassten Modelle ein Maß Δ , das anzeigt, wie gut das Modell die empirischen Daten wiedergab. Wenn aus dem Datensatz einer Vp eine beliebige Lotterie herausgegriffen wird, so gibt Δ an, welcher absolute Unterschied zwischen dem Rangplatz, den die Vp dieser Lotterie gegeben hat, und demjenigen, den das Modell repliziert, zu erwarten ist. Ist Δ also etwa .5, so macht das Modell im Mittel bei jeder Lotterie einen Vorhersagefehler von einem halben Rangplatz.

Welches Modell eignet sich am besten, das Verhalten der Vpn nachzuzeichnen? Um dieser Frage auf den Grund zu gehen, habe ich zunächst für alle vierundzwanzig Modelle den Mittelwert von Δ über die vier Ambiguitätsstufen gebildet; somit ergibt sich ein Gesamtmaß für die Güte des Modells, unabhängig vom Grad der Ambiguität. Dabei galt für alle acht Grundmodelle, dass die Variante, die einen linearen Zusammenhang zwischen ϵ und $u(\epsilon)$ annimmt, die beste Reproduktionsleistung erbrachte; dieser Befund ist nicht überraschend (vgl. 8.2.2). Da sich für alle Basismodelle die Annahme eines linearen Zusammenhangs zwischen ϵ und $u(\epsilon)$ als am günstigsten erwies, lohnte sich im Weiteren allein die Betrachtung dieser acht Modelle. Für diese ergab sich eine deutlich linksschiefe Verteilung der Δ -Werte. Von den acht Grundmodellen schnitt *Normal* am besten ab (vgl. Abb. 25); für dieses Modell betrug der Median von Δ .461, das zweitbeste Modell, *VentureI*, erreichte einen Δ -Wert von .469. Obwohl der Güteunterschied der beiden Modelle gering war, erwies er sich doch als statistisch signifikant ($p = .006$; Wilcoxon-Rangsummentest). Die Betrachtung der einzelnen Ambiguitätsstufen zeigte, dass lediglich auf einem der vier Niveaus eines der anderen Modelle *Normal* überlegen war: Bei einer Ambiguität von 25% schnitt *Empirisch* besser ab. Dieser Unterschied ging

jedoch bereits bei der Testung mittels Rangsummentest, der der Inflationierung des α -Fehlers nicht Rechnung trägt, mit einer sehr grossen Irrtumswahrscheinlichkeit einher ($p = .57$; Wilcoxon-Rangsummentest). Er bedarf daher keiner weiteren Beachtung. Abb. 26 zeigt, für wieviele Vpn jedes einzelne Modell die Daten am besten replizieren konnte. Hier ergibt sich ein deutlicheres Bild der Überlegenheit von *Normal*.

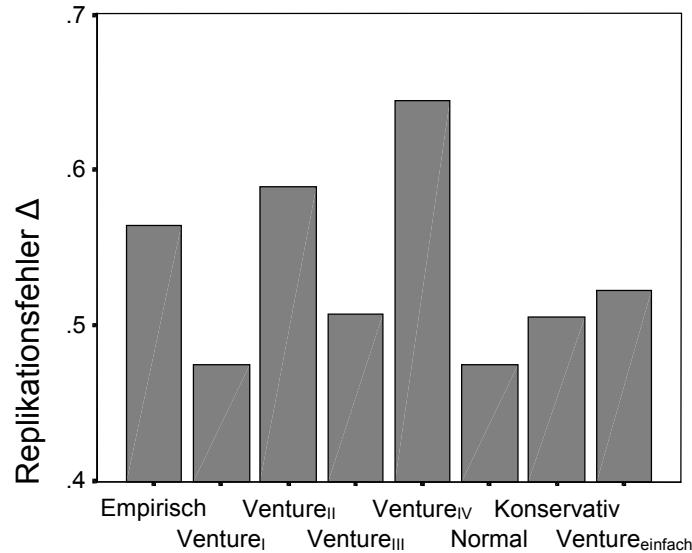


Abb. 25 Experiment I: Für die acht Basismodelle ist hier angegeben, wie gut sie das Verhalten der Vpn nachbilden konnten [Mediane von Δ].

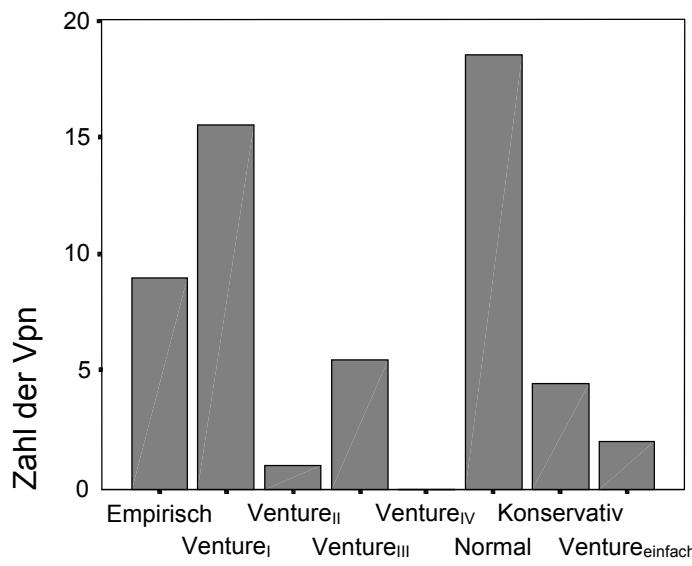


Abb. 26 Experiment I: Für jedes Basismodell ist die Zahl der Vpn abgetragen, deren Verhalten am besten durch dieses Modell erklärt werden konnte.

Das Bild wird noch eindeutiger, wenn wir uns vor Augen führen, dass die beiden besten Modelle, also *Normal* und *Venture_I*, einander äußerst ähnlich sind. Die Schätzer für θ , die beide Modelle

lieferten, korrelierten dem zufolge sehr hoch miteinander: zwischen $\rho = .87$ (25% Ambiguität) und $\rho = .97$ (85% Ambiguität). Das Verhalten von 61% der Vpn wurde am besten durch eines dieser beiden Modelle erklärt.

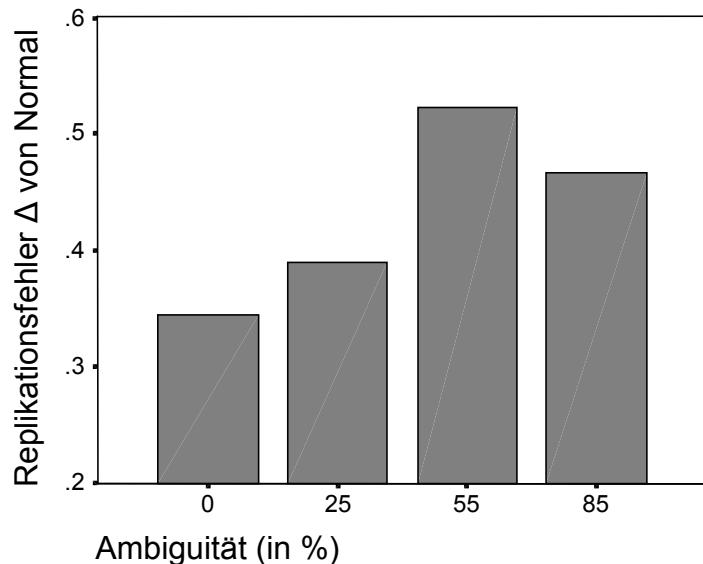


Abb. 27 Experiment I: Genauigkeit, mit der auf der Grundlage von *Normal* das Verhalten der Vpn auf den vier Ambiguitätsniveaus wiedergegeben werden kann [Δ Mediane].

Wie aus Abb. 27 zu ersehen ist, ließ sich das Verhalten der Vpn mit zunehmender Ambiguität weniger zuverlässig reproduzieren: Einem Δ um .37 bei fehlender und geringer Ambiguität stand ein Δ um .50 bei 55% und 85% Ambiguität gegenüber; der Fehler war hier also etwa um ein Drittel größer. Eine Rangvarianzanalyse für abhängige Daten nach Friedman wies, wie zu erwarten, einen statistisch signifikanten Effekt von Ambiguität auf Δ nach ($\chi^2 = 27,0$; $df = 3$; $p < .001$). Die einzelnen Unterschiede wurden nach dem von Conover vorgeschlagenen Verfahren getestet (vgl. Bortz, Liebert & Boehnke, 1991); dabei wurde ein Signifikanzniveau von 5% festgelegt. Statistisch signifikante Unterschiede fanden sich für die Ambiguitätskombinationen 0%-55%, 0%-85%, 25%-55% und 25%-85%.

Für den Anstieg des Reproduktionsfehlers gibt es zwei plausible Gründe. Zum einen könnte es sein, dass die Vpn ihre Strategie bei höherer Ambiguität wechselten, so dass *Normal* die Daten weniger gut reproduzierten konnte. Es wäre aber auch möglich, dass das Verhalten der Vpn hier insgesamt weniger strategiegeleitet und erratischer, willkürbestimmter war. Gegen die erste Annahme spricht, dass eine breite Palette von Basismodellen erprobt wurde und unter allen Versuchsbedingungen *Normal* nie von einem anderen Modell ernsthaft übertrroffen wurde. Hingegen wird die andere Vermutung, dass die Vpn sich bei erhöhter Ambiguität weniger konsistent verhielten, durch die Tatsache gestützt, dass die Leistungsdaten einen quasi identischen Verlauf zeigten (vgl. Abb. 24): Wenn die Vpn bei zunehmender Ambiguität weniger planvoll vorgingen, so wäre der gefundene Leistungsabfall wenig verwunderlich. Auch auf individueller Ebene ließ sich dieser Zusammenhang zwischen

der Reproduzierbarkeit des Verhaltens und der Leistung finden: Mittelt man für jede Vp Δ über die vier Ambiguitätsniveaus und korreliert dieses Gesamtmaß der Reproduktionsgüte mit ihrer Gesamtleistung, gemessen in *SFE*, so zeigte sich ein starker Zusammenhang ($r = .77$; $n = 56$; $p < .001$). Somit ist davon auszugehen, dass Δ mit wachsender Ambiguität größer wurde, weil hier in das Verhalten der Vpn eine stärkere Zufallskomponente eingeflossen war.

Insgesamt bewerte ich den Fehler, den *Normal* bei der Reproduktion des aktuellen Versuchspersonenverhaltens mit sich brachte, als relativ gering: Wenn bei der Reproduktion der Rangreihe eines Sechserblockes Lotterien der kleinstmögliche Fehler unterläuft, wenn also zwei benachbarte Ränge vertauscht werden, so führt dies bereits zu einem Δ von .33, der Median der empirischen Werte lag bei .46. Zum Vergleich: Würde man die von den Vpn vergebenen Ränge zufällig vorhersagen, resultierte ein Δ von 1,944; gebrauchte man die Rangreihen, die die ϵ -Strategie produziert, um das Vpn-Verhalten nachzuzeichnen, so resultierte ein Δ von 1,20.

Dass *Normal* das Versuchspersonenverhalten gut wiedergeben konnte, spricht dafür, dass auch die mit Hilfe dieses Modells gewonnenen Schätzwerte für θ brauchbar sind. Um deren Güte zu überprüfen, habe ich den Schätzprozess für θ (wiederum auf der Grundlage von *Normal*) noch einmal separat für gerade und ungerade Aufgabennummern durchgeführt. Die Korrelationen der so gewonnenen Split-Half-Schätzer betrugen .74, .73, .63 und .71 (Spearmans ρ , p stets $< .001$; in der Reihenfolge aufsteigender Ambiguität). Auf Grund der halbierten Aufgabenzahl wird dabei die Konsistenz der θ -Schätzungen unterschätzt; die Werte sollten somit gemäß der Spearman-Brown Formel korrigiert werden (z. B. Lienert & Raatz, 1994). Die berichtigten Werte für ρ betrugen .85, .84, .77 und .82. Diese Werte erscheinen mir insgesamt als zufriedenstellend.

9.3.4 Die relative Gewichtung der Wahrscheinlichkeit (θ)

Doch nun zur Kernfrage. Wie entwickelte sich θ bei steigender Ambiguität? Eine parametrische Analyse der abhängigen Variablen bot sich nicht an, da wir es hier mit einem Exponenten zu tun haben – dem zufolge ist etwa der Unterschied zwischen den θ -Werten 2 und 3 nicht derselbe wie der zwischen 4 und 5; zudem verteilten sich die Schätzwerte stark linksschief. Aus diesen Gründen gibt Abb. 28 in Abhängigkeit vom Grad der Ambiguität die Mediane der θ -Schätzwerte, die *Normal* lieferte, wieder. Bei fehlender Ambiguität ergab sich für θ ein Median von 1,55. Die Vpn machten ihre Entscheidungen also stärker von der Gewinnwahrscheinlichkeit als von der Gewinnhöhe abhängig. Ein solches konservatives Verhalten ist im Gewinnbereich die Regel (z. B. Tversky & Kahneman, 1981).

Mit zunehmender Ambiguität stieg θ tendenziell an. Dieses Muster war jedoch nicht eindeutig, da bei einer Ambiguität von 55% θ nicht nur absank sondern hier sogar den geringsten Median aufwies. Zu beachten ist hierbei jedoch, dass die Schätzprozedur bei einer Ambiguität von 55% und mehr für θ Werte liefert, die tendenziell etwas zu niedrig ausfallen (vgl. Abb. 20). Eine Rangvarianzanalyse nach Friedman zeitigte kein statistisch signifikantes Ergebnis ($\chi^2 = 5,4$; $n = 56$; $p = .145$). Von den Einzelvergleichen, die nach dem Verfahren von Conover durchgeführt wurden (vgl. Bortz, Lienert & Boehnke, 1990), erwies sich allein derjenige zwischen 0% Ambiguität und 85% Ambiguität als

signifikant ($p < .05$). Der Unterschied zwischen den Bedingungen 25% und 55% fiel zwar optisch markant aus, verfehlte den kritischen Wert, der ein auf dem 5%-Niveau signifikantes Ergebnis an-

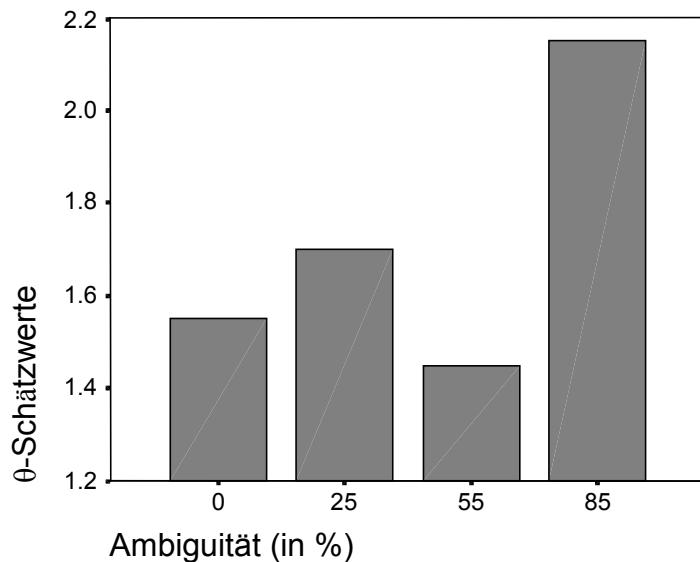


Abb. 28 Experiment I: Schätzwerte für θ , die auf der Grundlage von Normal gewonnen wurden [Mediane].

zeigt, jedoch am deutlichsten. Von daher mag es vertretbar erscheinen, trotz dieses Ausreißers einen kontinuierlichen Trend der Art, dass höhere Ambiguität generell zu einer stärkeren Gewichtung der Wahrscheinlichkeitskomponente führte, in den Daten zu sehen. Der Trendtest von Page macht es möglich, Rangdaten hinsichtlich einer solchen Trendhypothese zu testen (Bortz, Lienert & Boehnke, 1990). Der Testwert, der sich für die empirischen Daten ergab, lag dabei aber weit unterhalb des kritischen Wertes, der eine Irrtumswahrscheinlichkeit von .05 markiert.¹⁰ Von einem kontinuierlichen Zusammenhang ist also nicht auszugehen.

Das gefundene Muster für θ widerspricht klar der Gewichtungshypothese: Nichts in den vorliegenden Daten zeigt an, dass die Vpn ihre Urteile mit steigender Ambiguität stärker von den Lotteriewerten abhängig gemacht hätten.

9.3.5 Eine alternative Überprüfung der Gewichtungshypothese

Die Betrachtung der Schätzwerte für θ ist nur eine Möglichkeit, die Gewichtungshypothese zu überprüfen. Eine zweite Möglichkeit besteht darin zu ermitteln, welche Eigenschaften die Antwortdaten haben müssten, wenn θ sich mit zunehmender Ambiguität verändert, und dann zu überprüfen, inwieweit die Daten der Vpn diese Charakteristika tatsächlich aufweisen oder eben nicht. Falls hohe Ambiguität wirklich zu einer stärkeren Gewichtung der Wahrscheinlichkeitskomponente geführt hat, wie die Schätzungen für θ nahelegen, so sollten hier Größenunterschiede in der sichtbaren Gewinn-

¹⁰ Die kritischen Werte für den Trendtest nach Page sind nur bis zu einem N von 20 tabelliert (Bortz, Lienert & Boehnke, 1990). Da sich die kritischen Werte bis zu diesem N jedoch strikt linear entwickeln, ist eine Extrapolation für höhere Fallzahlen, wie ich sie vorgenommen habe, wohl unproblematisch.

fläche stärkere Konsequenzen für die Attraktivität der Lotterien haben als dies bei niedriger Ambiguität der Fall ist. Plottet man also für jedes Ambiguitätsniveau die Größe der sichtbaren Gewinnfläche gegen die Attraktivität der Lotterien (d. h. gegen die ihnen zugewiesenen Rangplätze), so sollte sich bei hoher Ambiguität ein steilerer Verlauf zeigen als bei niedriger. Abb. 29 ist zu entnehmen, dass dies tatsächlich der Fall war.

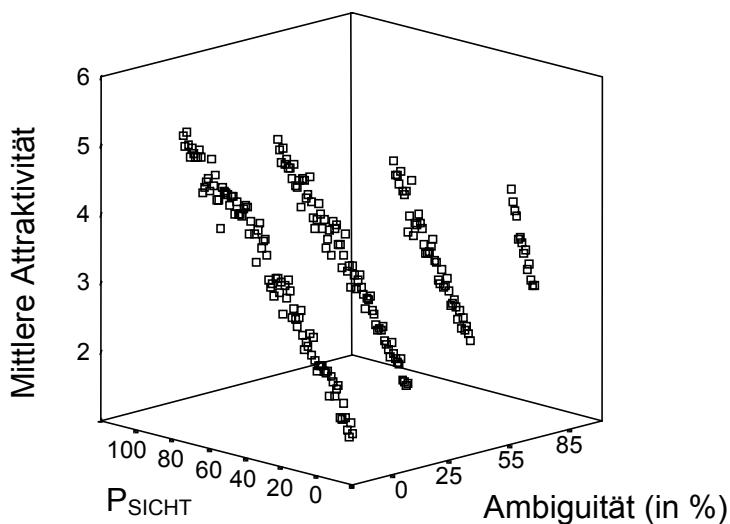


Abb. 29 Experiment I: Der Einfluss von P_{SICHT} auf die Attraktivität von Lotterien bei verschiedenen starker Ambiguität. Jeder einzelne Datenpunkt mittelt die Ränge aller Lotterien, die (auf diesem Ambiguitätsniveau) diese Größe von P_{SICHT} haben. Mit zunehmender Ambiguität werden die Verläufe steiler, d. h. gleiche Veränderungen in P_{SICHT} haben eine stärkere Wirkung auf die Attraktivität.

Simulation IV

Dies sollte man aber nicht voreilig als Bestätigung der θ -Schätzwerte gelten lassen. Denn vorher ist zu klären, wie sich mit steigender Ambiguität der Einfluss von P_{SICHT} auf die Attraktivität der Lotterien auswirkt, wenn sich die Entscheidungsstrategie nicht ändert, wenn θ also auf allen Ambiguitätsniveaus gleich ist – immerhin ist denkbar, dass sich dann ein ebensolches Muster wie in Abb. 29 dargestellt ergibt.

Die entsprechende Überprüfung erfolgte durch Computersimulation IV. Darin bearbeitete eine künstliche Vp alle Aufgaben, die in Experiment I verwendet worden waren. Diese „Vp“ folgte konsequent der Strategie *Normal* und gewichtete die geschätzte Gewinnwahrscheinlichkeit $P_{\text{SICHT}} + \text{Ambig}/2$ auf allen vier Ambiguitätsstufen mit $\theta = 1,55$. θ wurde also so gewählt, dass es dem Median der geschätzten θ -Werte bei fehlender Ambiguität in Experiment I entsprach.

Für jedes Ambiguitätslevel wurde dann die Regression von P_{SICHT} auf die Attraktivität der Lotterien (d. h. auf die ihnen zugewiesenen Rangplätze) bestimmt; gleiches erfolgte für die empirischen Daten aus Experiment I. In Abb. 30 sind die β -Gewichte der empirischen Daten denen für die

Simulationsdaten gegenübergestellt. Die Gewichte sind dabei so zu interpretieren wie die Steigungen in Abb. 29: Je höher sie ausfallen, desto stärker haben sich Unterschiede in P_{Sicht} auf die Attraktivität ausgewirkt. Wie an den Simulationsdaten zu sehen ist, steigen die β -Gewichte auch dann mit zunehmender Ambiguität, wenn θ konstant bleibt. Bis zu einer Ambiguität von 55% stimmen die β -Gewichte der empirischen Werte im Wesentlichen mit denen der simulierten Daten überein. Auf dem höchsten Ambiguitätsniveau zeigt sich jedoch ein deutlicher Unterschied – das Gewicht, das sich für die empirischen Daten ergab, fällt erheblich höher aus; der Unterschied beträgt ein mehrfaches des Standardfehlers, ist also auch statistisch bedeutsam.

Somit wurde das Ergebnis, welches die Schätzung von θ liefert hat, durch diese zweite Analyse bestätigt, dass nämlich in Experiment I hohe Ambiguität (85%) dazu geführt hat, dass bei der Attraktivitätsbewertung der Lotterien deutlich mehr Gewicht auf die Wahrscheinlichkeitskomponente gelegt wurde, als dies auf niedrigeren Ambiguitätsstufen der Fall war. Die Sichtweise, dass es sich dabei möglicherweise um ein kontinuierliches Phänomen handelt, wird durch diese zweite Analyse wiederum nicht unterstützt.

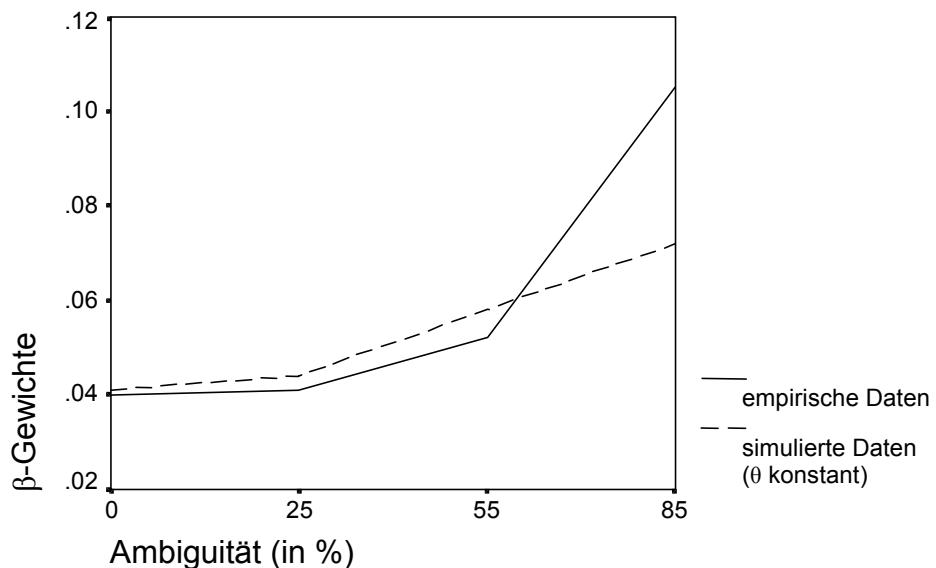


Abb. 30 Experiment I: Standardisierte β -Gewichte der Regression von P_{Sicht} auf die Attraktivität der Lotterien für empirische Daten und Daten der Simulation IV. Da in der Simulation θ konstant blieb, zeigt diese Linie, was passiert, wenn *kein* Strategiewchsel vollzogen wird. Dass der empirische Wert bei 85% Ambiguität viel höher liegt, zeigt somit, dass die Vpn der Wahrscheinlichkeitsinformation hier ein höheres Gewicht zukommen ließen.

9.3.6 Wodurch wird die Leistung der Vpn determiniert?

Nach der Erprobung der verschiedenen Basismodelle und nach den Schätzungen von θ ist es nun möglich, mehr darüber auszusagen, warum höhere Ambiguität zu einem Leistungsabfall führte. Des Weiteren erschließt sich jetzt auch, worauf die Leistungsunterschiede zwischen den Vpn beruhen.

Auf allen Ambiguitätsstufen war das Basismodell *Normal* am besten geeignet, das Antwortverhalten der Vpn zu reproduzieren. Es ist also nicht davon auszugehen, dass die Probanden hier eine entsprechende Verhaltensänderung zeigten, die die Leistungsunterschiede auf den verschiedenen Ambiguitätniveaus erklären könnte. Was erklärt die Leistungsunterschiede dann?

Interindividuelle Leistungsunterschiede

Zur Aufklärung interindividueller Leistungsunterschiede erwiesen sich zwei Variablen als bedeutsam: zum einen die Schätzwerte für θ und zum anderen die Werte für die Güte der Modellanpassung Δ (vgl. 9.3.3). Da sowohl das Leistungsmaß SFE als auch θ und Δ extrem schiefe Verteilungen aufwiesen und da θ zudem nicht intervallskaliert ist, beziehen sich alle nachfolgenden Aussagen über Zusammenhänge auf die Ränge der jeweiligen Variablen; die Variablen sind dementsprechend durch ein r indiziert. Auf jeder Ambiguitätstufe wurde ein Regressionsmodell zur Vorhersage von SFE_r aus θ_r und Δ_r aufgestellt, wobei beide Prädiktoren gleichzeitig in die Schätzung eingingen. Für alle Ambiguitätsniveaus ergab sich, dass die θ_r - und Δ_r -Werte praktisch unkorreliert waren; lediglich bei 85% Ambiguität zeigte sich ein zwar statistisch signifikanter aber relativ unbedeutender Zusammenhang ($r = .32; p = .02$). Da die Prädiktoren weitestgehend unabhängig voneinander waren, können die in Tab. 1 wiedergegebenen standardisierten β -Gewichte als Einflussstärke des jeweiligen Prädiktors auf das Kriterium interpretiert werden.

Tabelle 1: Standardisierte β -Gewichte für die Prädiktion der Leistung SFE_r aus θ_r und Δ_r .

	0% Ambiguität	25% Ambiguität	55% Ambiguität	85% Ambiguität
θ_r	.71	.58	n. s.	.21 ^a
Δ_r	.47	.43	.71	.72 ^a

a: In der Bedingung 85% Ambiguität korrelieren θ_r und Δ_r geringfügig miteinander ($r = .32$).

Die Interpretation der β -Gewichte von θ_r ist einfach; offensichtlich unterlagen die Vpn bei fehlender und geringer Ambiguität der Tendenz, die Wahrscheinlichkeitsdimension stärker zu gewichten als dies ihren Leistungen gut tat, so dass Teilnehmer, für die sich hohe θ -Schätzwerte ergaben, eher schlechter abschnitten. Umgekehrt erwies sich der Zusammenhang zwischen den Leistungen der Vpn (SFE_r) und der Modellierbarkeit ihrer Antworten durch die *normale* Strategie (Δ_r) auf den beiden unteren Ambiguitätsniveaus als mittelstark, während er auf den Stufen starker Ambiguität sehr hoch war. Generell galt hier also, dass Vpn, deren Antworten sich weniger gut durch das Basismodell *Normal* modellieren ließen, eher schwächere Leistungen erbrachten.

Dies kann zweierlei bedeuten: Zum einen, dass diese Personen eine andere, weniger erfolgreiche Strategie als *Normal* angewendet haben. Zum anderen, dass überhaupt weniger Strategie vorhanden war und mehr Willkür herrschte. Offensichtlich ist die zweite Interpretation deutlich eher zutreffend. Betrachtet man nämlich nicht nur *Normal* sondern alle acht Basismodelle und mittelt die Fehler, die diese bei der Replikation des Vpn-Verhaltens aufweisen, zu einem Gesamtmaß, so korre-

liert dieses genauso stark mit den Leistungen wie Δ_r . Das heißt verschiedene Basismodelle sagen ein sehr ähnliches Verhalten voraus. Kann das Verhalten einer Vp durch ein bestimmtes Modell nicht gut modelliert werden, so ist die Replikationsgüte, die sich bei der Verwendung anderer Basismodelle ergibt, ebenfalls niedrig. Es ist also nicht davon auszugehen, dass eine andere Strategie Verwendung fand sondern dass überhaupt wenig nachzeichbare Strategie vorlag.

Leistungsunterschiede in Abhängigkeit vom Ambiguitätsniveau

Somit wird auch klar, woher die kollektive Leistungseinbuße bei 55% und 85% Ambiguität herrührte (vgl. Abb. 24): Unter diesen Bedingungen ergaben sich für die Vpn deutlich höhere Δ -Werte (vgl. Abb. 27); wir können also davon ausgehen, dass hier insgesamt wahlloser agiert wurde als bei fehlender und niedriger Ambiguität, was zu einem Absinken der Leistung führte.

Dies vermag auch zu erklären, warum der Zusammenhang zwischen der Leistung und den Schätzwerten für θ auf den beiden oberen Ambiguitätsstufen so deutlich nachlässt (s. Tab. 1): Wenn das Verhalten hier insgesamt weniger strategiebestimmt war, konnten sich hier Strategieunterschiede, wie sie in unterschiedlicher Gewichtung der Wahrscheinlichkeitskomponente zum Ausdruck kommen, weniger stark auf die Leistung auswirken.

Probability Matching bei hoher Ambiguität? – Simulation V

Warum erscheint das Verhalten bei stärkerer Ambiguität als weniger strategiebestimmt? Im Vorversuch hatte sich gezeigt, dass die explizite Anweisung, ambige Wahrscheinlichkeiten aufzulösen, zu Probability Matching führte (vgl. 8.2.1). Probability Matching könnte auch die mit steigender Ambiguität nachlassende Regelhaftigkeit des Verhaltens in Experiment I erklären: Wenn nämlich die Vpn hier die Ambiguität nach dem Prinzip „ $P_{Sicht} +$ ein willkürlich bestimmtes Stück des Verdeckungsstückes“ aufgelöst hätten, so hätte sich die Varianz der so gewonnenen subjektiven Wahrscheinlichkeit aus zwei Quellen gespeist – P_{Sicht} und willkürlich bestimmtem Stück. Aus Sicht des Experiments ließe sich dann die Varianz der ersten Quelle nachverfolgen, die aus der zweiten nicht. Mit zunehmender Ambiguität wäre die Varianz von P_{Sicht} gesunken, die Fehlervarianz aber gestiegen. Infolge dessen müsste das resultierende Verhalten dann ungeordneter und willkürlicher wirken.

Um zu klären, ob die Vpn möglicherweise Probability Matching betrieben haben, soll Simulation V klären, welche Leistung bei dieser Strategie der Ambiguitätsauflösung im Vergleich zur Befol- gung von *Normal* zu erwarten ist. Die Simulation verwirklichte dazu das folgende 2×2 Design: (i) θ betrug 1 oder 2,15 (letzteres entspricht dem Median der empirischen Werte bei 85% Ambiguität), (ii) als Basismodell diente *Normal* oder ein Modell, das Probability Matching betreibt. In letzterem Fall bestimmte sich die vermutete Gewinnwahrscheinlichkeit als $P_{Sicht} + Rnd$, wobei Rnd eine Zufallszahl aus dem Intervall [0, .85] war. Dieselben 15 000 Aufgaben mit 85% Ambiguität wurden unter jeder Bedingung untersucht. Die Leistungen sind Tab. 2 zu entnehmen.

Tabelle 2: Ergebnis der Simulation V (SFE-Leistung).

	$\theta = 1$	$\theta = 2,15$
Normal	0,0	-0,2
Probability Matching	4,7	8,9

Wie zu ersehen ist, führte Probability Matching zu einer deutlichen Leistungseinbuße, insbesondere dann, wenn θ hoch war. Wir können somit praktisch ausschließen, dass das Gros der Vpn Probability Matching betrieben hat; denn die Leistung der mittleren Vp bei einem θ von 2,15 war mit 2,0 SFE erheblich besser als die in der Simulation für Probability Matching ermittelte. Inwieweit die bei hoher Ambiguität stärkere Regellosigkeit des Verhaltens Ausdruck einer mildereren Form von Probability Matching, also absichtsvoll, ist oder ob sich hier ein stärkerer unbeabsichtigter Fehler wiederspiegelt, lässt sich hier nicht entscheiden.

Zusammenfassung

Zusammenfassend kann man sagen, dass die Vpn die Wahrscheinlichkeitskomponente bei fehlender und geringer Ambiguität tendenziell überbewerteten. Dies zeigt sich schon darin, dass der Median der θ -Schätzwerte bei fehlender Ambiguität bei 1,55 lag. Das optimale Gewicht hätte hier jedoch bei 1 gelegen, da unter dieser Bedingung P und ϵ gleichermaßen zur Varianz der Erwartungswerte der Lotterien beitrugen und beide gleichermaßen fehlerfrei wahrgenommen werden konnten. Zusätzlich wurde die Leistung weitestgehend dadurch bestimmt, wie konsequent einer Strategie gefolgt wurde. Beide Faktoren gemeinsam konnten zwischen 57% (55% Ambiguität) und 66% (0% Ambiguität) der Leistungsvarianz der Vpn erklären.¹¹

9.4 Fazit und weiteres Vorgehen

Es zeigte sich, dass die Vpn die experimentelle Aufgabe erfolgreich bewältigen konnten; ihre Leistungen waren in der Regel nur geringfügig schlechter als die einer sehr guten Vergleichsstrategie. Mit zunehmender Ambiguität schwächte sich die Leistung jedoch ab, was in erster Linie darauf zurückzuführen war, dass die Vpn hier weniger konsistent geantwortet haben. Bei 85% Ambiguität fielen die Leistungen der Vpn im Mittel unter die der ϵ -Strategie.

Generell erwies es sich als sinnvoll, eine lineare Beziehung zwischen ϵ und $u(V)$ anzunehmen. Von den acht Basismodellen erwies sich *Normal* als den anderen klar überlegen; die Schätzungen für θ erfolgten somit auf Grundlage dieses Modells.

Die gefundenen Schätzwerte stehen in klarem Widerspruch zur Gewichtungshypothese – höhere Ambiguität führte nicht dazu, dass sich die Entscheidungen in höherem Maße auf die Werte der Lotterien stützten. Die Idee der Venture-Theorie, dass Ambiguität keinen Einfluss darauf hat, wie

¹¹ Diese Werte beziehen sich wiederum auf die in Ränge transformierten Daten.

stark sich die Entscheidungen auf die Wert- und wie stark auf die Wahrscheinlichkeitskomponente stützen, wurde teilweise bestätigt: Diese Beziehung galt bis zu einer Ambiguität von 55%. Sehr hohe Ambiguität (85%) führte jedoch zu einem Anstieg von θ .

Auf Grundlage einer neuen Methodik wurde ein neues Phänomen entdeckt: Der Grad der Ambiguität kann die Entscheidungsstrategie beeinflussen und zwar in einer Weise, die gegenintuitiv ist: Wenn die Information der Wahrscheinlichkeitskomponente besonders gering war, gründeten die Entscheidungen verstärkt auf dieser. Ist dieses Phänomen ernst zu nehmen? Ist der Befund stabil? Um auf diese Fragen eine Antwort geben zu können, zielte der zweite Versuch auf die Replikation dieses Phänomens. Gleichzeitig sollte dessen Deskription erweitert werden, indem nicht nur Gewinn- sondern auch Verlustaufgaben untersucht wurden. Diese Erweiterung erlaubt es, einen Schritt über die reine Deskription des Phänomens hinauszugehen und, zumindest ansatzweise, nach einer Erklärung für dasselbe zu suchen.

Zwei mögliche Erklärungen für das gefundene Phänomen

Denn wodurch kam es bei hoher Ambiguität dazu, dass die Vpn ihre Entscheidungen verstärkt auf die (nunmehr stark verrauschte) Wahrscheinlichkeitsinformation stützten? Zwei Möglichkeiten erscheinen mir dabei plausibel. Zum einen kann es sein, dass höhere Ambiguität im Rahmen der Disambiguierung eine stärkere Auseinandersetzung mit der Wahrscheinlichkeitskomponente anregt. Eine solche Position wird von der Venture Theorie vertreten (vgl. 3.4). Durch die vermehrte Aufmerksamkeit, die der Gewinnwahrscheinlichkeit zu Teil wird, könnte sie dann mental überrepräsentiert sein, was dazu führen könnte, dass ihr bei der Vergabe der Rangplätze eine verstärkte Bedeutung zukommt. Diese Möglichkeit soll *Aufmerksamkeitshypothese* heißen.

Eine andere Interpretation stützt sich darauf, dass ein höheres θ zu konservativeren, also vorsichtigeren, Entscheidungen führt: Ist θ größer 1, so führt das dazu, dass von zwei Lotterien mit gleichem Erwartungswert diejenige vorgezogen wird, in der der Gewinn kleiner aber wahrscheinlicher ist. Ein höheres θ bei hoher Ambiguität könnte also Ausdruck einer insgesamt vorsichtigeren Haltung sein. Mangelnde Sicherheit bezüglich der Gewinnwahrscheinlichkeit würde also zu einer vorsichtigeren Haltung bezüglich Risiko führen. Dies soll *Konservatismushypothese* heißen.

Möglichkeiten, zwischen Aufmerksamkeits- und Konservatismushypothese zu differenzieren

Dadurch, dass Experiment II den Verlustbereich mit einbezieht, lässt sich zwischen diesen beiden Erklärungen differenzieren. Während nämlich im Gewinnbereich ein steigendes θ verstärkten Konservatismus anzeigt, verhält es sich im Verlustbereich umgekehrt. θ -Werte größer 1 stehen hier nicht für risikomeidendes sondern für risikosuchendes Verhalten. Ein einfaches Rechenbeispiel macht dies klar. Die Attraktivität der beiden nachfolgenden erwartungswertgleichen Lotterien soll als $\epsilon \cdot P^{\theta}$ bewertet werden.

- (A) Sie verlieren 10,- DM mit einer Wahrscheinlichkeit von .8.
- (B) Sie verlieren 20,- DM mit einer Wahrscheinlichkeit von .4.

Wenn wir für θ beispielsweise 2 einsetzen, so ergibt sich $Attraktivität_A = -10 \cdot .64 = -6.4$ sowie $Attraktivität_B = -20 \cdot .16 = -3.2$. Die Abneigung gegen Lotterie B wäre somit die geringere; gleichzeitig ist dies jedoch auch die risikantere Lotterie, da der hier drohende Verlust größer ist.

Führte steigende Ambiguität bei Verlustaufgaben zu sinkenden θ -Werten, spräche dies also für die Konservatismushypothese, während steigende Werte die Aufmerksamkeitshypothese stützen würden.

Für die Aufmerksamkeitshypothese und die These der Venture-Theorie, dass vermehrte Ambiguität eine stärkere Auseinandersetzung mit der Wahrscheinlichkeitskomponente nach sich zieht, spräche außerdem, wenn sich die Bearbeitungszeiten mit steigender Ambiguität verlängerten. Die Konservatismushypothese legt hingegen keine verlängerten Zeiten nahe.

Schließlich gibt es eine dritte Möglichkeit, zwischen beiden Hypothesen zu unterscheiden. Wenn steigende Ambiguität die Aufmerksamkeit vermehrt auf die Wahrscheinlichkeitskomponente lenkt, so sollte diese besser erinnert werden, so dass eine Gedächtnisabfrage aufschlussreich sein könnte. Leider nimmt mit steigender Ambiguität aber auch die Variationsbreite von P_{Sicht} ab, was die Erinnerbarkeit einschränken sollte. Daher erscheint es wenig sinnvoll, die Erinnerungsaufgabe an die Wahrscheinlichkeitskomponente zu knüpfen. Bleibt also, den Gedächtnistest an ϵ zu knüpfen. Ein solches Unterfangen ist jedoch mit zwei möglichen Mängeln behaftet. Zum einen ist fraglich, ob eine vermehrte Aufmerksamkeit auf die Wahrscheinlichkeitskomponente auch mit einem Aufmerksamkeitsverlust bezüglich ϵ einhergeht. Ist dies nicht der Fall, so ist auch keine Veränderung in der Erinnerungsleistung zu erwarten. Das zweite Problem besteht im sogenannten *Zeigarnik-Effekt* (Zeigarnik, 1927). Damit bezeichnet man das Phänomen, dass aufgabenrelevante Gedächtnisinhalte rasch verloren gehen, wenn die Aufgabe erfüllt ist. Standardbeispiel hierfür ist der Kellner, der keines Schreibblocks bedarf, der sich aber nicht mehr erinnern kann, was an einem Tisch verzehrt worden ist, nachdem dort bezahlt wurde. Ähnlich könnte also im hier verwendeten Paradigma geschehen, dass die Erinnerung an die Lotteriewerte nach Beendigung einer Aufgabe sehr schnell verblasst, so dass ein Floor-Effekt eintrate – auf keinem Ambiguitätsniveau könnten also die ϵ -Werte erinnert werden.

Diesen Einwänden kann man jedoch so Rechnung tragen, dass man die möglichen Ergebnisse eines solchen Gedächtnistests asymmetrisch behandelt. Beide möglichen Schwächen führen nämlich, falls sie tatsächlich relevant wären, in dieselbe Richtung: Trotz Gültigkeit der Aufmerksamkeitshypothese blieben Unterschiede in der Erinnerungsleistung aus. Fehlende Unterschiede sollten somit nicht interpretiert werden; hingegen sprächen Merkleistungen, die mit steigender nachlassen, für die Aufmerksamkeitshypothese.

Experiment II griff die gerade angestellten Überlegungen auf und diente somit zwei Zielen. Zum einen sollte überprüft werden, inwiefern sich der in Experiment I gefundene Zusammenhang zwischen Ambiguität und θ replizieren lässt. Falls sich der Effekt als stabil erweisen sollte, soll erklärt

werden, ob sich dieses Phänomen eher mit der Aufmerksamkeits- oder eher mit der Konservatismushypothese erklären lässt.

10. EXPERIMENT II

Experiment II überprüft, ob sich der zentrale Befund von Experiment I (tendenzieller Anstieg von θ bei steigender Ambiguität) replizieren lässt. Falls dem so ist, können die θ -Daten aus dem hier mituntersuchten Verlustbereich Aufschluss darüber geben, ob dieser Befund eher mit der Aufmerksamkeits- oder der Konservatismushypothese zu erklären ist. Das zentrale Ergebnis ist aus Abb. 35 zu ersehen.

10.1 Materialien und Aufbau

Experiment II folgte dem Paradigma von Experiment I. Im Unterschied zum ersten Experiment gab es hier jedoch auch Verlustaufgaben. Diese folgten demselben Prinzip wie die Aufgaben des Gewinnbereichs: Falls das Glücksrad auf der gelben Fläche zu stehen kam, wurde die Punktzahl der entsprechenden Lotterie mal dem von der Vp gewählten Wertungsfaktor auf ihrem Punktekonto verrechnet, im Verlustfall also abgezogen; andernfalls veränderte sich der Punktestand nicht. Die Verlustaufgaben waren optisch genauso gestaltet wie die Gewinnaufgaben, die Punkteträge der Lotterien waren lediglich mit einem Minuszeichen versehen; zusätzlich stand in der Bildschirmmitte in roter Schrift und Umrandung dreimal „Verlust“.

Der Faktor Ambiguität (wie in Experiment I mit den Ausprägungen 0%, 25%, 55% und 85% verwirklicht) war mit dem Faktor Gewinn/Verlust gekreuzt, so dass ein 4×2 Within-Subjects Design resultierte. Jede Zelle umfasste zehn Aufgaben; diese wurden jeweils geblockt dargeboten. Die Reihenfolge der acht Blöcke war für jede Vp randomisiert. Wie schon in Experiment I wurde jede einzelne Lotterie zufällig und unabhängig entsprechend den unter 4.2 beschriebenen Maßgaben erzeugt.

Innerhalb eines jeden Blocks wurden zwei Gedächtnistests dargeboten. Diese sollten helfen, zwischen Aufmerksamkeits- und Konservatismushypothese zu diskriminieren, falls sich der Befund von Experiment I replizieren ließe (vgl. 9.4). Die Positionen der Tests wurden für jede Vp zufällig bestimmt, beschränkten sich aber auf die ersten acht Aufgaben eines Blockes. Wären nämlich andernfalls die ersten acht Aufgaben eines Blocks ohne Test geblieben, so hätte die Vp wissen können, dass nunmehr zwei Tests folgen, und sie hätte sich auf diese einstellen können; dies sollte verhindert werden.

Der Gedächtnistest sah so aus, dass auf einem neuen Bildschirm die sechs Kästchen in denen zuvor die Lotteriepunktzahlen standen, leer präsentiert wurden; in diese hatten die Vpn die erinnerten Punktzahlen einzutragen. Die Glücksräder wurden hier nicht mehr gezeigt. Der nächste Aufgabenblock wurde erst dann präsentiert, wenn der Gedächtnistest vollständig ausgefüllt war.

Die Zeit, die für jede einzelne Rankingaufgabe benötigt wurde, wurde automatisch erfasst.

10.2 Versuchsteilnehmer und Versuchsablauf

Über Kleinanzeigen in einem lokalen Veranstaltungskalender wurden 42 Frauen und 18 Männer als Vpn gewonnen. Die Teilnahme wurde mit 12,- DM vergütet; darüber hinaus konkurrierten die Teilnehmer um 80,- DM, 60,- DM und 40,- DM, die als Prämie für die drei Erstplazierten ausgelobt waren.

Die Vpn lasen zunächst die Versuchsinstruktion. Im Anhang (A2) ist exemplarisch die Anleitung von Experiment III wiedergegeben; die Instruktion von Experiment II bestand aus den entsprechenden Teilstücken. Nachdem die Vpn die Versuchsbeschreibung gelesen hatten, wurde ihnen die Vorgehensweise durch die Versuchsleiterin am Computer demonstriert und die Aufgabe dabei noch einmal mündlich erläutert. Auf die Gedächtnistests wurde nicht in der schriftlichen sondern nur in der mündlichen Instruktion hingewiesen. Der Grund hierfür bestand darin, dass die Versuchsaufgabe vielen Vpn nur nach alleinigem Durchlesen der Instruktion schwierig und verwirrend vorkam und erst durch die nachfolgende mündliche Erläuterung richtig klar wurde. Um die Vpn nicht zu entmutigen, beschränkte sich daher die schriftliche Instruktion auf die Kerninformationen. Den Vpn wurde gesagt, dass das Ergebnis ihres Gedächtnistests für die Bewertung ihrer Leistung irrelevant sein würde und die Prämien allein gemäß der erspielten Punktestände vergeben werden würden.

Der Versuch wurde in der Regel von vier gleichzeitig in einem Raum arbeitenden Vpn durchlaufen. Ehe sie die eigentlichen Aufgaben bearbeiteten, konnten sie, um sich mit der Bedienung des Programms vertraut zu machen, beliebig viele Übungsaufgaben absolvieren. Diese waren stets Gewinnaufgaben ohne Ambiguität; auch über das Ergebnis dieser Aufgaben erhielten die Vpn keine Rückmeldung.

Über einen ins Programm integrierten Aufgabenzähler konnten sich die Vpn stets orientieren, wieviele Aufgaben noch vor ihnen lagen.

Die freie Wahl des Arbeitstemos führte zu einer durchschnittlichen Bearbeitungszeit von 45 Minuten.

10.3 Ergebnisse

10.3.1 Fehlende Daten und Ausschluss von Daten von der weiteren Analyse

Aufgrund einer ungünstigen Gestaltung der „Versuchsleiterin-Versuchssteuerungs-Schnittstelle“ wurde bei 15 Vpn die Rangordnung des letzten Sechserblocks Lotterien nicht gespeichert. Aus demselben Grund fehlte bei 18 Vpn die Bearbeitungsdauer der letzten fünf Aufgaben sowie die beiden Gedächtnistests aus dem letzten Aufgabenblock. Da die Darbietungsreihenfolge der Blöcke zufällig variierte, verteilten sich die fehlenden Daten unregelmäßig über die Versuchsbedingungen. Aus

diesem Grunde und da nur ein geringer Teil der Gesamtdaten betroffen war (0,3%, 1,9% bzw. 3,8%), war dieser Datenverlust letztlich unerheblich.

Der Ausschluss von Vpn von der weiteren Analyse erfolgte nach demselben Kriterium wie in Experiment I: Für jede Vp wurden also für jede Aufgabe $V(C, \epsilon)$ und $V(C, P)$ bestimmt. Hierbei ist zu beachten, dass $V(C, P)$ sich im Verlustbereich genau anders herum verhält als im Gewinnbereich; orientiert sich die Rangordnung der Lotterien ausschließlich an den wahren Verlustwahrscheinlichkeiten, so ergibt sich hier für $V(C, P)$ ein Wert von 0, während der entsprechende Wert im Gewinnbereich 1 wäre. Daher wurde im Verlustbereich $V(C, P)$ zu $1 - V(C, P)$ transformiert. Der Mittelwert aller $V(C, \epsilon)$ - und $V(C, P)$ -Werte einer Vp wurde als Leistungsmaß herangezogen (vgl. 9.3.1). Dieser Wert zeigte wieder eine stark rechtsschiefe Verteilung, d. h. die meisten Vpn erbrachten einheitlich eine hohe Leistung, während die Ergebnisse weniger Teilnehmer stark abfielen. Die Daten von fünf Vpn mit einem Leistungswert kleiner .55 wurden von der weiteren Analyse ausgeschlossen, da bei diesen davon auszugehen ist, dass sie das Prinzip der Aufgabe nicht verstanden haben und rein willkürlich geantwortet haben.

Bei weiteren Vpn mit normaler Gesamtleistung tauchten für einzelne Versuchsbedingungen extrem niedrige Werte für $V(C, \epsilon)$ oder $V(C, P)$ auf; im Extremfall sanken diese Werte auf .08 ab. Hier sind Vpn also systematisch falsch vorgegangen (zur Erinnerung: wahlloses Verhalten führt zu Korrelationen um .5). Alles spricht dafür, dass die Vpn hier einen Wechsel von Verlust- zu Gewinnaufgaben oder umgekehrt verpasst haben: Erstens war die Gesamtleistung dieser Teilnehmer normal und lag somit deutlich über Zufallsniveau, sie hatten das grundlegende Prinzip also verstanden. Zweitens tauchten solche niedrigen Werte nur auf, wenn von einem Verlust- zu einem Gewinnblock gewechselt wurde oder umgekehrt. Drittens schließlich kamen so niedrige Werte in Experiment I, wo eine solche Verwechslung nicht auftauchen konnte, da hier nur Gewinnaufgaben zu bearbeiten waren, nicht vor; die niedrigste Korrelation für einen kompletten Aufgabenblock lag hier bei .33. Wenn bei einer Vp ein Aufgabenblock für $V(C, \epsilon)$ oder $V(C, P)$ einen kleineren Wert als .3 aufwies, so habe ich dies als Ergebnis einer solchen Verwechslung von Gewinn- und Verlustaufgaben interpretiert. Die Werte von fünf Vpn waren davon betroffen; ein systematischer Zusammenhang zur Versuchsbedingung war nicht zu erkennen. Bei allen nachfolgenden Analysen, in die die mittleren Ergebnisse jeder Vp pro Versuchsbedingung eingingen, habe ich bei diesen fünf Personen in den jeweiligen Blöcken den offensichtlich irrtumsbedingten empirischen Wert durch den Median, den die übrigen Vpn für diese Bedingung aufwiesen, ersetzt.

10.3.2 Die Leistung der Vpn

Die Abb. 31 gibt die Leistungen der Vpn, gemessen in *SFE*, für den Verlust- und den Gewinnbereich wieder. Da die Ergebnisse wiederum stark rechtsschief verteilt waren, sind hier die Mediane abgetragen. Die Leistung, die bei alleiniger Orientierung an den Lotteriepunkten erzielt worden wäre (ϵ -Strategie), ist zur Orientierung mit dargestellt.

Im Gewinnbereich zeigte sich ein Muster, welches dem aus Experiment I sehr ähnlich ist (vgl. Abb. 24): Die Fehlerwerte waren bei fehlender und niedriger Ambiguität geringer als bei hoher und

höchster. Obgleich Experiment II durchweg niedrigere *SFE*-Werte aufwies als Experiment I, deutet nichts auf einen systematischen Unterschied hin: Mittelt man für jede Vp die *SFE*-Werte über alle

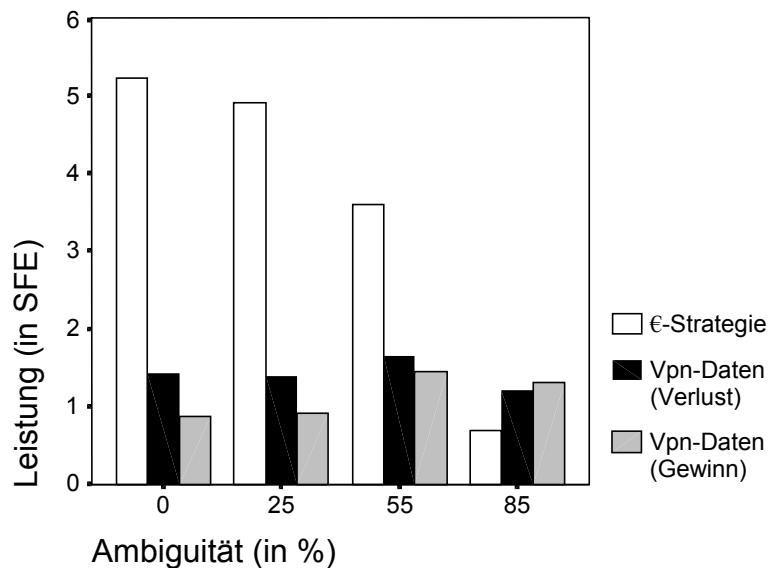


Abb. 31 Experiment II: Leistungsdifferenz der Vpn (und als Vergleichsmaßstab der ϵ -Strategie) zu *Normal*, gemessen in *SFE* [Mediane]. Für die ϵ -Strategie wurden die fast identischen Werte von Gewinn- und Verlustbereich gemittelt. Höhere Werte stehen für schwächere Leistungen.

Ambiguitätsstufen hinweg (in Experiment II natürlich nur im Gewinnbereich) und testet den Unterschied zwischen den so beschriebenen Leistungen in Experiment I und Experiment II mittels eines Wilcoxon Rangsummentests, so ergibt sich hier ein α -Fehler von .26. Auch in einem anderen Punkt glichen sich die Leistungen aus Experiment I und II sehr stark – der Einfluss der Ambiguitätshöhe auf die Leistung erwies sich als quasi identisch: Die Mediane der *SFE*-Werte beider Experimente korrelierten fast perfekt miteinander ($r = .97$; $n = 4$; $p = .03$). Der Leistungsabfall bei höherer Ambiguität war jedoch in Experiment II etwas schwächer ausgeprägt. Unterzieht man in Experiment II allein die *SFE*-Werte des Gewinnbereichs einer Rangvarianzanalyse nach Friedman, so wird der Faktor Ambiguität hier auch nicht signifikant ($\chi^2 = 3,7$; $n = 55$; $p = .30$). Vor dem Hintergrund der Ergebnisse von Experiment I erscheint es mir jedoch durchaus sinnvoll, auch hier einen entsprechenden Einfluss höherer Ambiguität zu vermuten.

Im Verlustbereich war kein entsprechender Trend erkennbar, die *SFE*-Werte waren auf allen Ambiguitätsniveaus etwa gleich hoch. Eine Rangvarianzanalyse nach Friedman, die allein für die Verlustaufgaben durchgeführt wurde, zeigte demzufolge keinen Effekt der Variable Ambiguität auf. Die Leistungen der Vpn waren im Verlustbereich insgesamt schwächer als im Gewinnbereich. Bildet man für jede Vp einen *SFE*-Gesamtwert für den Verlust- und einen für den Gewinnbereich, so liegen deren Mediane bei 1,70 und 1,31. Dieser Unterschied erwies sich als statistisch wenig stabil ($p = .19$; Wilcoxon Rangsummentest). Einzelvergleiche werden aber auf den Ambiguitätsstufen 0%, 25% und 55% signifikant (α -Niveau von .05; Verfahren für multiple Einzelvergleiche nach Conover). Einen Überblick über sämtliche Einzelvergleiche gibt nachfolgende Tabelle.

Tabelle 3: Signifikante Leistungsunterschiede in Experiment II ($p \leq .05$).

	0% (-)	25% (-)	55% (-)	85% (-)	0% (+)	25% (+)	55% (+)
25% (-)							
55% (-)							
85% (-)	x		x				
0% (+)	x	x	x				
25% (+)	x	x	x				
55% (+)			x		x		
85% (+)			x		x		

Was sich schon in Experiment I gezeigt hatte, war auch hier sowohl im Gewinn- als auch im Verlustbereich zu finden: Bis zu einer Ambiguität von 55% waren die Vpn deutlich besser als die simple ϵ -Strategie, bei höchster Ambiguität sank die Leistung jedoch unter das von dieser erreichte Level. Letztgenannter Unterschied war für beide Aufgabenbereiche statistisch signifikant (Verlust: $p = .02$; Gewinn: $p < .001$; beides Wilcoxon Rangsummentest). Insgesamt erwiesen sich aber auch die Vpn von Experiment II als erfolgreiche Maximierer.

10.3.3 Die Identifikation eines geeigneten Basismodells

Um festzustellen, welches Basismodell am besten zur Schätzung von θ geeignet ist, wurde genauso verfahren wie bereits in Experiment I: Für jeweils drei Varianten aller acht Basismodelle (vgl. 8.2) wurde derjenige θ -Wert bestimmt, der die beste Replikation des Vpn-Verhaltens erlaubte. Für jedes der 24 so optimierten Modelle ergab sich dann wieder ein Wert Δ , der angab, wie gut das Modell die von den Vpn gebildeten Rangreihen wiedergeben konnte (vgl. 9.3.3).

Jedoch mussten zuvor die meisten der Basismodelle für den Verlustbereich verändert werden. Es gibt zwei Möglichkeiten, die Strategie *Konservativ* auf Verlustaufgaben zu übertragen. Zum einen könnte die Attraktivität einer Lotterie genau wie im Gewinnbereich als $\epsilon \cdot P_{\text{Sicht}}$ berechnet werden. Die Wahrscheinlichkeit würde hier also nach dem Prinzip „Ich kümmere mich nur um das, wovon ich weiß, der Rest interessiert mich nicht“ behandelt. Die andere Übertragungsmöglichkeit besteht darin, die Attraktivität im Verlustbereich als $\epsilon \cdot (P_{\text{Sicht}} + \text{Ambig})$ zu berechnen, also wie auch im Gewinnbereich stets von der schlechtesten Möglichkeit auszugehen, was im Verlustbereich bedeutet anzunehmen, dass unter der Verdeckungsfläche ausschließlich Verlustfläche liegt. Um die Zahl der Basismodelle nicht noch weiter zu inflationieren, habe ich mich auf eine der beiden Möglichkeiten beschränkt, nämlich die letztgenannte. Keine der beiden Übertragungsmöglichkeiten erschien mir plausibler, so dass die Wahl willkürlich auf die pessimistische Variante fiel; dies erscheint mir aber insofern unproblematisch, als dass die Ergebnisse aus Experiment I nahelegen, dass *Konservativ*, unabhängig davon, welche Spielart man für den Verlustbereich wählt, kein besonders geeignetes Modell sein würde, um das Versuchspersonenverhalten nachzuzeichnen.

Auch sämtliche *Venture* Basismodelle bedurften einer Anpassung für den Verlustbereich. Für die Modelle *Venture*_{I-IV} ergaben diese sich aus den von Hogarth und Einhorn (1995) vorgeschlagenen Funktionen für den Verlustbereich (vgl. Abb. 6). Um die Idee von *Venture*_{einfach}, dass höhere Lotteriebeträge zu vorsichtigeren Einschätzungen der Lotteriewahrscheinlichkeiten führen, äquivalent auf den Verlustbereich zu übertragen, erhielt dieses Basismodell hier die Form

$$\text{Venture}_{\text{einfach}} \text{ (Verlustbereich): } \text{Attraktivität} = \epsilon \cdot (P_{\text{Sicht}} + (40 - V/2) \cdot Ambig / 100).$$

Mindestens 40% der Verdeckungsfläche werden also in diesem Basismodell der Verlustfläche zugeschlagen, bei höheren potenziellen Verlusten mehr.

Wie schon in Experiment I galt auch in Experiment II für alle acht Basismodelle, dass diejenige Variante, die einen linearen Zusammenhang zwischen ϵ und $u(\epsilon)$ annimmt, am besten abschnitt; deswegen wurden nur diese Varianten näher betrachtet.

Die Δ -Werte verteilten sich wiederum deutlich linksschief. Das Muster, das die Replikationsgüte der acht Basismodelle zeichnete, war dem aus Experiment I äußerst ähnlich; berücksichtigt man bei Experiment II nur den Gewinnbereich, so korrelieren die Mediane der Δ -Werte der acht Basismodelle fast perfekt mit denen aus Experiment I ($r = .97$; $n = 4$, $p = .035$).

Wie aus Abb. 32, die die Mediane der Δ -Werte für alle acht Basismodelle zeigt, zu ersehen ist, schnitt *Normal* mit einem Median von .371 auch in Experiment II am besten ab. Zweitbestes Modell mit einem Median von .383 war wieder *Venture*_I, das mit *Normal* eng verwandt ist. Der geringe Unterschied zwischen diesen beiden Modellen erwies sich (zumindest bei einseitiger Testung, die mir gerechtfertigt erscheint, da bereits ein entsprechendes Ergebnis aus Experiment I vorliegt) auch hier als statistisch signifikant ($p = .05$; Wilcoxon-Rangsummentest).

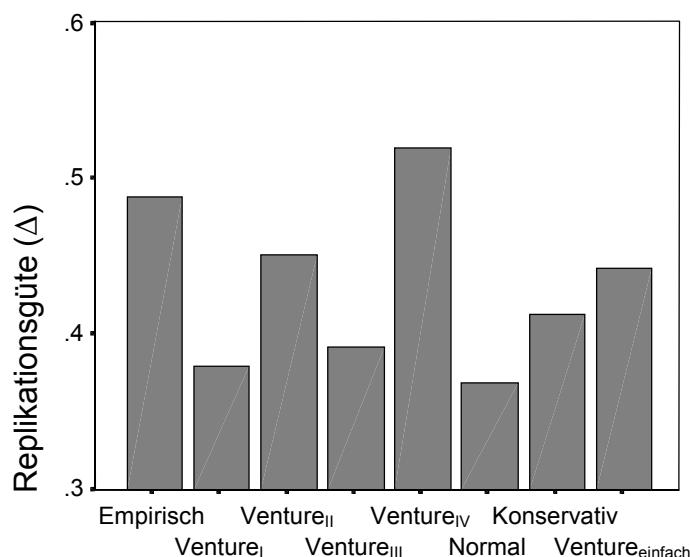


Abb. 32 Experiment II: Für die acht Basismodelle ist hier angegeben, wie gut sie das Verhalten der Vpn nachbilden konnten [Mediane von Δ].

Unter drei der acht Versuchsbedingungen schnitt *Normal* nicht am besten ab. In keinem dieser Fälle erwies sich jedoch die Überlegenheit des anderen Modells als statistisch signifikant.¹²

Der Vorsprung von *Normal* gegenüber den anderen Basismodellen wird wiederum deutlicher, wenn wir uns anschauen, für wie viele Vpn jedes Modell die beste Replikation des tatsächlich gezeigten Verhaltens lieferte. Abb. 33 zeigt diese Daten. Zu bedenken ist auch hier, dass *Normal* und *Venture_I* ähnliche Schätzer für θ lieferten. Die Korrelationen variierten je nach Bedingung zwischen $Q = .89$ (Gewinnbereich, 85% Ambiguität) und $Q = .97$ (Verlustbereich, 25% Ambiguität). Bei 56% der Vpn eignete sich eines dieser beiden Basismodelle am besten, ihr Verhalten zu replizieren.

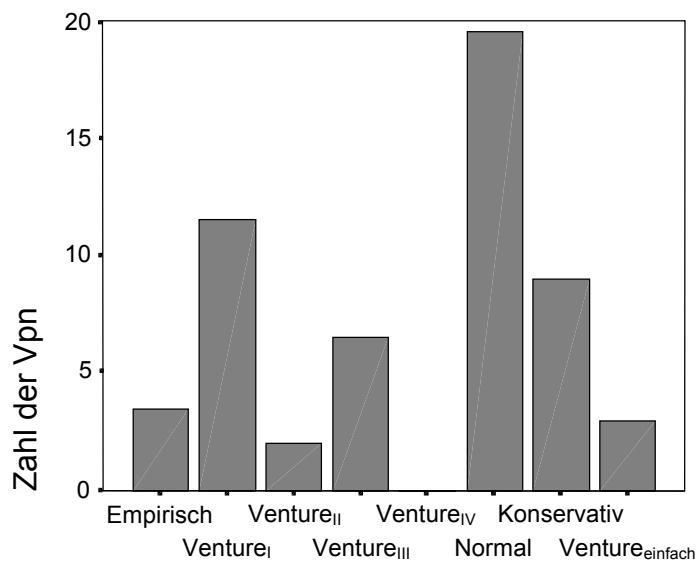


Abb. 33 Experiment II: Für jedes Basismodell ist die Zahl der Vpn abgetragen, deren Verhalten am besten durch dieses Modell erklärt werden konnte.

Wie gut die Replikation des Vpn-Verhaltens für die einzelnen Versuchsbedingungen gelang, ist Abb. 34 zu entnehmen. Der Modellfit war im Gewinnbereich generell besser als bei den Verlustaufgaben. Bildet man für jede Vp separat den Δ -Wert für den Gewinn- und den Verlustbereich, mittelt also über die Ambiguitätsniveaus, so ergibt sich für den Gewinnbereich ein Median von .34 und für die Verlustaufgaben ein Median von .40. Dieser Unterschied ist statistisch signifikant ($p = .02$; Wilcoxon-Rangsummentest). Der Verlauf für die Variable Ambiguität sah im Gewinnbereich deutlich anders aus als bei den Verlustaufgaben; daher ist eine getrennte Betrachtung sinnvoll. Im Verlustbereich ist kein Zusammenhang zwischen dem Ambiguitätsniveau und der Reproduzierbarkeit der Daten zu erkennen. Das Muster im Gewinnbereich glich hingegen demjenigen, das sich schon in Experiment I fand: Bei hoher und höchster Ambiguität lag der Reproduktionsfehler rund 40% höher als bei fehlender und niedriger Ambiguität. Dieser Effekt erwies sich jedoch als nicht so stabil wie im ersten Experiment: Eine Rangvarianzanalyse nach Friedman wies hier kein signifikantes Ergebnis auf ($p = .16$).

¹² D.h. alle α -Fehler waren größer als ,05. Getestet wurde dabei mittels Wilcoxon-Rangsummentest; der Inflationierung des α -Fehlers wurde dabei also noch nicht einmal Rechnung getragen.

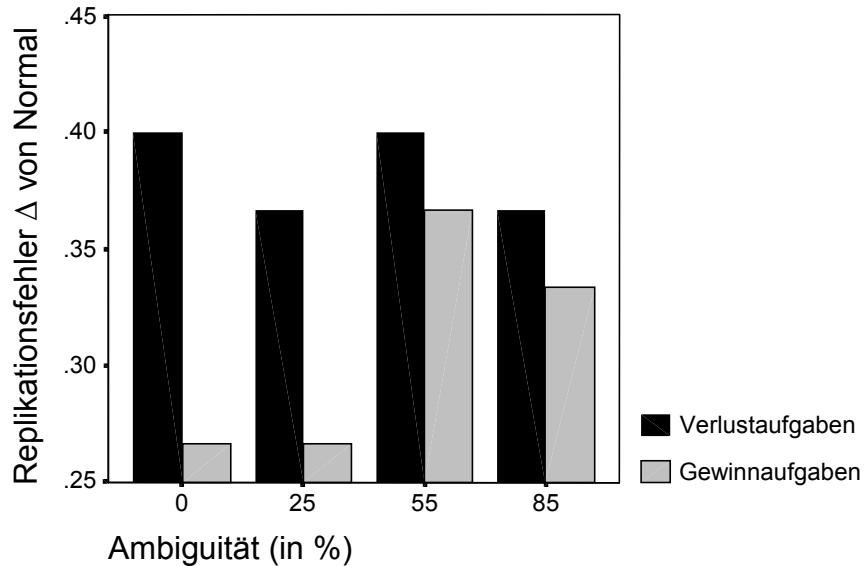


Abb. 34 Experiment II: Güte der Replikation des Verhaltens der Vpn auf der Grundlage von *Normal* [Mediane]. Höhere Werte stehen für eine geringere Güte.

In Experiment I hatte sich ein enger Zusammenhang zwischen der Leistung und der Reproduzierbarkeit des Verhaltens ergeben. Dieser fand sich auch in Experiment II, sowohl auf globaler wie auch auf individueller Ebene. So korrelierten zum einen die Mediane, die sich in den acht Versuchsbedingungen für die Leistung (*SFE*) ergeben haben, hoch mit den entsprechenden Medianen von Δ ($r = .93; n = 8; p = .001$). Stellt man zum anderen einen Vergleich auf individueller Ebene an und korreliert die Gesamtleistung der Vpn (wiederum in *SFE* gemessen) mit der Reproduzierbarkeit ihres Verhaltens (ebenfalls über alle Versuchsbedingungen gemittelt), so zeigte sich auch hier ein enger Zusammenhang ($r = .84; n = 55; p < .001$).

Wie schon in Experiment I steht auch hier zu vermuten, dass hinter der Leistung wie hinter der Reproduzierbarkeit des Verhaltens eine dritte Variable steht, die diese beiden gleichsinnig beeinflusste. Nämlich inwieweit die Vpn konsequent einer Strategie nachgegangen sind, oder, in anderen Worten, in welchem Grade die Rangreihen willkürlich vergeben worden sind.

Das Verhalten der Vpn ließ sich auf der Grundlage von *Normal* mit insgesamt recht hoher Genauigkeit wiedergeben. Für die mittlere Vp betrug $\Delta = .37$; im Mittel ergab sich also pro Sechserblock wenig mehr als eine einmalige Vertauschung zweier benachbarter Rangplätze. Im positiven Bereich lag die Reproduktionsgüte deutlich über derjenigen von Experiment I (Mediane: $\Delta = .46$ vs. $\Delta = .34$); der Unterschied ist auch statistisch bedeutsam ($p = .009$; U-Test). Um die Güte der Schätzwerte für θ , die das Basismodell *Normal* ergibt, besser einordnen zu können, wurde θ wieder separat auf Grundlage der Aufgaben mit geraden und ungeraden Nummern geschätzt. Die Korrelationen dieser Splithalf-Schätzer sind in Tabelle 4 wiedergegeben.

Die Korrelationen haben etwa dieselbe Höhe wie diejenigen, die sich für Experiment I ergeben hatten, obwohl die einzelnen Schätzungen von θ hier nur auf fünf Aufgaben beruhen im Gegensatz

Tabelle 4: Rangkorrelationen (ρ) der auf Grund von *Normal* gewonnenen Splithalf-Schätzer für θ sowie die nach Spearman-Brown berechneten Konsistenzschätzungen (in Klammern).

	Ambiguität 0%	Ambiguität 25%	Ambiguität 55%	Ambiguität 85%
Verlustaufgaben	.67 (.80)	.82 (.90)	.58 (.73)	.67 (.80)
Gewinnaufgaben	.65 (.79)	.65 (.79)	.65 (.79)	.75 (.86)

zu sieben bzw. acht Aufgaben in Experiment I. Die Qualität der Schätzer für θ betrachte ich als zufriedenstellend; sie kann (insbesondere vor dem Hintergrund niedrigerer Δ -Werte) als etwas höher angesehen werden als diejenige für die Schätzer aus Experiment I.

10.3.4 Die relative Gewichtung der Wahrscheinlichkeit (θ)

Abbildung 35 gibt die Mediane der Schätzer für θ wieder, die auf der Grundlage von *Normal* gewonnen wurden.

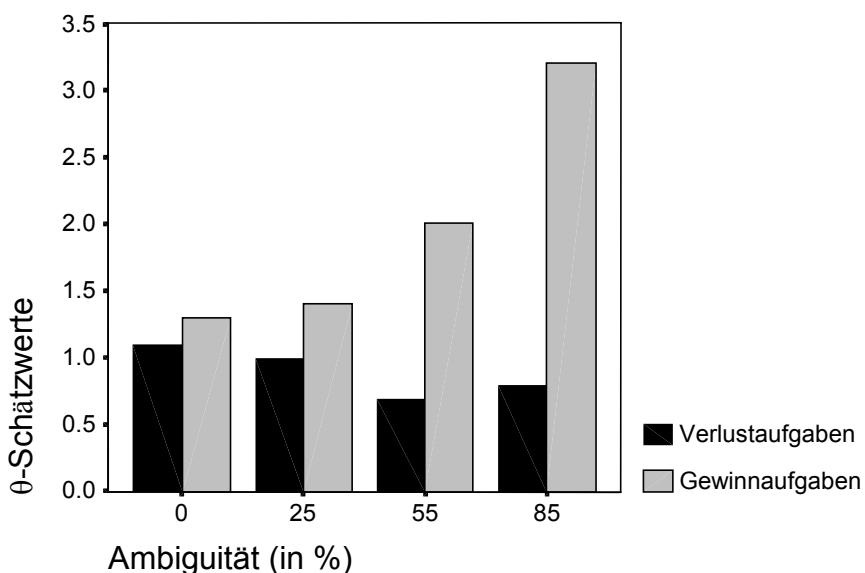


Abb. 35 Experiment II: Die Schätzwerte für θ , gewonnen auf der Grundlage von *Normal* [Mediane].

Wie zu sehen ist, unterschieden sich die Verläufe zwischen Gewinn- und Verlustbereich deutlich. Während θ bei den Gewinnaufgaben mit zunehmender Ambiguität kontinuierlich und deutlich stieg, ist im Verlustbereich, zumindest der Tendenz nach, ein Absinken von θ zu erkennen. Von daher ist es sinnvoll, den Einfluss der Ambiguität auf das Entscheidungsgewicht für den positiven und negativen Bereich gesondert zu betrachten. Im Verlustbereich weist der Friedman-Test einen moderaten aber stabilen Effekt auf ($\chi^2 = 10,7$; $df = 3$; $p = .013$). Einzelvergleiche nach dem Verfahren von Conover zeigen, dass alle Unterschiede mit Ausnahme von 25% vs. 85% auf dem 5%-Niveau signifikant sind. Im positiven Bereich zeigte sich ein stärkerer Einfluss der Ambiguität auf das Entscheidungs-

gewicht; hier erreicht der χ^2 -Wert 21,9 (Friedman-Test, $df = 3$; $p < .001$). Die Einzelvergleiche zeigen, dass alle Unterschiede auf dem 5%-Niveau signifikant sind. Der Effekt, dass höhere Ambiguität einen kontinuierlichen Anstieg der Ambiguität nach sich zog, kann also als stabil angesehen werden.

Im Gewinnbereich ergab sich bei fehlender Ambiguität für θ ein Median größer 1; die Vpn zeigten also insgesamt wieder ein eher risikomeidendes Verhalten (bei 62% der Teilnehmer lag der Schätzwert für θ über 1), was in Gewinnsituationen typisch ist (z. B. Tversky & Kahneman, 1981).

Im Verlustbereich betrug der Median der θ -Schätzer bei fehlender Ambiguität 1,1; es zeigte sich hier also eine leichte Tendenz zu risikosuchendem Verhalten. Dass die θ -Werte mit steigender Ambiguität sanken, zeigt jedoch, dass die Vpn hier zunehmend konservativer handelten. Dies spricht deutlich für die Konservatismus- und gegen die Aufmerksamkeitshypothese.

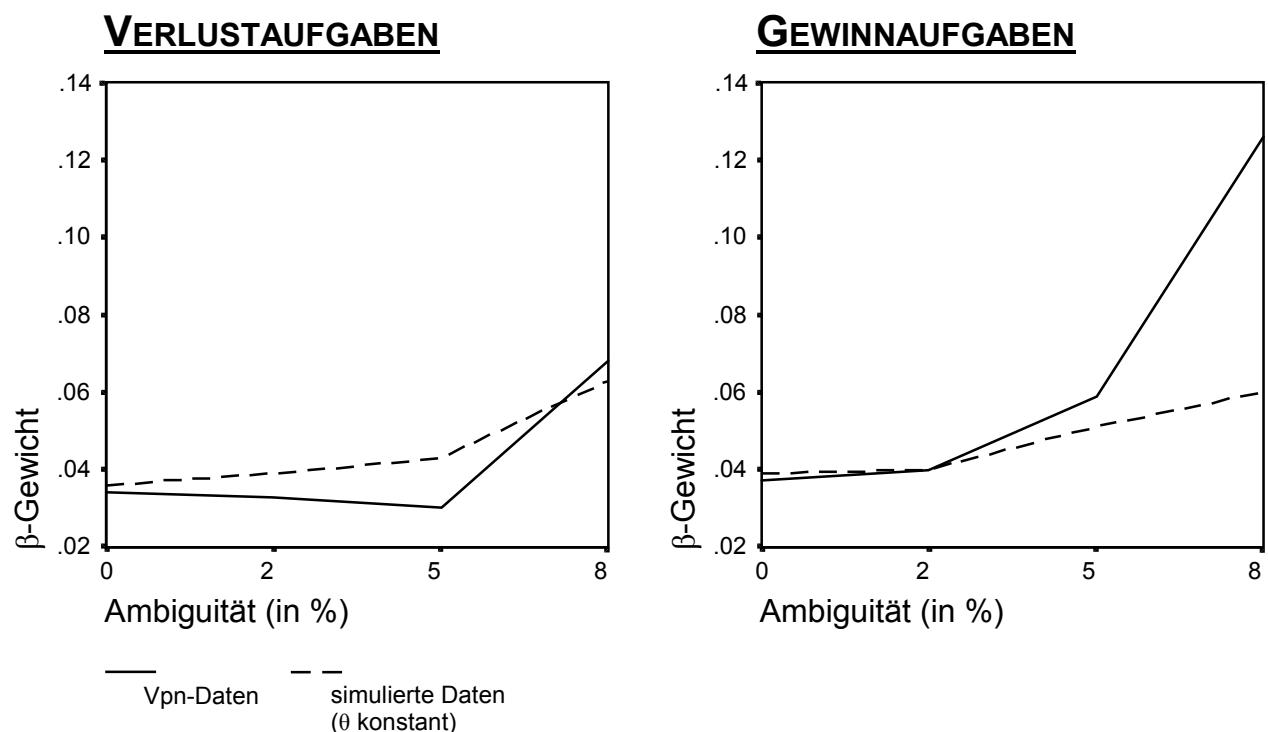


Abb. 36 Experiment II: Standardisierte β -Gewichte der Regression von P_{Sicht} auf die Attraktivität der Lotterien für empirische Daten und Daten der Simulation V. Da in der Simulation θ innerhalb des Gewinn- und Verlustbereiches konstant blieb, zeigen diese Linien, was passiert, wenn *kein* Strategiewechsel vollzogen wird.

10.3.5 Analyse des Einflusses von P_{Sicht} auf die Attraktivität der Lotterien

Auch für Experiment II soll das Bild, das aufgrund der Schätzwerte für θ entstanden ist, durch die zweite Analyseform abgesichert werden (vgl. 9.3.5). Abb. 36 zeigt dazu für die einzelnen Versuchsbedingungen, wie stark dort der Einfluss der sichtbaren Gewinnwahrscheinlichkeit auf die Attraktivität der Lotterien gewesen ist. Die Stärke dieses Einflusses wird dabei wieder über das β -Gewicht beschrieben, das sich für die Vorhersage der Attraktivität aus P_{Sicht} ergab. Prinzipiell gilt dabei, dass

höhere Werte einen stärkeren Einfluss bezeichnen, denn bei höheren β -Gewichten zog dieselbe Veränderung von P_{Sicht} eine stärkere Veränderung der Attraktivitätsbewertung nach sich. Für sich genommen sind diese β -Gewichte aber nicht aussagekräftig (vgl. 9.3.5); statt dessen ist es notwendig die Gewichte, die sich für die empirischen Daten ergeben haben, mit denjenigen zu vergleichen, die sich ergeben, wenn mit steigender Ambiguität *kein* Strategiewechsel vollzogen wird. Diese Vergleichsgewichte wurden mit Hilfe der Simulation VI ermittelt.

Simulation VI

Eine künstliche Vp bearbeitete alle Aufgaben, die in Experiment II verwendet worden waren. Dabei ging sie grundsätzlich nach der Strategie *Normal* vor; diese war gewählt worden, da dieses Basismodell am ehesten geeignet war, das Verhalten der Vpn nachzubilden (vgl. 10.3.3). Im Verlustbereich wurde dabei ein θ von 1,1 und im Gewinnbereich eines von 1,3 gewählt, da dies den Medianen der θ -Schätzer für die Experimentaldaten bei fehlender Ambiguität entsprach. Nachdem die „Vp“ der Simulation dergestalt alle Aufgaben bearbeitet hatte, wurden auch für diese Daten für jede der acht Versuchsbedingungen das β -Gewicht der Regression von P_{Sicht} auf die Attraktivität der Lotterien bestimmt. Im Verlustbereich ergaben sich dabei eigentlich negative β -Gewichte; denn eine Vergrößerung der Verlustfläche führte natürlich dazu, dass die Attraktivität der Lotterien sank und nicht stieg. Um den Vergleich mit dem Gewinnbereich und mit den θ -Schätzern zu erleichtern, wurden diese jedoch mit -1 multipliziert.

Die Abb. 36 stellt die β -Gewichte für die Experimentaldaten den aus der Simulation gewonnenen gegenüber. Im Gewinnbereich zeigte sich dasselbe Bild, das bereits die θ -Schätzer ergeben haben: Mit zunehmender Ambiguität wurde die Differenz aus empirischen und simulierten Werten immer größer; dies bedeutet, dass der Wahrscheinlichkeitskomponente mit steigender Ambiguität immer mehr Bedeutung zukam, so dass die Entscheidungen zunehmend konservativer ausfielen.

Im Verlustbereich stimmte das Ergebnis zum größten Teil mit den θ -Schätzern überein. Bis zu einer Ambiguität von 55% nahm der Einfluss der Verlustwahrscheinlichkeit kontinuierlich ab, um bei 85% Ambiguität wieder anzusteigen. Dieser Anstieg fiel jedoch, verglichen mit den θ -Werten, etwas zu steil aus. Während der θ -Wert hier unterhalb desjenigen für fehlende und 25% Ambiguität geblieben ist (vgl. Abb. 35), deutet die Analyse der β -Gewichte darauf hin, dass der Einfluss der Verlustwahrscheinlichkeit hier höher gewesen ist als bei 0% und 25% Ambiguität. Allerdings waren bei 85%-Ambiguität die Standardfehler für die geschätzten β -Gewichte so hoch, dass man nicht davon ausgehen muss, dass der empirische Wert tatsächlich *über* dem simulierten liegt. Ein ausgeprägter Widerspruch zu den Schätzwerten für θ liegt somit also in diesem Punkt nicht vor. Insgesamt stimmen die Ergebnisse dieser Analyse also wieder mit dem Bild, das die θ -Schätzwerte ergeben hatten, sehr gut überein.

10.3.6 Bearbeitungszeiten

Bei der Analyse von Bearbeitungszeiten tritt das Problem auf, welche längeren Zeiten als Ausreißer angesehen werden sollen; ob also eine längere Verweildauer bei einer Aufgabe eher auf angestrenгtes Nachdenken oder auf einen entspannenden Blick aus dem Fenster zurückzuführen ist. Um dieses Problem der Ausreißerselektion zu umschiffen, gingen in die Analyse der Bearbeitungszeiten für jede Vp die Mediane der zehn Bearbeitungszeiten in jeder Versuchsbedingung ein, also acht Werte pro Vp.

In allen Bedingungen zeigten sich die Zeiten als nur geringfügig linksschief verteilt und von ähnlicher Varianz. Einer parametrischen Auswertung steht somit nichts im Wege. Abb. 37 zeigt die Ergebnisse. Wie zu sehen ist, wurde im Gewinnbereich durchweg schneller gearbeitet als im Verlustbereich. Der Unterschied betrug im Mittel 4,6s pro Aufgabe, d. h. hier wurde 16% weniger Zeit benötigt als bei den Verlustaufgaben.

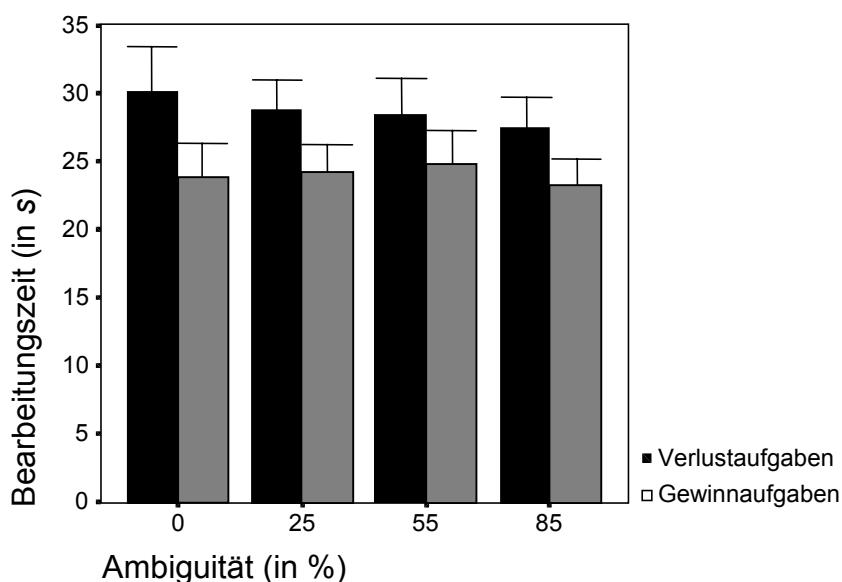


Abb. 37 Experiment II: Durchschnittliche Bearbeitungszeiten.
Die Winker markieren ± 2 Standardfehler.

Eine zweifaktorielle Varianzanalyse für Messwiederholungen zeigte, dass allein der Einfluss des Faktors „Gewinn/Verlust“ signifikant war ($F = 44,4; df = 1; p < .001$). Für den Faktor „Ambiguität“ war kein Einfluss erkennbar ($F = 1,4; df = 3; p = .27$), ebenso wenig für die Interaktion „Gewinn/Verlust \times Ambiguität“ ($F = .6; df = 3; p = .65$).

Dass Entscheidungen im Verlustbereich länger brauchen, ist durchaus typisch (z. B. Budescu, Weinberg & Wallsten, 1988). Da ein Ansteigen der Bearbeitungszeiten nicht einmal der Tendenz nach zu erkennen ist, spricht auch dieser Befund für die Konservatismus- und gegen die Aufmerksamkeitshypothese. Gleichzeitig konnte das Postulat der Venture-Theorie, dass höhere Ambiguität zu längeren Bearbeitungszeiten führen sollte, nicht bestätigt werden.

Für die Aufmerksamkeitshypothese würde ein positiver Zusammenhang zwischen der Bearbeitungszeit und der Höhe von θ sprechen. Dieser ließe sich nämlich so interpretieren, dass die bei einzelnen Vpn längere Beschäftigung mit der Disambiguierung die Aufmerksamkeit vermehrt auf die Wahrscheinlichkeitskomponente gelenkt und dieser darüber zu einem höheren Gewicht bei der Entscheidung verholfen hätte. Lediglich sechs der acht Versuchsbedingungen sind hierbei relevant, nämlich alle die, bei denen Ambiguität vorlag. Bei fünf dieser sechs Bedingungen lag die Korrelation zwischen Bearbeitungsdauer und θ nahe null. Im Verlustbereich zeigte sich bei 25% jedoch ein deutlicher Zusammenhang ($r = .57; p < .001$). Da aus den restlichen Daten jedoch kein Hinweis darauf zu finden ist, dass unter dieser Versuchsbedingung grundsätzlich andere Prozesse zum Tragen gekommen wären als unter den übrigen, und da die Korrelationen unter den anderen fünf Bedingungen nicht einmal eine Tendenz zu einem positiven Zusammenhang aufwiesen, sehe ich darin ein Zufallsergebnis. Die Aufmerksamkeitshypothese wird also wiederum nicht unterstützt.

Ein Zusammenhang dergestalt, dass langsamer arbeitende Teilnehmer bessere Leistungen erzielt hätten, fand sich nicht ($q = -.04, n = 55; p = .77$). Da die Unterschiede in den Bearbeitungszeiten beträchtlich waren (im neunzigsten Perzentil war die Bearbeitungszeit doppelt so hoch wie im zehnten) ist davon auszugehen, dass diese Differenzen zumindest teilweise Unterschiede im investierten kognitiven Aufwand widerspiegeln; ein höherer Aufwand kam offensichtlich der Leistung nicht zu gute. Ein deutlicher Zusammenhang zwischen Bearbeitungszeit und Leistung fand sich aber auf Ebene der Versuchsbedingungen: In den Bedingungen, die zu längeren Bearbeitungszeiten führten, wurden die eher schwächeren Leistungen erbracht ($r = .73; n = 8; p = .042$). Ein mäßiger Zusammenhang fand sich auch dahingehend, dass das Verhalten hier weniger gut replizierbar war ($r = .51; n = 8; p = .20$). Schwierigere, d. h. vor allem Verlustaufgaben, führten also zu weniger konsistentem Verhalten, zu schlechteren Leistungen und zu längeren Bearbeitungszeiten.

10.3.7 Gedächtnistest

Als Fehlerwert benutzt wurden hier die absoluten Differenzen zwischen tatsächlichen und erinnernten Lotteriewerten. Jede Vp hatte unter jeder Bedingung zwei Sechserblöcke als Gedächtnistest bearbeitet. Für jede Vp wurden die zwölf Fehlerwerte einer jeden Bedingung gemittelt; die nachfolgenden Analysen basieren auf diesen Werten.

Die Fehlerwerte erwiesen sich unter allen Versuchsbedingungen als näherungsweise normalverteilt und varianzhomogen. Abb. 38 gibt die mittleren Fehler wieder. Ein Effekt von Ambiguität ist nicht zu erkennen; ob es sich um Verlust- oder Gewinnaufgaben handelte, hatte offensichtlich ebenfalls keinen Einfluss auf die Gedächtnisleistung. Der Gesamtmittelwert lag bei 18,6. Der Fehler, den eine zufällige Strategie (wahllose Produktion von Werten zwischen 3 und 97) produzieren würde, liegt etwa bei 31,7 (Ergebnis einer Monte-Carlo Simulation). Die Gedächtnisleistungen sind insgesamt also als eher mäßig anzusehen.

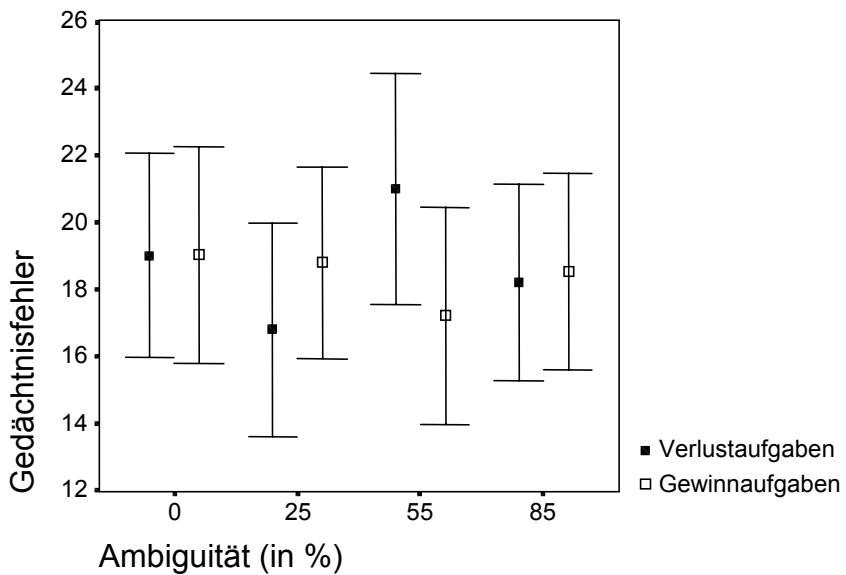


Abb. 38 Experiment II: Durchschnittliche Größe der Gedächtnisfehler bei der Erinnerung an die Lotteriewerte. Zufallsbestimmtes Verhalten hätte zu einer mittleren Größe von 31,7 geführt. Die Winker markieren ± 2 Standardfehler.

Da relativ viele Gedächtnistestdaten verloren gegangen sind (vgl. 10.3.1), habe ich hier darauf verzichtet, fehlende Werte durch den Stichprobenmedian zu ersetzen. Demzufolge konnten in die varianzanalytische Auswertung der Gedächtnisdaten lediglich die Werte von 34 Vpn einfließen. Da die Ausfälle nicht durch das Verhalten der Vpn oder durch irgendein Vpn-Charakteristikum verursacht worden sondern zufällig zustande gekommen waren, ist jedoch nicht davon auszugehen, dass sich die analysierte Untergruppe von der Gesamtstichprobe unterscheidet.

Im vorliegenden Kontext ist insbesondere interessant, inwieweit der Grad der Ambiguität einen Einfluss auf die Gedächtnisleistung hatte. Der entsprechende *F*-Wert fällt mit .3 extrem klein aus ($p = .86$). Der Faktor Gewinn/Verlust sowie die Interaktion Ambiguität \times Gewinn/Verlust erwiesen sich ebenfalls als völlig bedeutungslos. Aus den bereits diskutierten Gründen sollte ein solches Ergebnis, das keinen Einfluss des Grades der Ambiguität auf die Gedächtnisleistung zeigt, nicht weiter diskutiert werden (vgl. 9.4).

10.3.8 Wodurch wurde die Leistung der Vpn determiniert?

In einer post-hoc Analyse wurde wiederum untersucht, inwieweit sich Leistungsunterschiede zwischen den Vpn erklären lassen; alle Analysen beziehen sich wieder auf Rangdaten, die Variablen sind dem entsprechend durch ein r indiziert. Es zeigte sich, dass Bearbeitungszeit und Leistung unter allen Bedingungen unabhängig voneinander waren. Hingegen zeigten wiederum θ_r und vor allem Δ_r einen Einfluss. Für jede der acht Versuchsbedingungen wurde ein eigenes Regressionsmodell erstellt. In allen Fällen, in denen neben Δ_r auch θ_r einen signifikanten Beitrag zur Erklärung der SFE-Leistung beitrug, waren diese beiden Variablen unabhängig voneinander. Für die Variablen, die in

das jeweilige Regressionsmodell eingegangen sind, gibt nachfolgende Tabelle die Korrelationen mit der *SFE*-Leistung wieder.

Tabelle 5: Experiment II – Korrelationen von Δ_r und θ_r mit *SFE_r*. Δ_r und θ_r sind unter allen Bedingungen unkorreliert.

		0%	25%	55%	85%
Verlust	Δ_r	.66	.72	.43	.62
	θ_r	n.s.	n.s.	-.45	n.s.
Gewinn	Δ_r	.61	.38	.63	.45
	θ_r	.33	.32	n.s.	n.s.

Im Gewinnbereich galt dabei Ähnliches wie in Experiment I: θ konnte nur dann einen Effekt entwickeln, wenn das Verhalten in hohem Maße konsistent war, also bei fehlender und niedriger Ambiguität (vgl. Abb. 34). Dass der Zusammenhang zwischen Leistung und θ hier deutlich niedriger ausfiel als dies in Experiment I der Fall war, könnte daran liegen, dass dort die interindividuelle Variabilität in den θ -Schätzern erheblich höher war (in Experiment I lag die mittlere Standardabweichung bei 9,1, im Gewinnbereich von Experiment II lediglich bei 3,9).

Im Verlustbereich zeigte sich durchweg ein mittelstarker Zusammenhang zwischen Leistung und Reproduktionsgüte. θ spielte lediglich bei 55% Ambiguität eine moderate Rolle. Offensichtlich tendierten die Vpn hier dazu, die Wahrscheinlichkeitskomponente schwächer zu gewichten als ihren Leistungen gut tat; geringere θ -Werte gingen hier tendenziell mit höheren *SFE*-Werten einher. Dies war auch die Bedingung, in der insgesamt die niedrigsten θ -Werte zu finden waren (vgl. Abb. 35).

10.4 Fazit und weiteres Vorgehen

Erneut erwiesen sich die Vpn als erfolgreiche Maximierer, dabei zeigten sie im Gewinnbereich etwas bessere Leistungen als bei den Verlustaufgaben. Wie schon in Experiment I waren ihre Ergebnisse bis zu einer Ambiguität von 55% deutlich besser als die der ϵ -Strategie. Bei einer Ambiguität von 85% sank die Leistung jedoch erneut unter diesen Vergleichsmaßstab.

Erneut konnte das Vpn-Verhalten generell dann am besten wiedergegeben werden, wenn ein linearer Zusammenhang zwischen ϵ und $u(\epsilon)$ angenommen wurde. Von den acht Basismodellen erwies sich wiederum *Normal* als eindeutig bestes. Das Vpn-Verhalten ließ sich insgesamt besser replizieren, als dies im ersten Experiment der Fall war. Von daher erscheint es mir angemessen, den Ergebnissen des zweiten Experiments im Zweifelsfall ein stärkeres Gewicht beizumessen als denen des ersten.

Im Gewinnbereich zeigte sich mit steigender Ambiguität ein kontinuierliches Ansteigen des Entscheidungsgewichtes; der Wahrscheinlichkeitskomponente kam bei der Entscheidung also ein immer stärkeres Gewicht zu. In Experiment I hatte sich ein eher sprunghafter Verlauf gezeigt. Somit stehen die Ergebnisse nicht nur erneut im Gegensatz zu den Befunden von Gonzalez-Vallejo et al.

(1994), die zur Formulierung der Gewichtungshypothese geführt hatten sondern auch im klaren Gegensatz zur Venture-Theorie.

Im Verlustbereich zeigte sich die entgegengesetzte Tendenz; mit steigender Ambiguität sank das Entscheidungsgewicht hier tendenziell. Auch dieser Befund widerspricht der Venture-Theorie, da diese, zumindest implizit, davon ausgeht, dass es keinen Zusammenhang zwischen Ambiguität und Entscheidungsgewichten gibt.

In meinen Augen spricht der Befund im Verlustbereich in hohem Maße für die Konservatismushypothese: Im Gewinn- wie im Verlustbereich zog höhere Ambiguität ein vorsichtigeres Verhalten nach sich. Auch der fehlende Zusammenhang zwischen Ambiguität und Bearbeitungsdauer deutet in Richtung der Konservatismushypothese; die Gedächtnisdaten hatten sich als nicht aussagekräftig erwiesen.

Ich bewerte die Evidenz für die Konservatismushypothese als deutlich; sie ist aber keinesfalls zwingend. Das Sinken von θ im Verlustbereich kann nur dann als positive Evidenz herangezogen werden, wenn man davon ausgeht, dass die Entscheidungsprozesse in Gewinn- und Verlustbereich ähnlich waren – falls diese völlig verschieden waren, so können die Ergebnisse aus dem einen Bereich nicht zur Interpretation der Werte aus dem anderen herangezogen werden.

Nun sind aber systematische Unterschiede zwischen Gewinn- und Verlustentscheidungen gang und gäbe. Zu denken ist vor allem daran, dass Entscheidungen im Verlustbereich risikofreudiger ausfallen, aber auch daran, dass Entscheidungen, die Gewinne betreffen, in der Regel schneller gefällt werden. Beides hatte sich auch in Experiment II gezeigt. Des Weiteren ließ sich das Verhalten der Vpn im Gewinnbereich besser replizieren als bei den Verlustaufgaben, was dafür spricht, dass das Vorgehen der Teilnehmer hier systematischer war. Inwieweit sprechen diese Befunde dafür, dass den Entscheidungen im Gewinn- und Verlustbereich unterschiedliche Prozesse zu Grunde lagen?

Die Antwort auf diese Frage muss spekulativ bleiben. Denn die in diesem Bereich relevante Forschung hat ihrem Ansatz nach oft rein psychophysischen Charakter, das heißt es werden lediglich Beziehungen zwischen Reizen und Reaktionen beschrieben, ohne dass vermittelnde psychologische Prozesse thematisiert werden. Die Prospect Theorie oder das Contingent Weighting Modell sind Beispiele dafür; die Venture Theorie bleibt da, wo sie Aussagen über psychologische Mechanismen macht, eher schwammig. Von daher ist mir auch keine Arbeit bekannt, die sich mit der Frage, ob Entscheidungen im Verlustbereich anderen Prozessen unterliegen als bei Gewinnen, befasst. Typischerweise wird aber davon ausgegangen, dass Entscheidungen in beiden Bereichen sinnvoll mit denselben Modellen beschrieben werden können. Ich teile diese Auffassung. Und somit ist für mich das Sinken der θ -Werte im Verlustbereich ein klares Zeichen dafür, dass höhere Ambiguität in Experiment II generell zu einer vorsichtigeren Haltung gegenüber Risiko führte.

Die in den Experimenten I und II verwendete Operationalisierung von Ambiguität über eine zufällig plazierte Verdeckungsfläche ist sehr spezifisch. Somit eröffnet sich die Frage, ob der hier gefundene Effekt auch bei anderen Arten von Ambiguität auftritt. Mit dieser Frage verbunden ist diejenige, ob der Befund von Gonzalez-Vallejo et al. (1994), dass verbal über Wahrscheinlichkeiten informierte Entscheider im Vergleich zu numerisch informierten bei solchen Gewichtungsaufgaben besser abschnitten, in denen die Erwartungswerte der Lotterien höher mit den Auszahlungsbeträgen

korrelierten (vgl. Abb. 10), tatsächlich im Widerspruch zu den Ergebnissen der Experimente I und II steht. Denn dass ein Widerspruch vorliegen würde, setzte Folgendes voraus: (i) Die treibende Kraft, die die Vpn von Gonzalez-Vallejo et al. (1994) möglicherweise dazu geführt hatte, ihr Augenmerk vermehrt auf die Wertkomponente zu richten, wenn sie über die Wahrscheinlichkeiten verbal informiert worden waren, bestand in der *Ambiguität* der verbalen Information und nicht vielleicht in einem anderen Unterscheidungsmerkmal, etwa dem, dass Worte und keine Zahlen benutzt worden waren. (ii) Ambiguität wirkt immer gleich, egal wodurch sie verursacht wird; oder zumindest ist der Unterschied bedeutungslos, ob Ambiguität aus dem verbalen Ausdruck von Wahrscheinlichkeit oder aus der teilweisen Verdeckung von Glücksräädern resultiert.

Experiment III soll dieser Frage nachgehen: Lässt sich der gefundene Effekt auch dann erreichen, wenn die Ambiguität daher stammt, dass die Wahrscheinlichkeitskomponente der Lotterien in Worten beschrieben wird? Oder führt die Verwendung von Worten möglicherweise dazu, dass dann, wie von der Gewichtungshypothese postuliert, der Wertdimension bei der Entscheidungsfindung ein höheres Gewicht zukommt?

11. EXPERIMENT III

Experiment III prüft die Gewichtungshypothese unter Verwendung von graphisch und verbal erzeugter Ambiguität im Gewinn- und Verlustbereich. Das zentrale Ergebnis ist der Abb. 43 zu entnehmen.

11.1 Materialien

Das Paradigma von Experiment III war dasselbe wie bei den beiden vorangegangenen. Die Grundidee des Versuchs machte es notwendig, verbale Beschreibungen der Lotteriewahrscheinlichkeiten zu erhalten. Wenn wie bislang für jede Vp jede Lotterie neu ausgelost worden wäre, hätte dies zur Folge gehabt, dass Beschreibungen von 96 verschiedenen Wahrscheinlichkeiten (von .03 bis .97) benötigt worden wären. Um dies zu verhindern, wurde mit lediglich zehn verschiedenen Aufgaben à sechs Lotterien gearbeitet. Diese beinhalteten insgesamt 46 verschiedene Wahrscheinlichkeiten. Die Aufgaben sind im Anhang in Tab. A wiedergegeben.

Gewinnung der Aufgaben

Zur Gewinnung dieser zehn Aufgaben wurden jeweils sechs Lotterien, die gemeinsam eine Aufgabe bildeten, zufällig ausgelost. Aus solcherart gewonnenen Aufgaben wurden dann zehn Stück dergestalt ausgewählt, dass die folgenden Bedingungen erfüllt wurden: (i) Der Mittelwert von P liegt nahe .5, derjenige von ϵ nahe 50 (für die Aufgaben von Experiment III galt $M_P = .529$, $M_\epsilon = 50,4$), (ii) die Varianzen von P und ϵ sind derjenigen von $SD = 27,4$ ähnlich, die sich bei einer Auslosung der Lotterien, wie sie in den vorangegangenen Experimenten erfolgt war, ergibt ($SD_P = 30,1$,

$SD\epsilon = 29,7$), sowie (iii) die Mittelwerte von $V(EV, P)$ und $V(EV, \epsilon)$ sind einander ähnlich ($M_{V(EV, P)} = .77$, $M_{V(EV, \epsilon)} = .75$).

Verbale Beschreibungen der Wahrscheinlichkeiten

Um verbale Beschreibungen der 46 verwendeten Wahrscheinlichkeiten zu erhalten, wurden diese in Form von Glücksräder drei Personen vorgelegt. Deren Aufgabe bestand darin, die Wahrscheinlichkeiten so gut wie möglich in Worte zu fassen. Auf diese Weise entstanden drei Listen, deren jede die Übertragung aller 46 Wahrscheinlichkeiten in Worten enthielt.

Die verbalen Beschreibungen mussten dabei folgenden Restriktionen Rechnung tragen: (i) es durften keine Zahlen benutzt werden – Beschreibungen wie „knapp 40%“ waren also tabu; als Ausnahme davon war die Formulierung „fifty-fifty“ gestattet, da diese alltäglich ist; (ii) die Beschreibungen durften sich nicht auf eine flächenartige Darstellung von Wahrscheinlichkeit beziehen – somit waren Beschreibungen wie „bis sieben Uhr“ oder dergleichen nicht statthaft; (iii) aus pragmatischen Gründen sollten die Beschreibungen kurz sein; (iv) alle Beschreibungen mussten auf Ereignisse mit positiven wie negativen Konsequenzen anwendbar sein – alle wertenden Begriffe wie „Chance“, „Risiko“, „Glück“, „Hoffnung“, „Gefahr“, „toll“, „hervorragend“, „schlecht“ usw. mussten somit außen vor bleiben; (v) die Formulierung sollte auch ohne Kontext eindeutig als Beschreibung einer Wahrscheinlichkeit verständlich sein – somit war „sehr geringe Wahrscheinlichkeit“ erlaubt, einfach „sehr gering“ aber nicht. Schließlich wurden die Wahrscheinlichkeitsbeschreiber darauf hingewiesen, dass es keinen Zwang zur Abwechslung gebe, dass es somit völlig in Ordnung sei, eine Formulierung öfter zu verwenden.

Da vorherige Bemühungen gezeigt hatten, dass diese Aufgabe vielen Leuten erhebliche Schwierigkeiten bereitete, habe ich mich diesbezüglich an drei mir bekannte Akademiker gewandt, die diese Aufgabe unentgeltlich erledigten. Wie diese die Wahrscheinlichkeiten in Worte übersetzten, ist dem Appendix, Tabelle B, zu entnehmen. Die drei Codierer gebrauchten zwischen sechs und achtzehn Begriffen.

Wie hoch ist die Ambiguität der Wahrscheinlichkeitsbeschreibungen?

Um die Effekte, die verbal erzielte Ambiguität auf das Entscheidungsverhalten ausübt, besser einschätzen zu können, ist es sinnvoll, mit einer Kontrollbedingung zu arbeiten. In dieser sollten die Lotteriewahrscheinlichkeiten mit derselben Ambiguität wie in der verbalen Bedingung präsentiert werden; die Darstellung sollte jedoch wiederum wie in den Experimenten I und II graphisch erfolgen. Um eine solche Kontrollbedingung herstellen zu können, ist es notwendig zu ermitteln, wie hoch die Ambiguität ist, die den verbalen Wahrscheinlichkeitsbeschreibungen zu eigen ist.

Wie lässt sich nun die Ambiguität bestimmen, die diesen Beschreibungen innewohnt? Zwei Wege sind denkbar. Zum einen könnte man Ambiguität mit Informationsverlust gleichsetzen; dann würde man eine Reihe von Beurteilern bitten, die Begriffe in numerische Wahrscheinlichkeiten rückzuübersetzen. Der aufgetretene Informationsverlust lässt sich dann durch den Vergleich der Originalwahr-

scheinlichkeiten mit den rückübersetzten ermitteln, und es kann dann bestimmt werden, wie groß im Glücksradparadigma die Verdeckungsfläche sein muss, um dort einen ebenso großen Informationsverlust zu bewirken. Eine andere mögliche Vorgehensweise bestünde darin, Ambiguität als individuell wahrgenommene Vagheit zu betrachten. Diese ließe sich ermitteln, indem man Beurteiler bittet, für jede der Wahrscheinlichkeitsumschreibungen eine Spannbreite anzugeben, die angibt, was man sinnvollerweise unter der jeweiligen Formulierung verstehen kann.

Wie zwei Voruntersuchungen ergaben, kommen beide Methoden zu deutlich verschiedenen Einschätzungen der Ambiguität der verbalen Umschreibungen.

Ambiguität als Informationsverlust

Um den Informationsverlust zu ermitteln, zu dem der verbale Ausdruck der Wahrscheinlichkeiten führt, wurden die drei Listen fünfzehn Beurteilern vorgelegt; jeder Beurteiler erhielt dabei nur eine der drei Listen. Für jede Wahrscheinlichkeitsbeschreibung ihrer Liste gaben die Beurteiler eine Punktschätzung darüber ab, welche Wahrscheinlichkeit der Übersetzer bei der Verwendung der Formulierung im Sinn hatte. Die Beurteiler waren nicht darüber informiert, dass manche der Formulierungen für verschiedene Wahrscheinlichkeiten verwendet worden waren. Für jede Liste lagen somit fünf Korrelationen zwischen Ursprungs- und rückübersetzten Wahrscheinlichkeiten vor. Von diesen Korrelationen wurde jeweils der Median gebildet. Die Mediane betrugen $r = .96$ für die Listen 1 und 3 sowie $r = .89$ für die Liste 2 (da die Streuungen der Korrelationen innerhalb jeder Liste äußerst gering waren, ist der Unterschied zwischen Liste 2 und den anderen beiden trotz der geringen Beurteilerzahl durchaus ernst zu nehmen). Geht man davon aus, dass im Glücksradparadigma die wahren Wahrscheinlichkeiten anhand von *Normal* erschlossen werden, so erzeugen Verdeckungen von 27% bzw. 45% entsprechende Korrelationen¹³.

Ambiguität als subjektive Bedeutungsspannbreite

In einer zweiten Untersuchung wurde überprüft, als wie vage die verbalen Beschreibungen empfunden wurden. Fünf Beurteilern wurde dazu eine Liste vorgelegt, die sämtliche Wahrscheinlichkeitsumschreibungen der drei Übersetzer enthielt. Hier wurden die Beurteiler gebeten, für jede der Formulierungen eine Spannbreite der möglichen Bedeutung anzugeben (s. o.). Aus diesen Angaben ließ sich nun für jede der drei Listen, über Formulierungen und Beurteiler gemittelt, die durchschnittliche Spannbreite errechnen. Bei dieser Mittelwertsberechnung wurden die einzelnen Formulierungen mit ihrer Auftretenshäufigkeit gewichtet; denn während viele Formulierungen pro Übersetzer nur einmal verwendet worden waren, kamen andere bis zu zwölf mal vor. Zwischen den drei Listen zeichneten sich keine nennenswerten Unterschiede ab; der Mittelwert über die drei Listen hinweg betrug 12,5. D. h. um die Formulierung einer beliebigen der 46 verwendeten Wahrscheinlichkeiten wurde im Mittel lediglich ein Intervall von 12,5% gelegt.

¹³ Diese Werte wurden durch eine Monte-Carlo Simulation ermittelt.

Zu der Frage, wie Leute verbale Beschreibungen von Wahrscheinlichkeiten verstehen, ist mir nur eine Studie bekannt, die Bedeutungsspannbreiten erhoben hat und von diesen auch die Mittelwerte angibt (Reagan, Mosteller & Youtz, 1989). Über sämtliche Beurteiler und verwendeten Formulierungen hinweg betrug der mittlere Range hier 12%; er stimmt also sehr gut mit meinem Befund überein.

Wie hoch sollte nun die Ambiguität in der Kontrollbedingung sein? 12% – oder 27% bzw. 45%? Ich habe mich für den ersten Wert entschieden, der auf der Angabe von Bedeutungsspannbreiten beruht. Denn für die Wirkung, die Ambiguität auf das Entscheidungsverhalten ausübt, sollte weniger der tatsächliche Informationsverlust als vielmehr die subjektiv wahrgenommene Vagheit der Information ausschlaggebend sein.

Die Darbietung der Aufgaben entsprach derjenigen in den Experimenten I und II. Wie die Aufgaben, bei denen die Gewinn- bzw. Verlustwahrscheinlichkeiten in Worten dargestellt worden waren, aussahen, ist aus Abb. 39 zu ersehen. Die Schriftgröße und die Platzierung der Wahrscheinlichkeitsbeschreibungen habe ich so gewählt, dass mir die Aufmerksamkeit, die diese auf sich zieht, mit derjenigen, die die Glücksräder hervorrufen, vergleichbar erschien.

In Experiment II waren offensichtlich vereinzelt Versuchspersonen des Wechsels zwischen Gewinn- und Verlustaufgaben nicht gewahr geworden. Um auf diese Wechsel noch klarer hinzuweisen, wurden sie daher in Experiment III von einem kurzen Signalton begleitet.

Um auch für die Aufgaben, in denen die Lotteriewahrscheinlichkeiten in Form verbaler Beschreibungen präsentiert werden, Schätzwerte für θ bestimmen zu können, ist es notwendig zu wissen, welche Bedeutung diese Beschreibungen für eine bestimmte Person haben. Dieses individuelle Verständnis wurde mit Hilfe eines Fragebogens erfasst, auf dem alle verbalen Beschreibungen eines Versuches vorkamen. Neben der Angabe einer besten Schätzung für die Bedeutung eines jeden Begriffs ($P_{\text{Schätz}}$) wurde zusätzlich ein Intervall abgefragt. Ober- und Untergrenze ($P_{\text{upper}}, P_{\text{lower}}$) sollten dabei den Wahrscheinlichkeitsbereich abstecken, für den die entsprechende Formulierung in den Augen der Vp sinnvoller Weise verwendet werden kann. Dadurch sollte auch überprüft werden, ob die für die Kontrollbedingung gewählte Ambiguität von 12% tatsächlich adäquat war. Diesem Ziel diente auch die in den Bogen aufgenommene Schlussfrage, ob sich der Teilnehmer besser durch die verdeckten Glücksräder oder die verbalen Wahrscheinlichkeitsbeschreibungen über die Chancen bzw. Risiken der einzelnen Lotterien informiert gefühlt hat.

Für jede der drei Listen mit verbalen Umschreibungen wurde der Fragebogen in sechs verschiedenen Versionen erstellt: Drei jeweils zufällig erzeugte Itemreihenfolgen wurden gekreuzt mit den beiden möglichen Reihenfolgen der Antwortvorgaben für die Schlussfrage. Eine Fragebogenversion für die Liste 1 ist exemplarisch im Anhang (A3) wiedergegeben.

11.2 Aufbau

In Experiment III wurde ein 3×2 Within-Subjects Design verwirklicht. Für die Darstellung der Lotteriewahrscheinlichkeiten gab es die Bedingungen „graphisch ohne Ambiguität“, „graphisch mit

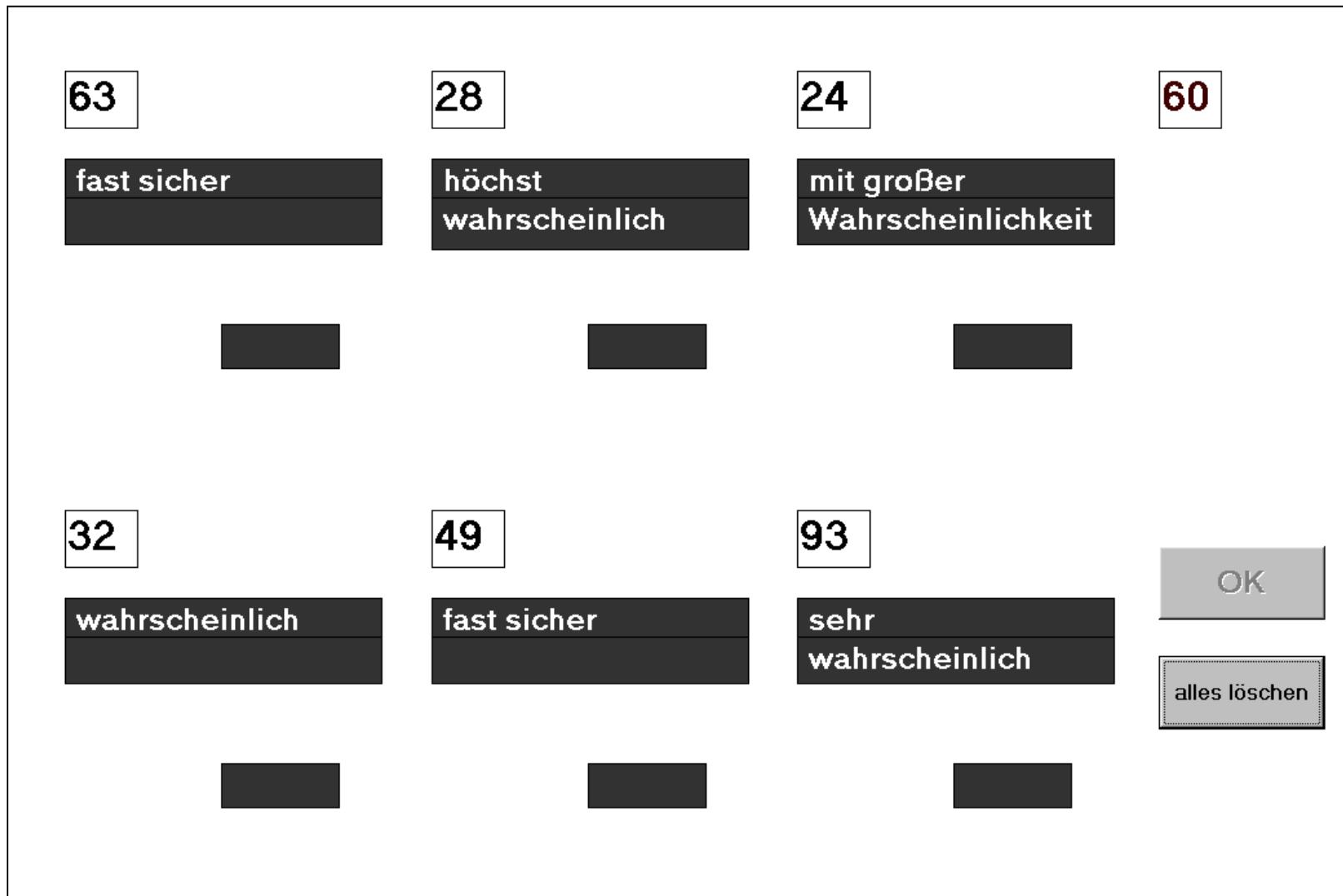


Abb. 39 Experiment III: Exemplarische Bildschirmgestaltung unter der Bedinung „verbal“. In die kleinen dunkelgrauen Kästchen wurde durch Mausklicken der Wertungsfaktor gesetzt. Das Kästchen mit der 60 zeigt den Rundenzähler. Im Experiment waren die hier dunkelgrau abgebildeten Flächen gelb, die Wahrscheinlichkeitsbeschreibungen und die Wertungsfaktoren waren schwarz dargestellt.

12% Ambiguität“ und „verbal“. Der zweite Faktor resultierte daraus, dass wiederum Gewinn- und Verlustaufgaben einbezogen wurden.

In der Bedingung „verbal“ erhielt jeweils ein Drittel der Versuchsteilnehmer die verbalen Beschreibungen der Liste 1, 2 und 3. Die Zuordnung erfolgte zufällig.

In den Experimenten I und II waren die Aufgaben für jede Vp neu und zufällig erzeugt worden. Aus bereits erwähnten Gründen wurde von diesem Vorgehen für die Bedingung „verbal“ abgewichen; hier habe ich mich mit einem festen Satz von zehn Aufgaben begnügt. Um die Vergleichbarkeit der verschiedenen Bedingungen zu gewährleisten, wurden dieselben Aufgaben auch in den anderen Blöcken verwendet. Alle sechs Blöcke bestanden also aus denselben zehn Aufgaben (bei den Verlustaufgaben waren die Lotteriewerte natürlich negativ). In der Bedingung „graphisch mit Ambiguität“ erfolgte die Platzierung der Verdeckungsflächen jedoch wie gehabt zufällig. Somit variierte in dieser Bedingung P_{Sicht} von Teilnehmer zu Teilnehmer.

Die sechzig Aufgaben wurden blockweise dargeboten. Die Präsentationsreihenfolge der Blöcke wurde für jede Vp neu randomisiert. Ebenso wurde die Reihenfolge der Lotterien innerhalb einer Aufgabe für jede Versuchsperson und jede Aufgabe neu und zufällig bestimmt. Durch diese Maßnahme wurde es den Teilnehmern praktisch unmöglich gemacht zu erkennen, dass sie jede Aufgabe sechs mal bearbeiteten.

Der Fragebogen zu den subjektiven Bedeutungen der verbalen Wahrscheinlichkeitsbeschreibungen wurde stets zum Schluss gegeben, da eine Untersuchung von Erev, Bornstein und Wallsten (1993) gezeigt hat, dass Entscheidungen, die auf verbaler Wahrscheinlichkeitsinformation beruhen, systematisch verändert werden, wenn man von den Vpn vorher eine explizite numerische Übersetzung einfordert. Da Menschen eine solche Übersetzungsaufgabe im Alltag nicht abverlangt wird, würde man die Daten also mit einem artifiziellen Effekt belasten, wenn man die Hälfte der Vpn die Übersetzungsaufgabe vorab erledigen ließe.

11.3 Versuchsteilnehmer und Versuchsablauf

Über Kleinanzeigen und Aushänge an der Universität konnten 58 Teilnehmer (45 Frauen und 13 Männer) gewonnen werden. Bei einer Teilnehmerin musste der Versuch abgebrochen werden, da ihr jegliches Verständnis für die letzte Aufgabe, die verbalen Wahrscheinlichkeitsbeschreibungen in Zahlen rückzuübersetzen, fehlte. Somit gingen die Daten von 57 Personen in die Auswertung ein. Deren Altersmedian lag bei 21,0 Jahren. Sofern die Teilnehmer sich im Grundstudium Psychologie befanden, wurde ihnen die Versuchsteilnahme auf ihre Zulassungsvoraussetzung zum Vordiplom angerechnet. Alle anderen erhielten 8,- DM. Zusätzlich konkurrierten alle Teilnehmer um Prämien in Höhe von 80,- DM, 60,- DM und 40,- DM, die für die drei Erstplatzierten ausgelobt waren.

Zu Beginn des Versuchs lasen die Teilnehmer die Instruktion, die im Anhang unter A2 wiedergegeben ist. Im Anschluss daran wurde die Bedienung des Programms demonstriert und dabei die Logik des Versuchs noch einmal erläutert. Die Teilnehmer wurden darauf hingewiesen, dass sie nach der Bearbeitung der 60 Aufgaben noch einen kurzen Fragebogen auszufüllen hätten. Zu Beginn des eigentlichen Versuchs konnten die Vpn nach Wunsch Probefürchläufe absolvieren, um sich mit der

Handhabung des Programms vertraut zu machen. Hier wurden zufällig generierte Gewinnaufgaben mit graphisch dargestellten, nichtambigen Wahrscheinlichkeiten dargeboten. Weder in den Probendurchgängen noch in den relevanten Aufgaben erhielten die Teilnehmer Rückmeldung über die Folgen ihrer Entscheidungen; ihnen wurde lediglich zum Schluss ihr Gesamtpunktestand angezeigt. Im Anschluss daran füllten sie dann den Fragebogen zu den subjektiven Bedeutungen der Wahrscheinlichkeitsbeschreibungen aus.

Die Teilnehmer absolvierten den Versuch alleine oder in kleinen Gruppen von bis zu vier Personen. Im letzteren Fall konnten sie keinen Blick auf die Bildschirme ihrer Mitstreiter werfen. Alle arbeiteten im selbstgewählten Tempo. Die reine Aufgabenbearbeitungszeit wies dabei einen Median von 27 Minuten auf.

11.4 Ergebnisse

11.4.1 Ausschluss von Daten von der weiteren Analyse

Der Ausschluss von Daten von der weiteren Analyse erfolgte nach demselben Prinzip wie in den Experimenten I und II (vgl. 9.3.1). Das Leistungsmaß erwies sich wiederum als stark rechtsschief verteilt. Das heißt fast alle Versuchspersonen erbrachten sehr gute Leistungen, einige wenige fielen in der Leistung deutlich ab. Bei dreien lag das Leistungsmaß unter dem in den Experimenten I und II gewählten Kriterium von .55. Der Leistungskennwert einer vierten Person lag mit .563 zwar knapp über diesem alten Kriterium, war aber im Scatterplot klar als Ausreißer zu erkennen. Von der weiteren Analyse wurden daher die Daten dieser vier Teilnehmer ausgeschlossen.

11.4.2 Die Leistung der Vpn

Die Analyse der Leistungen erfolgte wieder auf Basis von *SFE* (vgl. 9.3.2). Für alle drei Ambiguitätsbedingungen musste dabei erneut bestimmt werden, wieviele Punkte einer *SFE* entsprechen. Dies war auch für die Bedingung ohne Ambiguität nötig, da die Aufgaben in Experiment I und II anders generiert worden waren als die in Experiment III. Für die in Experiment III verwendeten Aufgaben wurde der Wert also nach dem oben beschriebenen Verfahren neu bestimmt (vgl. 9.3.2). 10,61 Punkte entsprachen dabei einer *SFE*. Auf gleiche Art wurde der Wert für 12% Ambiguität bestimmt: $SFE = 10,55$ Punkte. Für die Bedingung „verbal“ unterschied sich der Punktewert, der einer *SFE* entspricht, von Versuchsperson zu Versuchsperson. Dies lag daran, dass jeder Teilnehmer ein anderes Verständnis der verbalen Wahrscheinlichkeitsbeschreibungen hatte¹⁴. Es ergaben sich *SFE*-Äquivalente zwischen 4,47 und 10,78 Punkten; der Median lag bei 10,26 Punkten¹⁵.

¹⁴ Wie die Werte für die Bedingung „verbal“ ermittelt wurden, ist im Appendix unter A4 geschildert.

¹⁵ Dass dieser Wert unter demjenigen für 12% Ambiguität liegt, macht Sinn, denn in der Bedingung „verbal“ ist der Informationsverlust höher (vgl. 11.1). Je höher der Informationsverlust ist, desto weniger fällt jedoch eine Abweichung von einer guten Referenzstrategie ins Gewicht (vgl. 9.3.2). Dies spiegelt sich auch in dem engen Zusammenhang zwischen der Güte des Verständnisses der Wahrscheinlichkeitsbeschreibungen einer Versuchsperson und ihrem *SFE*-Äquivalent. Misst man die Güte des Verständnisses als mittlere absolute Differenz zwischen tatsächlicher und rückübersetzter Wahrscheinlichkeit, so korreliert dieser Wert mit $r = -.82$ mit dem *SFE*-Äquivalent ($n = 53, p < .001$).

Da die *SFE*-Werte unter allen Versuchsbedingungen stark linksschief verteilt waren, gibt Abb. 40 die Mediane wieder. Wie zu erkennen ist, übte keiner der experimentellen Faktoren einen Einfluss auf die Leistung aus; die Mediane lagen für alle Bedingungen zwischen 1,3 und 1,5 *SFE*, keiner der Unterschiede erwies sich als statistisch signifikant. Die Bedingung 0% Ambiguität hatte Experiment III mit den Experimenten I und II gemein. Hier zeigten sich keine interpretierbaren Unterschiede zwischen den Experimenten: Die entsprechenden Tests ergaben, dass die Unterschiede in den *SFE*-Werten statistisch nicht signifikant waren.

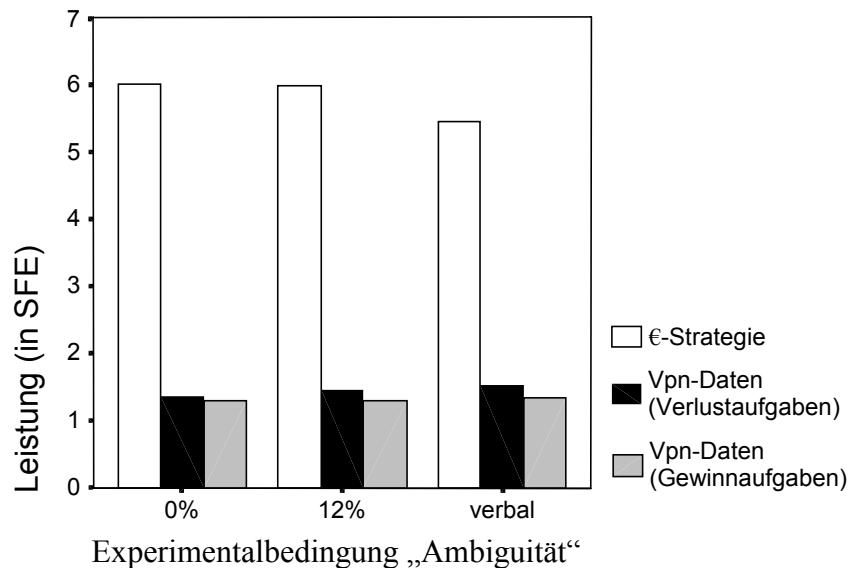


Abb. 40 Experiment III: Leistungsdifferenz der Vpn und der ϵ -Strategie zu *Normal*, gemessen in *SFE* [Mediane]. Die Leistungen der ϵ -Strategie waren für Gewinn- und Verlustbereich nahezu identisch – in der Abb. sind die Mediane abgetragen.

Ein Einfluss der verwendeten Liste mit Wahrscheinlichkeitsbeschreibungen auf die Leistung war weder im Gewinn- noch im Verlustbereich zu finden: Separat durchgeführte Kruskall-Wallis Tests, dem nichtparametrischen Pendant zur einfaktoriellen ANOVA, mit dem dreistufigen Faktor „Liste“, führten zu keinen signifikanten Ergebnissen (Gewinn: $\chi^2 = .8$; $df = 2$; $p = .66$. Verlust: $\chi^2 = 1,9$; $df = 2$; $p = .39$).

Unter allen sechs Versuchsbedingungen waren die Leistungen der Vpn derjenigen der ϵ -Strategie weit überlegen, die Unterschiede zur Leistung von *Normal* waren vergleichsweise gering. Wie bereits in den vorherigen Experimenten zeigten sich die Vpn also als erfolgreiche Maximierer.

11.4.3 Die Identifikation eines geeigneten Basismodells

Die Suche nach einem geeigneten Basismodell erfolgte nach demselben Schema wie in den Experimenten I und II (vgl. 9.3.3). Um θ schätzen zu können, war es jedoch zunächst notwendig, die Basismodelle in geeigneter Weise an die Bedingung „verbal“ anzupassen.

Im Falle graphisch dargestellter Ambiguität nimmt *Empirisch* eine Abhängigkeit zwischen P_{Sicht} und dem Anteil der Verdeckungsfläche, der P_{Sicht} zugeschlagen wird, an (vgl. 8.2.1). Dieses Konzept ist in meinen Augen auf verbale Beschreibungen von Wahrscheinlichkeiten nicht sinnvoll übertragbar. In der Bedingung „verbal“ wurde *Empirisch* daher so abgeändert, dass die Lotteriewahrscheinlichkeit mit dem Schätzwert ($P_{\text{Schätz}}$), den die Vpn für die Wahrscheinlichkeitsbeschreibung abgegeben hatten, gleichgesetzt wurde. Somit galt

$$\text{Empirisch} \text{ (Versuchsbedingung „verbal“): Attraktivität} = \epsilon \cdot P_{\text{Schätz}}.$$

Die Basismodelle *VentureI-IV* sind auf einen Ankerwert angewiesen (vgl. 8.2.2). Wenn in einer Lotterie die Wahrscheinlichkeit durch Worte beschrieben wird, so liegt es in meinen Augen nahe, einer Evaluation der Wahrscheinlichkeitskomponente zunächst den Wert zu Grunde zu legen, der subjektiv am ehesten die verbale Beschreibung trifft. Als Ankerwert wurde hier also $P_{\text{Schätz}}$ benutzt, ansonsten blieben die *Venture*-Modelle I bis IV unverändert.

Konservativ geht bei graphisch dargestellter Ambiguität vom jeweils schlechtesten Fall aus; im Gewinnbereich wird die Wahrscheinlichkeit also als P_{Sicht} geschätzt, im Verlustbereich als $P_{\text{Sicht}} + \text{Ambig}$. Eine solche pessimistische Sichtweise ist mühelos auf den Fall verbaler Wahrscheinlichkeitsbeschreibungen übertragbar. Im Gewinnbereich wird die Wahrscheinlichkeit als unterer Wahrscheinlichkeitsschätzwert (P_{lower}) angesehen, im Verlustbereich als oberer (P_{upper}).

Im Fall graphisch erzeugter Ambiguität geht *Normal* davon aus, dass die Wahrscheinlichkeit als $P_{\text{Sicht}} + \text{Ambig}/2$ geschätzt wird, da dieser Wert dem subjektiven Erwartungswert entsprechen dürfte und ihm deshalb eine besondere Plausibilität zukommt. Äquivalent dazu kann in der Bedingung „verbal“ $P_{\text{Schätz}}$ als subjektiv plausibelster Wert für die Lotteriewahrscheinlichkeit angenommen werden. Somit ergibt sich

$$\text{Normal} \text{ (Versuchsbedingung „verbal“): Attraktivität} = \epsilon \cdot P_{\text{Schätz}}.$$

Normal ist hier also mit *Empirisch* identisch.

Venture einfach geht davon aus, dass die Schätzung der Lotteriewahrscheinlichkeit zum einen konservativ ist und zum anderen auf einfache Weise vom Lotteriewert beeinflusst wird (vgl. 8.2.2). Beides lässt sich gut auf die Bedingung „verbal“ übertragen; hier gilt

$$V_{\text{einfach}} \text{ (Gewinnbereich): Attraktivität} = \epsilon \cdot (P_{\text{lower}} + (60 - \epsilon/2) \cdot (P_{\text{upper}} - P_{\text{lower}})/100)$$

$$V_{\text{einfach}} \text{ (Verlustbereich): Attraktivität} = \epsilon \cdot (P_{\text{lower}} + (40 - \epsilon/2) \cdot (P_{\text{upper}} - P_{\text{lower}})/100)$$

Wie in beiden Vorgängerexperimenten zeigte sich auch in Experiment III ausnahmslos, dass diejenige Variante eines Basismodells das Versuchspersonenverhalten am besten wiedergeben konnte, die einen linearen Zusammenhang zwischen ϵ und $u(\epsilon)$ annahm. Daher wurden im Weiteren von allen Basismodellen allein die linearen Varianten näher betrachtet.

Wie in beiden vorherigen Experimenten auch verteilten sich die Δ -Werte der acht Basismodelle deutlich linksschief. Dem entsprechend gibt Abb. 41 die Mediane der Replikationsfehler wieder. Wie schon in den Experimenten I und II eignete sich *Normal* am besten, das Verhalten der Vpn zu reproduzieren. Dies galt nicht nur bei der Betrachtung über alle Versuchsbedingungen hinweg, wie sie Abb. 41 zeigt. Auch bei einer getrennten Betrachtung von „alten“ Versuchsbedingungen (d. h. solchen mit graphischer Darstellung der Wahrscheinlichkeit) und der neu eingeführten Bedingung „verbal“ schnitt das Basismodell *Normal* jeweils am besten ab. Insgesamt lag allerdings der Replikationsfehler dieses Modells mit einem Median von .353 nur minimal unter dem der Basismodelle *Empirisch* ($\Delta = .358$), *Konservativ* ($\Delta = .368$) und *Venture_{einfach}* ($\Delta = .369$); der Vorsprung gegenüber diesen Modellen erwies sich als nicht signifikant. Erst der Unterschied zum nächsten Modell, *Venture_{III}*, kann als stabil angesehen werden (Wilcoxon Rangsummentest: $p < .001$). Aus zwei Gründen erscheint es mir dennoch gerechtfertigt, die Analyse von θ wieder allein anhand der Schätzwerte, die auf der Grundlage von *Normal* gewonnen wurden, durchzuführen. Zum einen hat sich dieses Basismodell in den beiden vorangegangenen Experimenten in besonderer Weise bewährt, so dass man eher geneigt sein kann, dessen Vorsprung gegenüber den anderen Modellen als nicht rein zufällig zu betrachten. Wichtiger wiegt jedoch ein zweiter Grund: Zur Validierung der *Normal*-Schätzung von θ wurde diese mit einer anderen, gemischten Schätzung kontrastiert. Bei dieser wurde für jede Vp individuell das beste Basismodell bestimmt und die sechs θ -Schätzer dieses Modells betrachtet. Das θ -Muster, welches dieses gemischte Modell über die sechs Versuchsbedingungen produzierte, war quasi identisch mit dem Muster, das sich auf der Grundlage von *Normal* ergab ($r = .92$; $n = 6$; $p = .01$). Somit erscheint es mir unproblematisch, im Folgenden allein die Schätzer, die *Normal* lieferte, zu betrachten.

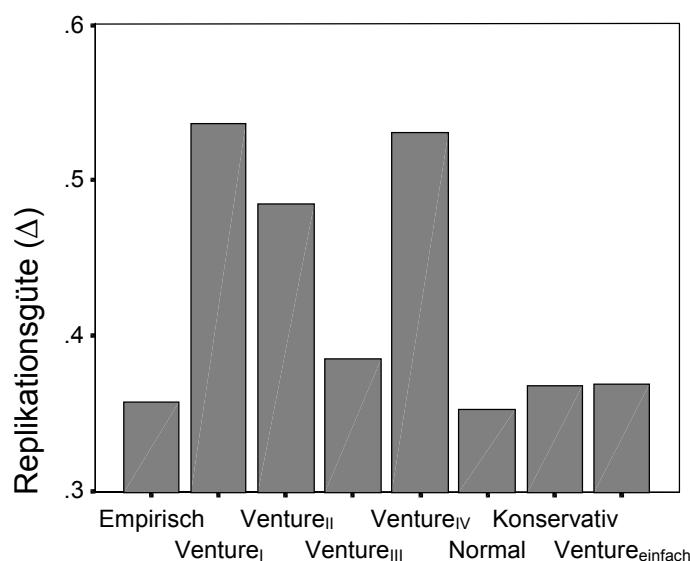


Abb. 41 Experiment III: Für die acht Basismodelle ist hier angegeben, wie gut sie das Verhalten der Vpn nachbilden konnten [Mediane von Δ]. Kleinere Werte stehen für eine höhere Güte.

Wie gut sich das Wahlverhalten der Vpn in den einzelnen Versuchsbedingungen auf der Grundlage von *Normal* reproduzieren ließ, ist aus Abb. 42 zu ersehen. Wie bereits in Experiment II ließen sich die Wahlen im Gewinnbereich durchweg besser wiedergeben als im Verlustbereich. Mittelt man für jede Vp die drei Δ -Werte des Gewinnbereichs, so weisen diese Werte einen Median von .29 auf. Der äquivalente Wert betrug für den Verlustbereich .39; die Differenz ist statistisch signifikant ($p < .001$; Wilcoxon Rangsummentest). Separat für den Gewinn- und Verlustbereich wurde eine Rangvarianzanalyse für die drei Ambiguitätsbedingungen vorgenommen. In beiden Fällen übten diese einen signifikanten Einfluss aus ($\chi^2 = 14,3$; $df = 2$; $p = .001$ [Verlust]; $\chi^2 = 24,1$; $df = 2$; $p < .001$ [Gewinn]). Einzelvergleiche nach Conover (vgl. Bortz, Lienert & Boehnke, 1991) ergaben, dass im Gewinn- wie im Verlustbereich die Unterschiede zwischen den Bedingungen „0%“ und „verbal“ sowie „12%“ und „verbal“ signifikant waren (p stets $< .002$).

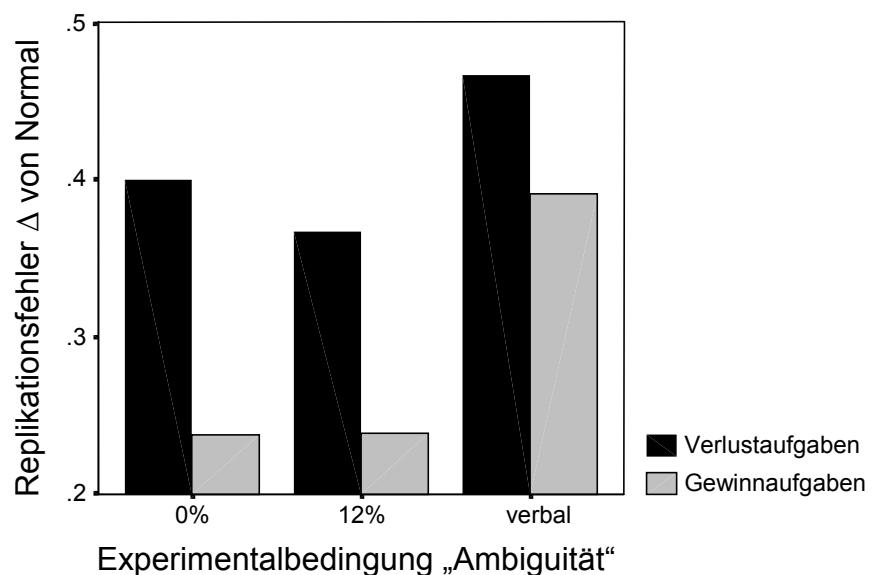


Abb. 42 Experiment III: Güte der Replikation des Verhaltens der Vpn auf der Grundlage von *Normal* [Medianen]. Höhere Werte stehen für eine geringere Güte.

Wie schon in den vorangegangenen Experimenten zeigte sich auch hier auf globaler wie individueller Ebene ein enger Zusammenhang zwischen der Leistung der Vpn und der Replizierbarkeit ihres Verhaltens. Die Medianen, die sich unter den sechs Versuchsbedingungen für die Leistung (*SFE*) ergeben hatten, korrelierten hoch mit den entsprechenden Medianen von Δ ($r = .79$; $n = 6$; $p = .03$ [einseitige Testung]). Der Zusammenhang auf individueller Ebene zeigt sich, wenn man die Gesamtleistung der Vpn (gemessen in *SFE*) mit der Reproduzierbarkeit ihres Wahlverhaltens (ebenfalls gemittelt über alle sechs Versuchsbedingungen) korreliert: $r = .71$ ($n = 53$; $p < .001$).

Wie schon bei den Vorgängerexperimenten legt dieser Befund nahe, dass Leistung und Reproduzierbarkeit gleichsinnig von einer dritten Variable abhingen, nämlich von dem Grad, in dem die einzelnen Vpn konsequent einer Strategie folgten, oder anders, inwieweit die Rangreihen willkürlich vergeben worden sind.

Die Replikationsgüte unter den Bedingungen „0%“ war quasi identisch mit den entsprechenden Werten aus Experiment II (vgl. Abb. 34). Ich sehe die Δ -Werte als zufriedenstellend gering an; in allen Bedingungen blieb der Median deutlich unter .5. Noch einmal zwei Werte zur Orientierung: Der kleinstmögliche Fehler, der bei einer Aufgabe auftauchen kann, d. h. die Vertauschung zweier benachbarter Ränge, führt zu einem Δ von .33; die Vorhersage der Rangfolgen der Vpn allein aufgrund der Lotteriewerte ϵ hätte hier zu einem Δ von 1.26 geführt.

Um die Konsistenz der θ -Schätzung zu überprüfen wurden die θ -Werte auf Grundlage von *Normal* wiederum separat für die geraden und die ungeraden Aufgaben geschätzt. Die Split-Half Reliabilitäten sowie die Konsistenzschätzungen sind der nachfolgenden Tab. 6 zu entnehmen.

Tabelle 6: Rangkorrelationen (ρ) der auf Grund von *Normal* gewonnenen Split-Half Schätzer für θ sowie die nach Spearman-Brown korrigierten Konsistenzschätzungen (in Klammern).

	Ambiguität 0%	Ambiguität 12%	Verbal
Verlustaufgaben	.33 (.50)	.61 (.75)	.62 (.81)
Gewinnaufgaben	.66 (.80)	.69 (.82)	.62 (.81)

Die Konsistenzschätzungen fielen hier ähnlich aus wie in den vorangegangenen Experimenten. Eine Ausnahme bildete die Bedingung „0% Ambiguität“ bei den Verlustaufgaben. Die Reliabilität fiel hier deutlich geringer aus und ist als eher unbefriedigend einzuschätzen. Ein Scatterplot zeigte, dass für diesen geringen Zusammenhang nicht einzelne Ausreißer verantwortlich waren. Möglicherweise war die Aufteilung der Aufgaben nach geraden und ungeraden Nummern hier zufällig besonders ungünstig; eine erneute Reliabilitätsanalyse, bei der die Aufgaben nach ihrer Präsentationsreihenfolge aufgesplittet worden waren (erste Hälfte gegen zweite Hälfte), führte zu einer Split-Half Korrelation von .45, was einer Konsistenz von .62 entspricht. Jedoch ist auch dieser Wert im Vergleich zu den anderen eher niedrig. Eine verminderte Reliabilität dieser Messung sollte jedoch nicht dazu führen, dass der Median von θ in dieser Bedingung falsch geschätzt wird, da sich die Schätzfehler bei einem n von 53 gut ausgleichen sollten. Die erhöhte Fehlervarianz sollte jedoch zur Folge haben, dass etwaige Unterschiede zu anderen Bedingungen weniger schnell signifikant werden.

11.4.4 Die relative Gewichtung der Wahrscheinlichkeit (θ)

Abbildung 43 gibt die Schätzwerte für θ wieder, die auf der Grundlage des Basismodells *Normal* gewonnen worden waren. Zunächst fällt auf, dass wie in Experiment II die Werte im Verlustbereich durchweg niedriger lagen als im Gewinnbereich. Einzelvergleiche nach dem Verfahren von Conover zeigten dabei signifikante Unterschiede für die Bedingungen „0% Ambiguität“ ($p < .001$) und „12% Ambiguität“ ($p < .05$).

Getrennt nach Gewinn- und Verlustbereich wurden die Daten dann einer Rangvarianzanalyse unterzogen; dabei zeigte sich weder im ersten noch im zweiten ein signifikanter Effekt ($\chi^2 = 4,7$; $df = 2$; $p = .10$ bzw. $\chi^2 = .3$; $df = 2$; $p = .87$).

Welche der drei Listen mit Wahrscheinlichkeitsbeschreibungen verwendet worden war, hatte keinen Einfluss auf θ : Für die Bedingung „verbal“ wurden die Schätzwerte, getrennt für Gewinn- und Verlustbereich, einem Kruskall-Wallis Test, dem nichtparametrischen Pendant zur einfaktoriellen ANOVA, unterzogen, wobei die Daten anhand des Faktors „Liste“ in drei unabhängige Gruppen geteilt waren; die Ergebnisse waren nicht signifikant (Gewinn: $\chi^2 = .05$; $df = 2$; $p = .82$. Verlust: $\chi^2 = .06$; $df = 2$; $p = .82$).

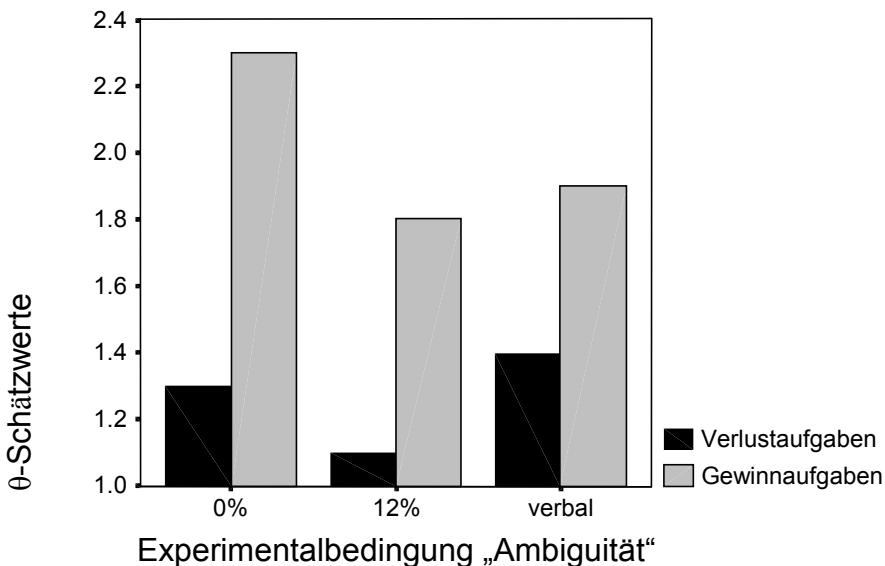


Abb. 43 Experiment III: Die Schätzwerte für θ , gewonnen auf der Grundlage von *Normal* [Medianen].

Das für Gewinnlotterien typische risikomeidende Verhalten zeigte die absolute Mehrheit der Vpn (bei 92% war $\theta > 1$). Im Verlustbereich waren die Risikoeinstellungen hingegen ausgewogen: 27 Teilnehmer zeigten ein eher risikosuchendes Verhalten ($\theta > 1$) und 26 ein eher risikomeidendes ($\theta < 1$).

Das Ergebnis, dass θ in der Bedingung „12% Ambiguität“ nicht verschieden von der Bedingung „0% Ambiguität“ war, passt gut in das Bild, das die Vorgängerexperimente gezeichnet haben. Der Effekt von 25% Ambiguität wurde in Experiment I nicht signifikant, in Experiment II fiel er sehr gering aus; somit nimmt es nicht Wunder, wenn ein deutlich geringerer Ambiguitätsgrad hier keinen Effekt nach sich zog. Es liegt nahe, diese Interpretation auch auf den fehlenden Effekt verbaler Wahrscheinlichkeitsbeschreibungen zu übertragen. Demnach wäre auch hier die Ambiguität zu gering gewesen, um einen Effekt im Sinne der Konservatismushypothese (Ansteigen von θ im Gewinnbereich und Absinken bei Verlustaufgaben), zu erreichen. Misst man die Ambiguität der Beschreibungen nach derselben Logik wie im Vorversuch, nämlich als mittlere Bedeutungsbandbreite $P_{\text{upper}} - P_{\text{lower}}$, so ergaben sich für die drei Listen mittlere Ambiguitätswerte von 13,9% (Liste 1), 11,1% (Liste 2) und 10,5% (Liste 3)¹⁶. Mit 12% Ambiguität scheint die graphische Bedingung also

¹⁶ Dabei wurde die Differenz $P_{\text{upper}} - P_{\text{lower}}$ einer jeden Wahrscheinlichkeitsbeschreibung mit deren Auftretenshäufigkeit gewichtet.

eine gute Vergleichsbedingung abgegeben zu haben, und es ist nicht erstaunlich, dass sich auch für die Bedingung „verbal“ kein Effekt gezeigt hat.

Ganz so eindeutig lassen sich die Daten aber nicht interpretieren, denn eine sehr deutliche Mehrheit von 82% der Vpn erklärte, dass sie sich durch die verdeckten Glücksräder besser informiert gefühlt hätten als durch die Wahrscheinlichkeitsbeschreibungen (interpretierbare Unterschiede zwischen den Listen gab es dabei nicht). Man muss dies nicht zwingend so auslegen, dass diese Beschreibungen als vager empfunden wurden als die verdeckten Glücksräder, aber diese Interpretation liegt zumindest nahe. Somit zeigt sich, dass ein eindeutiger Vergleich des Ambiguitätsgrades zwischen den Präsentationsmodi „verdeckte Glücksräder“ und „verbale Beschreibungen“ kaum möglich ist. Man kann also argumentieren, dass die Vagheit in beiden Bedingungen vergleichbar war und zu gleichen Effekten führte; man kann sich aber auch auf den Standpunkt stellen, dass die „wahre“ Ambiguität in der Bedingung „verbal“ viel höher gelegen hat (vielleicht bei 35%) und dass Ambiguität hier somit anders wirkte als bei den verdeckten Glücksräder, da Experiment II bei vergleichbar hoher graphischer Ambiguität Effekte auf θ gefunden hat.

Ein solcher Streit erscheint mir jedoch müßig. Festzuhalten bleibt, dass die durch verbale Wahrscheinlichkeitsbeschreibungen erzeugte Ambiguität nicht zu einer Veränderung von θ im Sinne der Konservatismushypothese geführt hat – möglicherweise, weil die den Beschreibungen inhärente Ambiguität zu gering war. Für den von Wallsten, Budescu und Tsao (1997) behaupteten Effekt, dass die Verwendung verbaler Beschreibungen, im Sinne einer strategischen Umorientierung, zu einer stärkeren Anlehnung an die Wertkomponente der Lotterien führt, gab es, wie in den vorherigen Experimenten auch, nicht den leisen Hinweis.

Da mir die Validität der θ -Schätzungen mittlerweile hinreichend gesichert erscheint, habe ich hier auf eine separate Analyse der Daten über den Einfluss von P_{Sicht} auf die Attraktivität der Lotterien (vgl. 9.3.5 und 10.3.5) verzichtet.

11.4.5 Bearbeitungszeiten

Um den Einfluss von Ausreißern zu kontrollieren, ging in die Analyse der Bearbeitungszeiten für jede Versuchsbedingung wiederum der Median einer jeden Vp ein. Diese mittleren Werte erwiesen sich als näherungsweise normalverteilt, so dass eine parametrische Auswertung erfolgen konnte. Eine zweifaktorielle Varianzanalyse mit Messwiederholung ergab einen signifikanten Haupteffekt für den dreistufigen Faktor „Darstellung der Lotteriewahrscheinlichkeit“ ($F = 21,7; df = 2; p < .001$); für den zweiten Faktor „Gewinn/Verlust“ zeigte sich ein noch stärkerer Haupteffekt ($F = 75,0; df = 1; p < .001$). Die Interaktion war nicht signifikant ($F = .4; df = 2; p = .66$). Die Mittelwerte sind der Abb. 44 zu entnehmen. Wie schon in Experiment II brauchten die Vpn für Verlustaufgaben deutlich länger als für Gewinnaufgaben; der Unterschied betrug hier im Mittel 8,6 Sekunden oder 39%. Wie deutlich zu erkennen ist, geht der Effekt von „Darstellung der Lotteriewahrscheinlichkeit“ allein darauf zurück, dass die Vpn deutlich mehr Zeit für die Aufgaben mit verbalen Wahrscheinlichkeitsbedingungen brauchten, im Mittel 5,5 Sekunden, was 22% entspricht.

Um zu überprüfen, ob die Art der Wahrscheinlichkeitsbeschreibung einen Einfluss auf die Bearbeitungszeit hatte, wurden die mittleren Bearbeitungszeiten unter der Bedingung „verbal“ einer zweifaktoriellen gemischten Varianzanalyse unterzogen. Die verwendete Liste der verbalen Beschreibungen bildete dabei einen Zwischensubjektfaktor, „Gewinn/Verlust“ ging als Messwiederholungsfaktor ein. Selbstverständlich war der Faktor „Gewinn/Verlust“ signifikant ($F = 40,1; df = 1; p < .001$). Ein signifikanter Effekt zeigte sich jedoch auch für die verwendete Liste ($F = 7,2; df = 2; p = .002$) und ebenso für die mäßig stark ausgeprägte Interaktion ($F = 4,2; df = 2; p = .02$). Abb. 45 zeigt, dass Liste 3 zu schnelleren Bearbeitungszeiten führte als Liste 2, diese wiederum zu kürzeren Zeiten als Liste 1; die Effekte waren vor allem im Verlustbereich ausgeprägt. Es ist denkbar, dass diese Unterschiede auf unterschiedliche Grade von Ambiguität zurückgehen, dass also etwa die Umbeschreibungen aus Liste 1 als vager empfunden wurden als die aus Liste 3; die Venture Theorie würde einen solchen Zusammenhang zwischen Vagheit und Bearbeitungszeit nahelegen. Um dies zu überprüfen, ging für jede Vp die mittlere wahrgenommene Vagheit als abhängige Variable in eine Varianzanalyse mit dem dreistufigen Faktor „Liste“ ein; letzterer hatte in der Tat einen signifikanten Einfluss auf die subjektive Vagheit ($F = 4,7; df = 2; p = .014$) in der erwarteten Richtung: Die Beschreibungen aus Liste 1 wurden als am vagesten empfunden, diejenigen aus Liste 3 als am wenigsten vase. Die mittleren Bearbeitungszeiten unter den drei verschiedenen Listen korrelierten fast perfekt mit deren mittlerer wahrgenommener Vagheit ($r = .96; n = 3, p = .19$). Somit erscheint es plausibel, die Ambiguitätsunterschiede zwischen den Listen als ursächlich für die unterschiedlichen Bearbeitungszeiten anzusehen.

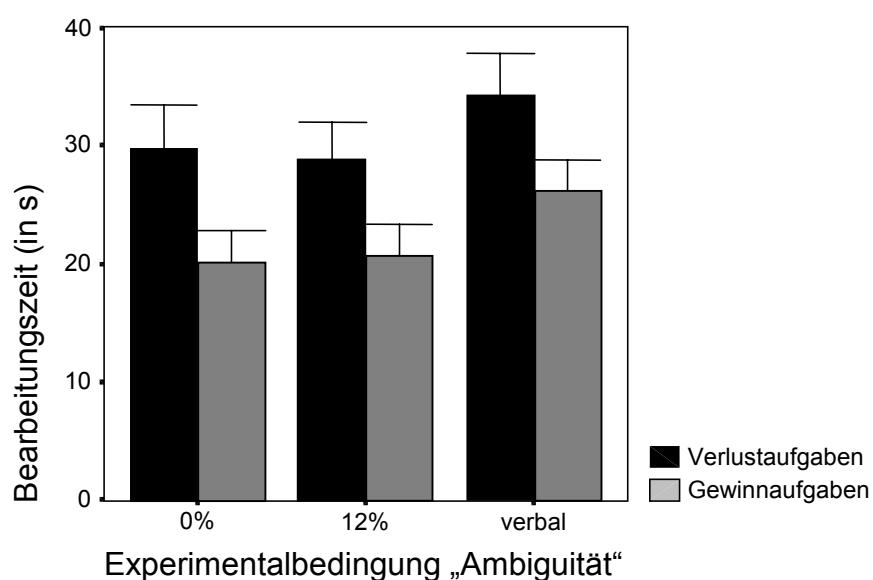


Abb. 44 Experiment III: Durchschnittliche Bearbeitungszeiten unter den einzelnen Versuchsbedingungen. Die Winker markieren jeweils +2 Standardfehler.

Zumindest in der Rückschau ist leicht verständlich, dass die Wahrscheinlichkeitsbeschreibungen der Liste 1 als vager angesehen wurden; diese Liste bediente sich nämlich nur sechs verschiedener

Formulierungen, um den Wahrscheinlichkeitsraum von .03 bis .97 abzudecken, im Gegensatz zu dreizehn und achtzehn Begriffen der anderen Listen. Somit war für die Vpn klar, dass hier jeder Begriff einen relativ breiten Range abdecken musste, während dies für die Begriffe der anderen beiden Listen nicht galt.

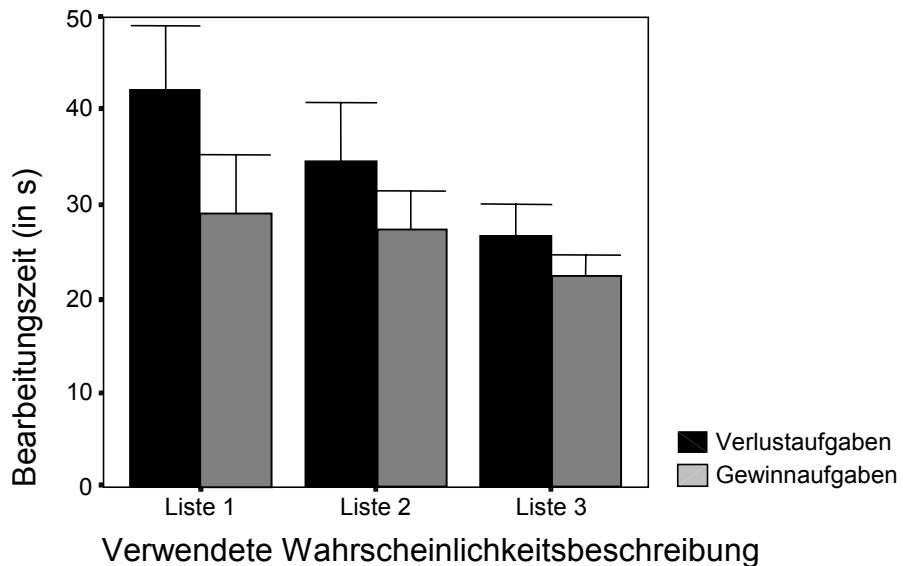


Abb. 45 Experiment III: Durchschnittliche Bearbeitungszeiten in Abhängigkeit von der verwendeten Liste mit Wahrscheinlichkeitsbeschreibungen. Die Winker entsprechen jeweils +2 Standardfehlern.

Im Prinzip könnte auch eine andere Größe als die wahrgenommene Ambiguität für die längeren Bearbeitungszeiten unter Liste 1 verantwortlich gewesen sein; die Tatsache nämlich, dass Vpn hier innerhalb einer Aufgabe auf Grund der niedrigen Zahl der verwendeten Beschreibungen häufiger auf Lotterien trafen, die dasselbe Wahrscheinlichkeitslabel trugen. Dies sollte in meinen Augen aber die Bearbeitung allenfalls beschleunigen, da diese Lotterien dann nur noch auf der Wertdimension miteinander verglichen werden mussten.

Zusammengefasst fanden sich zwei grundlegende Unterschiede zwischen den Bearbeitungsdauern graphisch und verbal dargestellter Lotterien. Letztere waren länger; zudem zeigte sich hier ein Zusammenhang zur Ambiguität der Wahrscheinlichkeitsinformation – Listen mit höherer subjektiv wahrgenommener Ambiguität führten zu längeren Zeiten; ein solcher Effekt fand sich für graphisch dargestellte Lotterien weder in Experimenten II noch in Experiment III.

Es gab keinen Zusammenhang zwischen Bearbeitungszeit und Leistung dergestalt, dass die Teilnehmer, die sich mehr Zeit gelassen hatten, auch bessere Leistungen erbracht hätten ($\rho = -.17$; $n = 53$; $p = .21$). Wie schon in Experiment II wird auch hier nahegelegt, dass sich ein erhöhter kognitiver Aufwand, der eine der Ursachen für längere Bearbeitungszeiten darstellen dürfte, nicht ausgezahlt hat¹⁷. Enge Zusammenhänge zwischen der Bearbeitungszeit und der Leistung gab es jedoch wieder auf Ebene der Versuchsbedingungen. Die Bedingungen, in denen schnell gearbeitet

¹⁷ Mangelnde Variabilität in den Zeiten lag nicht vor, im Gegenteil waren die Unterschiede in den Bearbeitungszeiten beträchtlich. So lagen die Zeiten des neunzigsten Perzentils mehr als doppelt so hoch wie die des zehnten.

wurde, führten auch zu besseren Leistungen ($r = .88$; $n = 6$; $p = .019$). Gleichzeitig waren dies auch die Bedingungen, in denen sich das Verhalten der Versuchsteilnehmer am besten replizieren ließ ($r = .98$; $n = 6$; $p < .001$). Für Verlustaufgaben sowie die Gewinnaufgaben in der Bedingung „verbal“ galt also, dass sie schwieriger waren, insofern sie nämlich zu weniger geordnetem Verhalten, schlechteren Leistungen und längeren Bearbeitungszeiten führten.

11.4.6 Wodurch wurden die Leistungen der Vpn bestimmt?

Die nähere Analyse der Leistungen erfolgte wieder auf Rangdatenniveau. Unter keiner der sechs Versuchsbedingungen war ein Zusammenhang zwischen Bearbeitungszeit und Leistung zu finden. Hingegen konnte über θ_r und Δ_r wiederum ein beträchtlicher Teil der interindividuellen Leistungsvarianz aufgeklärt werden. Für alle sechs Versuchsbedingungen getrennt war ein Regressionsmodell zur Vorhersage von SFE_r auf Grundlage von θ_r und Δ_r erstellt worden. Beide Prädiktoren wurden in allen sechs Modellen signifikant; bis auf die Bedingung „0% Ambiguität/Gewinn“, in der θ_r und Δ_r einen (allerdings nur mäßigen) Zusammenhang zeigten ($r = .33$; $p = .016$), erwiesen sich die Prädiktoren als unabhängig voneinander, so dass Tab. 7 die einfachen Korrelationen mit SFE_r wiedergibt.

Tabelle 7: Experiment III – Korrelationen von θ_r und Δ_r mit SFE_r .

		0% Ambiguität	12% Ambiguität	verbal
Verlust	Δ_r	.73	.50	.54
	θ_r	.33	.45	.55
Gewinn	Δ_r	.41 ^a	.56	.40
	θ_r	.82 ^a	.67	.28

a: In dieser Bedingung waren Δ_r und θ_r nicht unabhängig voneinander, sondern korrelierten mit .33.

Zwischen 26% („verbal/Gewinn“) und 70% („0% Ambiguität/Gewinn“ und „0% Ambiguität/Verlust“) der interindividuellen Leistungsvarianz konnten so aufgeklärt werden. Über alle Versuchsbedingungen hinweg dämpften ein zu hohes Augenmerk auf die Wahrscheinlichkeitsdimension und mangelnde Konsistenz im Verhalten die Leistung etwa gleich stark. Wie schon in den beiden Vorgängerexperimenten galt auch hier, dass sich individuelle Unterschiede in der Gewichtung der Wahrscheinlichkeitskomponente eher unter denjenigen Versuchsbedingungen auf die Leistung auswirkten, in denen sich ein konsistentes Versuchspersonenverhalten zeigte (vgl. hierzu Abb. 42).

12. FAZIT AUS DEN DREI EXPERIMENTEN

Kapitel 12 zieht zusammenfassend Bilanz aus den Experimenten I – III.

Wie lassen sich die Ergebnisse der drei Experimente zusammenfassen? Alle vorliegenden Befunde sprechen konsistent gegen die Gewichtungshypothese. Das auf der Grundlage methodologisch unzureichender Experimente basierende spezifischere Postulat, dass die verbale an Stelle der numerischen Beschreibung von Wahrscheinlichkeiten zu einer stärkeren Orientierung der Entscheider an der Wertdimension führt (Gonzalez-Vallejo et al., 1994; Wallsten, Budescu & Tsao, 1997), muss ebenfalls zurückgewiesen werden. Experiment III hat zwar diesen Vergleich nicht unmittelbar gezogen, sondern an Stelle numerischer Wahrscheinlichkeitsinformation nichtambige graphische gesetzt, aber diese beiden Modi sind nicht nur logisch äquivalent, sondern empirische Untersuchungen belegen auch deren faktische Bedeutungsgleichheit; so zeigte Wallsten (1971), dass graphisch dargestellte Gewinnwahrscheinlichkeiten akkurat eingeschätzt werden, und Budescu und Weiss (1987) demonstrierten, dass numerisch und graphisch dargestellte Chancen zu sehr ähnlichen Entscheidungsmustern führen. Genau konträr zur Gewichtungshypothese führte hohe graphisch erzeugte Ambiguität im Gewinnbereich dazu, dass sich die Vpn verstärkt an der Wahrscheinlichkeitsdimension orientierten. Das in Experiment II gefundene Absinken von θ sowie die Befunde zu den Bearbeitungszeiten in diesem Versuch legen den Schluss deutlich nahe, dass dieser Effekt darauf zurückzuführen ist, dass hohe Ambiguität eine vorsichtigere Einstellung gegenüber Risiko nach sich zog.

Für die Qualität der vorliegenden Daten spricht, dass typische Befunde auf dem Gebiet von Entscheidung unter Unsicherheit durchweg repliziert werden konnten: In Experiment I fand sich das generell im Gewinnbereich typische risikomeidende Verhalten; auch in den Experimenten II und III agierten die Vpn erwartungsgemäß bei Gewinnaufgaben risikomeidender als im Verlustbereich. Ebenfalls typisch ist der Befund der Experimente II und III, dass Entscheidungen, die potenzielle Verluste betrafen, länger brauchten als solche, die Gewinnoptionen betrafen. Die von den Vpn vorgenommene Quantifizierung verbaler Ambiguität kam zu denselben Ergebnissen, die in der Literatur genannt werden. Wo immer dies möglich war, passten sich die Ergebnisse also nahtlos in den Kontext bestehender Forschungsergebnisse ein, was für die vorliegenden Experimente spricht. Für deren Güte spricht ebenso, dass die Versuchspersonen der gestellten Aufgabe nicht nur nachkommen konnten sondern durchweg sehr gute Leistungen erbrachten: Dies zeigt etwa ein direkter Vergleich mit den Resultaten des Experimentes von Gonzalez-Vallejo et al. (1994), in dem die Vpn eine Aufgabe zu lösen hatten, die fast identisch mit denen der hier vorgelegten Experimente ist. In der Studie von Gonzalez-Vallejo und Mitarbeitern lag der maximal zu erzielende Erwartungswert bei 12,5\$ während ein rein zufallsgesteuertes Verhalten zu einem Erwartungswert von 9,9\$ führte. Setzt man den ersten Wert mit einer Leistung von 100% gleich und letzteren mit 0%, so lag die mittlere Leistung der Vpn von 10,78\$ bei 34%. In der verbalen Gewinnbedingung von Experiment III lag der entsprechende Wert bei 83% (in den Experimenten I und II erreichten die Vpn selbst bei 85% deutlich höhere Leistungen als die Vpn von Gonzalez-Vallejo et al., nämlich 51% und 54%).

Mit den guten Leistungen der Vpn in den hier dargestellten Experimenten verknüpft war, dass deren Verhalten insgesamt konsistent und gut replizierbar war. Der Replikationsfehler Δ war in allen Experimenten niedrig, und in allen drei Versuchen erwies sich dasselbe Basismodell am besten geeignet, das Verhalten der Teilnehmer nachzuzeichnen. Die Konsistenz der Schätzwerte für θ erwies sich als zufriedenstellend, und eine alternative Auswertungsstrategie, die zur Validierung der θ -Schätzer in den Experimenten I und II herangezogen worden war, führte zu denselben Schlussfolgerungen. Auch diese Tatsachen sprechen für die Qualität der Versuchsdaten.

Während die hier vorliegenden Experimente Wahlen zwischen Optionen gleicher Ambiguität untersuchten, befassen sich die meisten Untersuchungen, die Entscheidungen bei ambiger Wahrscheinlichkeitsinformation zum Thema haben, mit Wahlen zwischen Optionen, die im Grad ihrer Ambiguität variieren. Typischerweise zeigt sich dabei eine Präferenz für die weniger ambigen Optionen (einen Überblick geben Camerer & Weber, 1992). In diesem Paradigma induziert Ambiguität also in der Regel ein konservativeres Verhalten. Die Interpretation, dass die in den Experimenten I und II gefundenen Effekte von Ambiguität auf θ ebenfalls darauf zurückzuführen sind, dass die Vpn hier bei steigender Unsicherheit vorsichtiger agierten, fügt sich gut in dieses Bild.

Der Befund, dass graphisch erzeugte Ambiguität tendenziell ein konservativeres Verhalten nach sich zog, ließ sich für verbal erzeugte Ambiguität nicht finden. Ob dies daran lag, dass die den Wahrscheinlichkeitsaussagen inne wohnende Unsicherheit zu gering war oder der Effekt tatsächlich an die graphische Erzeugung von Ambiguität gebunden bleibt, kann nicht entschieden werden, da die Ambiguität der Wahrscheinlichkeitsbeschreibungen nicht ohne weiteres so gemessen werden kann, dass sie eindeutig mit der graphisch erzeugten verglichen werden kann.

Die Personen, die die Wahrscheinlichkeitsbeschreibungen für Experiment III lieferten hatten, waren um eine hohe Verständlichkeit der von ihnen gegebenen Information bemüht. Wenn Menschen im Alltag Grade subjektiver Sicherheit beschreiben, mag dies nicht immer der Fall sein. Denn während die Aussagen im Experiment eine präzise Grundlage hatten, nämlich graphisch dargestellte Wahrscheinlichkeiten, gründen entsprechende Alltagsaussagen selten auf genauen Informationen. Somit neigen Menschen dann vielleicht eher dazu, Zuflucht zu schwammigeren Begriffen zu nehmen (an Vagheit kaum zu überbieten ist etwa „möglich“), anstatt möglichst genaue Beschreibungen zu geben. In dieselbe Richtung könnte auch das Bestreben wirken, sich – etwa wenn es um eine Prognose geht – nicht festnageln zu lassen. Dafür, dass Sender ein hohes Maß an Genauigkeit scheuen, spricht zumindest, dass sie entsprechende Information lieber verbal als numerisch abgeben (Erev & Cohen, 1990; Wallsten, Budescu, Zwick & Kemp, 1993). Es besteht also die Möglichkeit, dass im Alltag unschärfere Formulierungen gebraucht werden als dies in Experiment III der Fall war; diese könnten wiederum ein konservativeres Gewichtungsverhalten nach sich ziehen. Im letzten Abschnitt dieser Arbeit werde ich auf die Effekte, die dies hätte, noch einmal eingehen.

13. WAS KOSTET AMBIGUITÄT, WAS KOSTEN SUBOPTIMALE ENTSCHEIDUNGSSTRATEGIEN? SIMULATION VIII

Kapitel 13 wendet sich dem Aspekt der Leistung zu. Mit Hilfe einer Simulation wird erklärt, wie gut sich das von den Vpn gezeigte Entscheidungsverhalten unter einer Fülle von Randbedingungen bewährt. Die Ergebnisse der Simulation weisen eine Reihe praktischer Implikationen auf, die in 13.3.3 diskutiert werden.

Aus logischen Gründen bin ich bislang nur auf die relativen Leistungen der Vpn (gemessen in SFE) eingegangen. Ich habe also nie die Frage gestellt, wieviele Punkte sie unter verschiedenen Bedingungen erzielt haben, sondern statt dessen gemessen, wieviele „Verwechslungsschritte“ die Vpn von hinter einem angemessenen Leistungsmaßstab zurückblieben. Die ökonomische Frage, wie stark die Ungenauigkeit von Wahrscheinlichkeitsinformation die Qualität von Entscheidungen beeinträchtigt, ist aber aus sich heraus interessant. Ihr soll der Schluss dieser Arbeit gewidmet sein.

Ich möchte mich dabei nicht darauf beschränken, die aktuellen Ergebnisse der drei Experimente zu betrachten. Vielmehr soll in einem ersten Schritt das von den Vpn gezeigte Verhalten möglichst genau beschrieben werden. In zwei wesentlichen Aspekten ist dies bereits geschehen: *Normal* ist in besonderer Weise geeignet nachzuzeichnen, wie Ambiguität aufgelöst wurde; und zum zweiten liegen Schätzwerte für θ vor. Offensichtlich wurde das Verhalten der Vpn aber zusätzlich durch eine Fehlerkomponente geprägt – sie soll hier *error* heißen. Für die Existenz dieser Fehlerkomponente spricht erstens, dass sich das Verhalten der Vpn nicht hundertprozentig genau replizieren ließ (die Δ -Werte waren stets größer null), und zweitens gab es keine Vp, die die Leistung, die die Strategie *Normal* erbracht hätte, erreichte. Die Fehlerkomponente *error* gilt es also zunächst zu modellieren, dies geschieht im nächsten Abschnitt 13.1.

Wenn dies geschehen ist und somit ein realistisches Modell des Vpn-Verhaltens erstellt wurde, soll dieses dann in einem zweiten Schritt mittels einer Computersimulation auf verschiedenste Aufgaben übertragen werden, die eine Vielzahl von Randbedingungen widerspiegeln, deutlich mehr, als man, wegen des damit verbundenen Aufwandes, experimentell untersuchen könnte. Diese Randbedingungen werden später, im Abschnitt 13.2, genauer beschrieben. Grundsätzlich betrachtet die Simulation jedoch Aufgaben, die den in den Experimenten verwendeten ähnlich sind.

Mit Hilfe der Simulation sollen fünf Fragen beantwortet werden:

- Welchen Einfluss auf Leistungen hat die insbesondere bei hoher Ambiguität gefundene Änderung von θ ? Welchen Einfluss auf die Leistung hätte es dem gegenüber, wenn θ mit steigender Ambiguität sinken würde, wie die Gewichtungshypothese vermutete?
- Wie bereits in 9.3.2 angedeutet wurde, ist *Normal* nicht die beste Art, Ambiguität aufzulösen. Inwieweit wird die Leistung also dadurch gemindert, dass Ambiguität nicht optimal aufgelöst wird?
- Des Weiteren muss natürlich *error* die Leistung schmälern. Wie stark wirkt sich diese Fehlerkomponente aus?

– Die letzten beiden Fragen zielen nicht auf Strategieaspekte sondern auf den Einfluss der Ambiguität selbst: Inwieweit wird die Güte von Entscheidungen gemindert, wenn die Wahrscheinlichkeitsinformation unpräzise ist?

– Und schlussendlich die noch etwas spezifischere Frage, welcher Einfluss auf die Leistung ist von der Ambiguität, die verbalen Wahrscheinlichkeitsbeschreibungen inne wohnt, zu erwarten?

Die letztgenannte Frage ist vor allem deswegen interessant, weil Menschen wie bereits erwähnt dazu neigen Wahrscheinlichkeitsurteile in verbaler Form zu geben und nicht numerisch (Wallsten, Budescu, Zwick & Kemp, 1993). Dass die Bedeutung solcher in Worte gekleideter Wahrscheinlichkeitsaussagen nicht klar umrissen ist, braucht dabei nicht problematisch zu sein, da die zu kommunizierenden Wahrscheinlichkeiten oft auch dem Sender nicht genau bekannt sind. Die Verwendung einer einzelnen Zahl würde hier oft nur einen ungerechtfertigten Grad an Präzision vortäuschen. Problematisch ist jedoch, dass verschiedene Menschen denselben Ausdrücken systematisch unterschiedliche Bedeutungen beimessen (z. B. Mosteller & Youtz, 1990). Wenn also ein Sender seine mehr oder minder genaue Schätzung einer Wahrscheinlichkeit in Worten vermittelt, so sollten wir erwarten, dass die Schätzung, die der Empfänger auf Grundlage dieser Information macht, im Durchschnitt mit einem stärkeren Fehler behaftet ist als die des Senders. Darunter sollte dann die Qualität von Entscheidungen, die der Empfänger auf der Grundlage dieser Information trifft, leiden. Während es vereinzelt Stimmen gibt, die davon ausgehen, dass Menschen dennoch auf der Grundlage verbaler Wahrscheinlichkeitsinformation zu besseren Entscheidungen kommen (Zimmer, 1983), plädieren viele Autoren für eine Veränderung unserer alltäglichen Praxis. So schlagen etwa Mosteller und Youtz (1990) vor, verbale Wahrscheinlichkeitsbeschreibungen zu normieren und Beyth-Marom (1982) spricht sich dafür aus, auf verbale Umschreibungen zu verzichten und statt dessen auf die Angabe von Wahrscheinlichkeitsintervallen auszuweichen.

Mit der Beantwortung der letzten der oben aufgeführten fünf Fragen möchte ich versuchen abzuschätzen, welcher Gewinn von solchen Maßnahmen zu erwarten wäre.

13.1 Ein realistisches Modell menschlichen Verhaltens – Simulation VII

Wie bereits ausgeführt, ist es für die Simulation notwendig, das Verhalten der Vpn adäquat modellieren zu können. Was hierzu noch fehlt, ist die Strategie *Normal* um eine Fehlerkomponente *error* zu ergänzen. Eine solche lässt sich an vielen Stellen in das Modell integrieren, etwa bei der Wahrnehmung der Gewinnwahrscheinlichkeit oder bei der Verrechnung der subjektiv erwarteten Gewinnwahrscheinlichkeit mit dem Lotteriewert. Ich habe mich willkürlich für die letztgenannte Variante entschieden.

Der Algorithmus, der das Vpn-Verhalten modelliert, soll *Real* heißen, und er ordnet die Lotterien gemäß ihrer wahrgenommenen Attraktivität. Diese ergibt sich dabei wie folgt:

$$Real \quad \text{Attraktivität} = \epsilon \cdot (P_{Sicht} + Ambig/2)^0 + error \cdot \epsilon \cdot (P_{Sicht} + Ambig/2)^0,$$

wobei θ jeweils so gewählt werden kann, dass es den empirisch gefundenen θ -Werten entspricht; *error* ist eine gleichverteilte Zufallsvariable mit dem Erwartungswert null. D. h. die Attraktivität bestimmt sich zunächst gemäß der Strategie *Normal* (erster Summand); von dieser Bewertung wird dann aber um einen zufälligen Prozentwert nach oben oder unten abgewichen (zweiter Summand) – es macht sich also ein gewisser Schätzfehler bemerkbar. Wie hoch sollte nun dieser Schätzfehler sein? Generell führen höhere Werte von *error* dazu, dass erstens die Leistung sinkt und dass zweitens das Verhalten weniger gut replizierbar wird, Δ also ansteigt. *Error* ist also dann optimal gewählt, wenn damit sowohl die empirisch gefundenen Δ -Werte als auch die in den Experimenten erbrachten Leistungen wiedergegeben werden.

Simulation VII

Welcher Wert für *error* ist angemessen, um das Verhalten der Vpn realitätsgerecht nachbilden zu können? Um diese Frage beantworten zu können, habe ich mich einer Computersimulation bedient.

Zweihundertfünfzig künstlich erzeugte „Vpn“ bearbeiteten dabei vierzig individuell verschiedene Gewinnaufgaben, jeweils zehn auf dem Ambiguitätsniveau 0%, 25%, 55% und 85%. Die Aufgaben wurden dabei nach derselben Prozedur erzeugt wie in den Experimenten I und II. Die „Vpn“ bearbeiteten diese Aufgaben so, dass sie strikt der oben beschriebenen Strategie *Real* folgten. Die θ -Werte waren dabei nicht für alle „Vpn“ gleich, sondern sie wurden auf jedem Ambiguitätsniveau so gewählt, dass sich eine Verteilung ähnlich derjenigen ergab, wie sie in den gepoolten Daten der Experimente I und II zu finden war. *error* fungierte als freier Parameter, der so lange verändert wurde, bis die simulierten Vpn ein Verhalten erzeugten, das demjenigen der realen Versuchsteilnehmer möglichst ähnlich war.

Die Antwortdaten der künstlichen Vpn wurden derselben Auswertungsprozedur unterzogen, die auch die empirischen Daten durchlaufen hatten, d. h. für jede der 250 „Vpn“ wurde θ auf der Grundlage von *Normal* geschätzt. Da das Verhalten durch den Zufallsfehler *error* mitbeeinflusst war, konnte es natürlich nicht fehlerfrei reproduziert werden, sondern es ergaben sich auch für die simulierten Vpn Δ -Werte größer null. Für verschieden starke Zufallseinflüsse, *error*, konnte somit überprüft werden, inwieweit sie zu Δ -Werten und Punktgewinnen führten, die den gepoolten Ergebnissen der Experimente I und II entsprachen (von Experiment II flossen dabei nur die Daten für den Gewinnbereich ein). Es zeigte sich, dass sich beide Kennmarken der empirisch gewonnenen Daten auf allen vier Ambiguitätsniveaus sehr gut reproduzieren ließen, wenn für *error* eine Ober- bzw. Untergrenze von $\pm .55$ gewählt wurde (vgl. Abb. 46).

In der nachfolgenden Simulation VIII war *Real* daher derart gestaltet, dass für *error* genau diese Grenzen galten. Das hatte zur Folge, dass von der „wahren“ Attraktivitätseinschätzung einer Lotterie, die der erste Summand von *Real* widerspiegelt, im Durchschnitt um 27,5% nach oben oder unten abgewichen wurde.

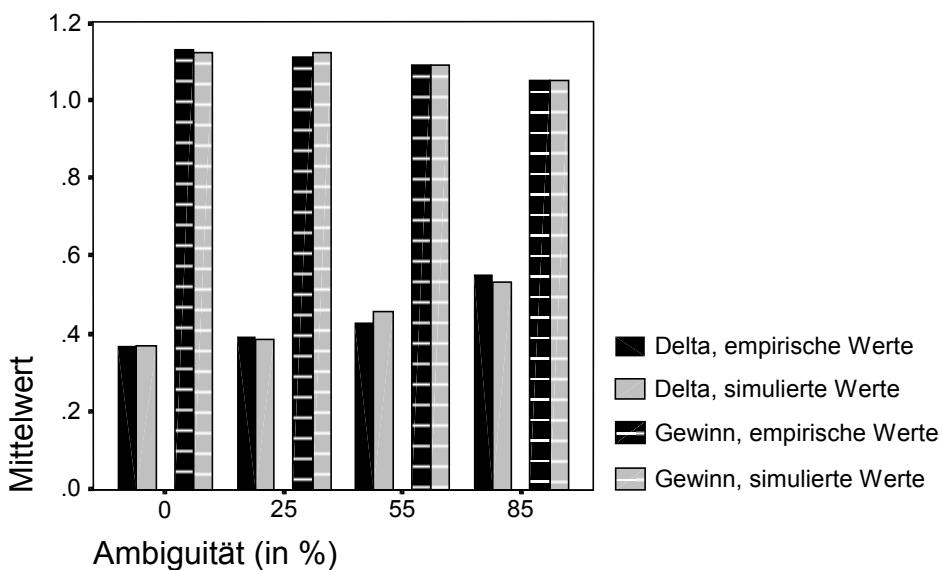


Abb. 46 Simulation VII: Wird in *Real* für *error* eine Ober- bzw. Untergrenze von $\pm .55$ gewählt, so lassen sich die empirisch gefundenen Gewinne und Δ -Werte hervorragend wiedergeben. Die Gewinne wurden hier abbildungsgerecht transformiert.

13.2 Beschreibung der Simulation VIII

In der Simulation fällten drei virtuelle Entscheidungsfinder Entscheidungen bezüglich einfacher Lotterien, ähnlich denen der hier berichteten Experimente. Die Lotterien wurden auf dieselbe Weise erzeugt wie für Simulation I beschrieben (vgl. 4.3), mit dem einzigen Unterschied, dass die Lotteriewerte ϵ aus dem veränderlichen Intervall $[\epsilon_{\min}, \epsilon_{\max}]$ gezogen wurden (s. u.) und nicht aus dem Intervall [3, 97]. Die den „Entscheidungsfindern“ zugängliche Information über P war wiederum entweder präzise oder (mehr oder weniger) ambige. Ambiguität wurde dabei wieder darüber erzeugt, dass an Stelle von P zwei Intervallgrenzen P_{lower} und P_{upper} angegeben wurden, innerhalb derer P lag (der genaue Prozess ist unter 4.3 beschrieben). Der Grad der Ambiguität α ergab sich dabei als Differenz dieser beiden Werte. Wie auch in den Experimenten formte eine Reihe von Lotterien gemeinsam eine Aufgabe. Alle Lotterien einer Aufgabe hatten dabei dasselbe Ambiguitätsniveau. Drei „Entscheidungsfinder“ hatten bezüglich der Lotterien einer Aufgabe eine Entscheidung zu treffen; dabei folgte jeder der drei einer anderen Strategie – diese werden nachher genau vorgestellt. Als abhängige Variable wurden die Erwartungswerte für die Punktzahlen analysiert, die sich für die Wahlen der drei künstlichen Entscheidungsfinder ergaben.

13.2.1 Berücksichtigte unabhängige Variablen

Die Variablen, die in der Simulation systematisch variiert wurden, sollen im Folgenden vorgestellt werden. Deren wichtigste war natürlich α , der *Grad der Ambiguität*; er variierte von 0 bis .9 in Stufen von .1. Zusätzlich wurden fünf andere Variablen betrachtet, von denen zu vermuten ist, dass sie eine

vermittelnde Wirkung auf den Einfluss ausüben, den die Ambiguität der Wahrscheinlichkeitsinformation auf die Entscheidungsgüte hat. Zwei davon bezogen sich auf die verwendete Entscheidungsstrategie, die anderen drei beschrieben Aufgabencharakteristika. Die letztgenannten sollen zuerst vorgestellt werden.

Zwei verschiedene *Aufgabentypen* wurden berücksichtigt. Äquivalent zur Aufgabe der Vpn in den Experimenten erforderte die *Rankingaufgabe*, dass alle Lotterien einer Aufgabe in eine Rangreihe gebracht wurden. Der Punktwert ϵ der gewinnenden Lotterien wurde multipliziert mit dem vergebenen Rangplatz dieser Lotterie gutgeschrieben. Die Rankingaufgabe beschreibt in ihrer Struktur solche Entscheidungsprobleme, bei denen mehrere Optionen gleichzeitig aber mit unterschiedlichem Elan verfolgt werden. Denken wir etwa an eine Managerin, die für die Entwicklung neuer Produkte verantwortlich ist. Sie wird in der Regel gut beraten sein, wenn sie nicht alle Ressourcen der vielversprechendsten Idee zukommen lässt sondern statt dessen eine Vielzahl von Entwicklungsansätzen fördert, wobei die Höhe der Mittel auf den wahrgenommenen Chancenreichtum des jeweiligen Ansatzes abgestimmt ist. Ähnlich wird Leuten, die ihr Geld in Aktien investieren möchten, geraten, ihr Kapital zu streuen und nicht alles auf ein Pferd zu setzen.

Der zweite untersuchte Aufgabentyp erforderte, dass von allen Lotterien einer Aufgabe lediglich eine ausgewählt wurde. Diese wurde dann gespielt, und im Gewinnfall wurde ihr Punktwert ϵ gutgeschrieben. Die zuletzt beschriebene *Wahlauflage* gleicht also in ihrer Struktur solchen Alltagsentscheidungen, bei denen nur eine der offenstehenden Handlungsalternativen wahrgenommen wird. Wir können etwa an einen Patienten denken, der überlegt, ob er sich einer medizinischen Routineuntersuchung unterziehen soll oder nicht. Oder an die Mitglieder eines Firmenvorstandes, die darüber entscheiden müssen, ob eine andere Firma aufgekauft werden soll oder nicht.

Eine weitere Aufgabencharakteristik, die Beachtung erfordert, ist die *Variabilität der Lotteriewerte* ϵ . Je geringer diese bei gleichbleibender Variabilität der Lotteriewahrscheinlichkeiten P ist, desto stärker hängen die Erwartungswerte der Lotterien von P ab. Somit sollte der Informationsverlust bezüglich P , den Ambiguität mit sich bringt, hier gravierendere Konsequenzen haben; steigende Ambiguität sollte hier also zu einer stärkeren Leistungsverschlechterung führen. Gemessen wurde die Variabilität von ϵ als der Erwartungswert für den Quotienten zweier zufällig ausgewählter Lotteriewerte. Die in der Simulation verwendeten Variabilitätsgrade sind der Tabelle 8 zu entnehmen, mitsamt den Ziehungsranges $[\epsilon_{\min}, \epsilon_{\max}]$, die zu diesen führen¹⁸.

Als dritte Aufgabencharakteristik wurde schließlich die *Aufgabengröße*, d. h. die Zahl der Lotterien, die eine Aufgabe bildeten, variiert. Es wurden Aufgaben mit zwei, drei, vier, sechs, acht und zehn Lotterien berücksichtigt.

Die erste der beiden Variablen, die Unterschiede in Entscheidungsstrategien beschreiben, betraf die *Art, in der die Ambiguität aufgelöst wurde*. Neben der bereits beschriebenen Strategie *Real* wurde

¹⁸ Die entsprechenden Erwartungswerte waren über Monte-Carlo Studien ermittelt worden.

Tabelle 8: Die in der Simulation VIII verwendeten zwölf Variabilitätsgrade der Lotteriewerte ϵ .

ϵ_{\min}	ϵ_{\max}	Erwarteter Quotient zweier zufällig ausgewählter Lotteriewerte
3	3	1,0
3	8	1,5
3	11	2,1
3	32	2,4
3	47	2,7
3	67	3,0
3	97	3,3
3	135	3,6
3	190	3,9
3	262	4,2
3	365	4,5
3	670	5,1

auch die optimale Strategie erprobt. Die Kenntnis des auf jedem Ambiguitätsniveau optimal erzielbaren Ergebnisses hilft, die Ergebnisse, die mit *Real* erreicht werden, zu bewerten. Es gilt:

$$\text{Optimal: Attraktivität} = \epsilon \cdot E(P | P_{\text{lower}}, a)^{\theta},$$

wobei $E(P | P_{\text{lower}}, a)$ der Erwartungswert für die Lotteriewahrscheinlichkeit P gegeben die Untergrenze für diese Wahrscheinlichkeit (P_{lower}), sowie gegeben den aktuellen Ambiguitätsgrad a ist. Falls es eine einfache mathematische Beziehung zwischen $E(P | P_{\text{lower}}, a)$ einerseits und P_{lower} sowie a andererseits gibt, so hat sie sich vor meinem begrenzten mathematischen Talent gründlich versteckt. Ich kann diese Beziehung daher nur in graphischer Form im Appendix wiedergeben (Abb. A1). Diese Werte wurden ermittelt, indem auf jedem Ambiguitätsniveau alle möglichen Permutationen von P und P_{lower} gebildet und mit ihren leicht zu berechnenden Auftretenswahrscheinlichkeiten gewichtet wurden.

Schließlich wurde noch als dritte Strategie *Normal* ins Rennen geschickt. *Normal* ist bis auf die Fehlerkomponente identisch mit *Real*. Der Sinn, auch die Leistung von *Normal* zu betrachten, liegt darin, dass dies ermöglicht abzuschätzen, inwieweit der Leistungsabfall von *Real* gegenüber *Optimal* auf die suboptimale Schätzung von P oder auf den Fehlerterm zurückzuführen ist.

Als zweite Strategievariable wurde die *Höhe von θ* einbezogen. Für θ wurden die Werte .33, .4, .5, .67, 1, 1.5, 2, 2.5 und 3 einbezogen. Der Bereich der in den Experimenten gefundenen Werte für θ wurde somit abgedeckt.

Die nachfolgende Tab. 9 gibt noch einmal einen Überblick über sämtliche unabhängigen Variablen.

Tabelle 9: Überblick über die in Simulation VIII variierten Variablen.

Variable	Berücksichtigte Ausprägungen
Grad der Ambiguität	0, .1, .2, .3, .4, .5, .6, .7, .8, .9
Aufgabentypen	Rankingaufgabe, Wahlaufgabe
Variabilität der Lotteriewerte ^a	1.0, 1.5, 2.1, 2.4, 2.7, 3.0, 3.3, 3.6, 3.9, 4.2, 4.5, 5.1
Zahl der Lotterien pro Aufgabe	2, 3, 4, 6, 8, 10
Strategie	<i>Optimal, Real, Normal</i>
Höhe von 0	.33, .4, .5, .67, 1, 1.5, 2, 2.5, 3

a: angegeben als erwarteter Quotient zweier zufällig gezogener Werte – vgl. auch Tab. 8.

13.2.2 Design

Alle Variablenkombinationen wurden berücksichtigt, so dass die Simulation in einem $10 \times 2 \times 12 \times 6 \times 3 \times 9$ Design resultierte (Ambiguitätsniveau, Aufgabentyp, Variabilität von €, Aufgabengröße, Strategie zur Ambiguitätsauflösung, Höhe von 0). Für jede Zelle wurden 50 000 zufällig konstruierte Aufgaben analysiert. Wann immer die Randbedingungen dies gestatteten, enthielten die Zellen identische Aufgaben, also immer dann wenn sie dieselbe Zahl von Lotterien pro Aufgabe und dieselbe Variabilität von € aufwiesen.

13.3 Ergebnisse

Die Darstellung der Ergebnisse gliedert sich in drei Teile. In 13.1 wird ein allgemeiner Überblick über die Wirkung der untersuchten unabhängigen Variablen gegeben. In 13.2 werde ich der Frage nachgehen, ob möglicherweise eine andere Wahl von Parameterwerten ein völlig anderes Bild zeichnen würde. Schließlich werde ich in 13.3 speziell auf die fünf eingangs gestellten Fragen eingehen.

In einem ersten Schritt war es notwendig, die Daten zu normieren. Dies ist für Vergleiche zwischen den beiden Aufgabentypen, verschiedenen Lotterieanzahlen pro Aufgabe sowie verschiedenen Graden der Variabilität von € notwendig. Denn hält man alle anderen Größen konstant, so erbrachte die Rankingaufgabe mehr Punkte als die Wahlaufgabe, Aufgaben mit vielen Lotterien erzielten höhere Gewinne als solche mit wenigen, und ebenso führten Aufgaben, in denen die Variabilität von € hoch war, zu höheren Erträgen als solche, bei denen sie gering war. Daher wurde die Gesamtdatenmenge entlang der Variablen Aufgabentyp, Aufgabengröße und Variabilität von € aufgesplittet. Dies resultierte in 144 Gruppen, deren jede 270 Zellen enthielt. In jeder dieser Gruppen wurde die Zelle, die den höchsten Ertrag erbrachte, auf 100% gesetzt¹⁹. Die Werte der anderen Zellen wurden dann ebenfalls in Prozentwerte transformiert. Alle folgenden Analysen beruhen auf diesen standardisierten Werten.

¹⁹ Es handelte sich dabei immer um die Bedingung 0% Ambiguität, 0 = 1 und *Optimal*, bzw. *Normal* (beide Strategien sind bei fehlender Ambiguität identisch).

Die Erträge der einzelnen Entscheidungen entsprachen deren Erwartungswerten. Der Erwartungswert einer Entscheidung ist natürlich nicht immer ein geeignetes Kriterium, deren Erfolg zu bewerten. Wenn ich etwa die Möglichkeit zwischen zwei Anlagestrategien habe und dabei diejenige, die den höheren Erwartungswert hat, das Risiko des finanziellen Ruins birgt, werden die meisten Leute es als vernünftig ansehen, dieser Strategie nicht zu folgen. Die Wahl des Erwartungswertes als Kriterium ist aber insofern sinnvoll, als dass es per se kein optimales Evaluationskriterium gibt, dem hier verwendeten aber auf der Grundlage des Gesetzes der großen Zahl ein besonderer Charme zu eigen ist (vgl. 2.2). Ein noch sinnhafteres Kriterium ist natürlich der erwartete Nutzen, zu dem die Entscheidung führt. Dieser erfordert die Messung von Nutzenwerten; wenn diese bekannt sind, können die Lotteriewerte der Simulation jedoch als Nutzenwerte interpretiert und die diesen zu Grunde liegenden realen Werte dann berechnet werden. Eine Interpretation der Ergebnisse als erwarteter Nutzen ist also ohne weiteres möglich.

13.3.1 Überblick über die Wirkung der unabhängigen Variablen

Ein Design, bei dem alle Kombinationen von sechs unabhängigen Faktoren verwirklicht werden, produziert eine unabsehbare Fülle möglicher Interaktionen höherer Ordnung. Die Datenanalyse bedarf daher einer Beschränkung. Die Daten wurden separat für alle sechs Kombinationen von Aufgabentyp \times Strategie analysiert. Innerhalb jeder dieser Untergruppen habe ich die Haupteffekte der verbleibenden Variablen untersucht sowie deren einfache Interaktion mit dem Ambiguität niveau. Ich habe generell keine inferenzstatistischen Tests durchgeführt, da aufgrund des hohen N a priori davon auszugehen ist, dass alle Effekte, die hinreichend groß sind, um von Interesse zu sein, auch signifikant sind.

Strategieunterschiede und der Gesamteffekt von Ambiguität

Abb. 47 gibt den Haupteffekt von Ambiguität auf die Leistung wieder. Die Ergebnisse sind dabei getrennt nach Aufgabentyp und Strategie dargestellt. Wie zu sehen ist führte Ambiguität in allen sechs Bedingungen zu einem exponentiell anwachsenden Leistungseinbruch; dieser war für die Wahlaufgabe deutlich stärker ausgeprägt als für die Rankingaufgabe. Während *Optimal* und *Real* auf allen Ambiguitätsstufen praktisch identische Leistungen erbrachten (die maximalen Unterschiede betrugen .6% [.2%])²⁰, erzielte *Real* deutlich schwächere Ergebnisse, im Durchschnitt 6,2% [2,1%] weniger. Durch den Vergleich mit *Normal* wird klar, dass die Leistungseinbußen, die *Real* gegenüber *Optimal* aufwies, fast ausschließlich auf die Wirkung des Fehlerterms zurückgingen, die suboptimale Schätzung von P jedoch fast keinen Einfluss hatte.

Niedrige und mittlere Grade von Ambiguität führten insgesamt bei allen drei Strategien zu nur geringen Leistungseinbußen; verglichen mit der Bedingung 0% Ambiguität betrugen diese bei 30%

²⁰ Die erste Zahl bezieht sich auf die Wahlaufgabe, die in Klammern auf die Rankingaufgabe. Diese Lesart ist im Folgenden beizubehalten.

Ambiguität nur rund 1,7% [$<1\%$], und 50% Ambiguität reduzierte die Leistungen um rund 5% [2%]. Bei 90% Ambiguität brach die Leistung aller drei Strategien um 19% [7%] ein.

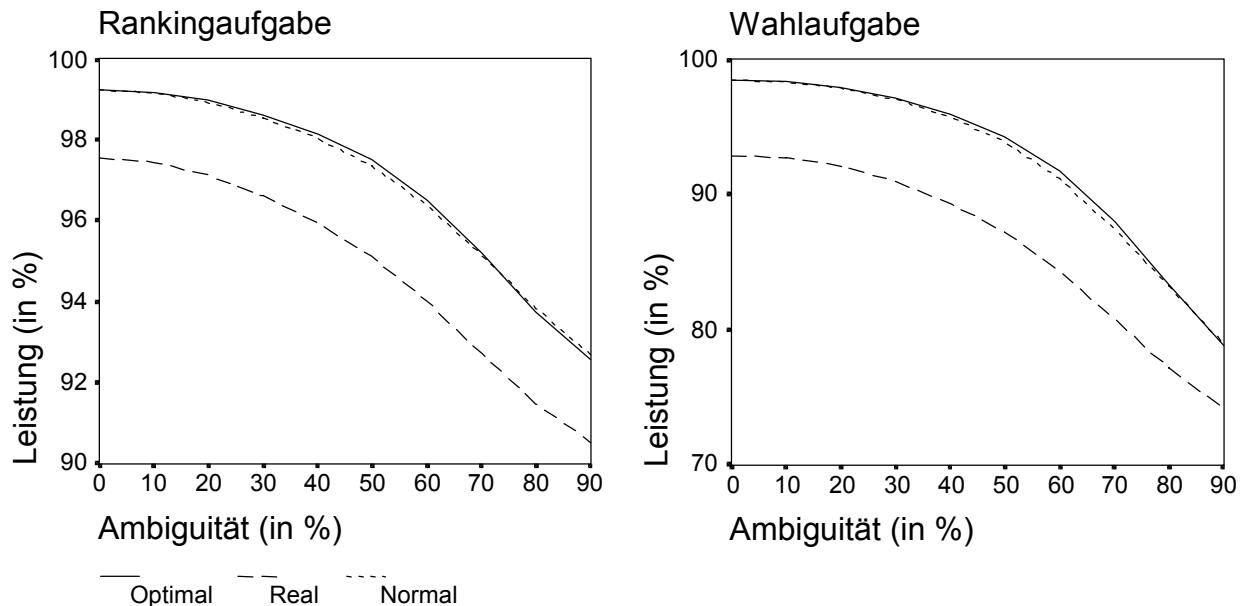


Abb. 47 Simulation VIII: Die Leistungen in Abhängigkeit von der Strategie und vom Grad der Ambiguität, getrennt für Ranking- und Wahlaufgaben.

Bei den Rankingaufgaben erzielte die ϵ -Strategie auf allen Ambiguitätsniveaus eine Leistung von 92,2%. Das bedeutet, dass die Leistung von *Real* ab einer Ambiguität von etwa 75% unter diejenige der ϵ -Strategie sinkt. In den Experimenten I und II war genau dies eingetreten: Die Vpn schnitten bei 55% Ambiguität noch deutlich besser ab als die ϵ -Strategie, bei 85% Ambiguität konnten sie deren Leistung nicht mehr erreichen. Interessanterweise ist dieser Effekt für Wahlaufgaben nicht zu erwarten: Hier erreichte die ϵ -Strategie durchgängig eine Leistung von 72,7%. Die Leistung von *Real* lag selbst bei 90% Ambiguität noch über diesem Level. Zieht man also die Leistung der ϵ -Strategie als Maßstab heran, so sollten Vpn bei hoher Ambiguität im Wahlparadigma deutlich besser abschneiden als im Rankingparadigma (unter der Voraussetzung, dass das Verhalten im ersten ebenfalls durch *Real* gut beschrieben werden kann).

Aus der Abb. 47 wird ersichtlich, dass *Normal* bei extrem hoher Ambiguität ein geringfügig besseres Resultat erzielte als *Optimal*. Dies nimmt zunächst Wunder, da *Optimal* doch immer die genaueren Inferenzen über P anstellt. Der Grund hierfür liegt darin, dass diese Strategie bei extrem hoher Ambiguität auf zu hohe θ -Werte mit einer starken Leistungsverschlechterung reagierte, während *Normal* für Veränderungen von θ relativ unsensibel war, wie Abb. 48 am Beispiel der Wahlaufgabe zeigt.

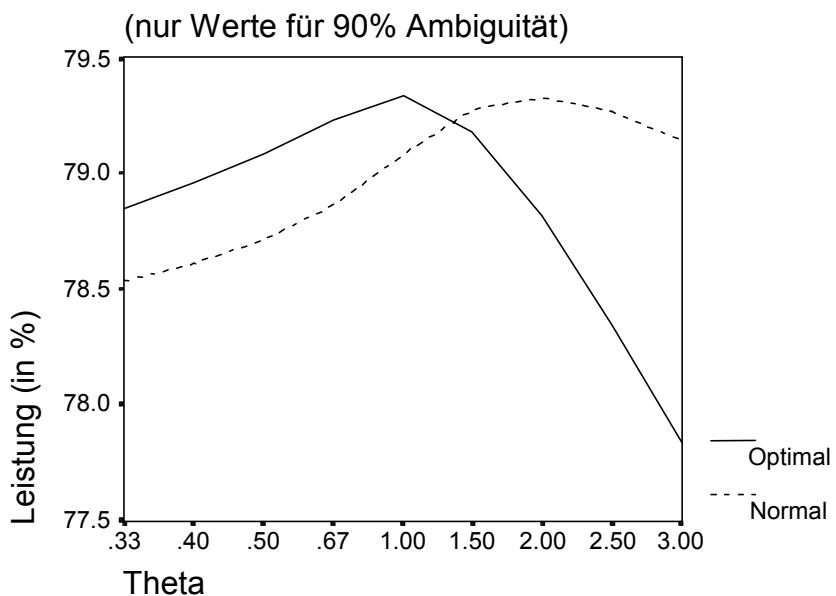


Abb. 48 Simulation VIII: Bei hoher Ambiguität reagiert *Optimal* vergleichsweise empfindlich auf hohe θ -Werte. Daher schneidet bei hoher Ambiguität *Normal* insgesamt besser ab (vgl. Abb. 48), obwohl diese Strategie die Ambiguität nicht auf optimale Art auflöst. (Abb. zeigt die Werte für Wahlausgaben).

Der Einfluss der Variabilität von ϵ

Die Analyse der verbleibenden Variablen ergab, dass diese für die Wahl- und die Rankingaufgabe den gleichen Einfluss hatten, nur dass dieser im Falle der Wahlausgaben stärker war. Deshalb geben die folgenden Abbildungen nur noch die Resultate für diesen Aufgabentyp wieder. Abb. 49 zeigt den

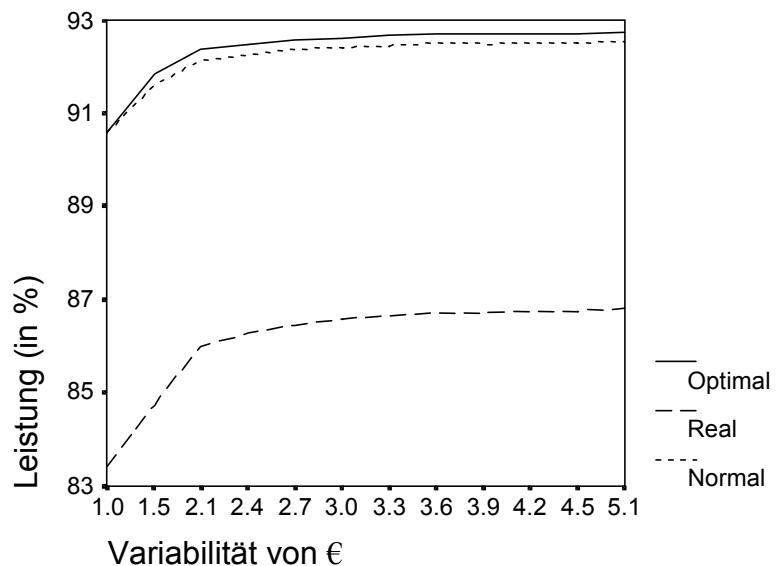


Abb. 49 Simulation VIII: Die Leistung in Abhängigkeit von der Variabilität der Lotteriewerte (vgl. auch Tab. 8). Bei 3,3 ist diese identisch mit der Variabilität der Wahrscheinlichkeiten. (Abb. zeigt die Werte für Wahlausgaben).

Einfluss der Variabilität von ϵ . Bei einem Wert von 3,3 entsprach diese der Variabilität der Lotteriewahrscheinlichkeiten. Wurde die Variabilität von ϵ größer als die von P , hatte dies praktisch keinen leistungssteigernden Effekt (maximale Leistungszuwächse .1%). Nennenswert leistungsmindernd wirkte sich nur aus, wenn die Variabilität von ϵ erheblich geringer ist als die von P war. Bei dem Extremfall einer Variabilität von 1 (d. h. ϵ ist hier bei allen Lotterien konstant) fällt die Leistung zwischen 2% [1%] (*Optimal* und *Normal*) und 3% [3%] (*Real*). Dieser Effekt war bei zunehmender Ambiguität stärker ausgeprägt: Betrachten wir nur die Fälle, in denen diese mindestens 50% betrug, so resultierten Leistungseinbußen von 5% [3%] (*Optimal* und *Normal*) sowie 6% [5%] (*Real*).

Der Einfluss der Aufgabengröße

Die Abb. 50 zeigt den Effekt, den die Anzahl der Lotterien, aus denen eine Aufgabe bestand, auf die Leistung hatte. Wie zu sehen ist, nahm die Leistung bei allen drei Strategien mit steigender Aufgabengröße ab; der Leistungsabfall zeigte dabei einen asymptotischen Verlauf. Mit einer maximalen Leistungsminderung von 11% [3%] war *Real* dabei stärker betroffen als *Optimal* und *Normal*, bei denen die Leistungen rund 5,5% [2%] schwächer ausfielen. Auch der Einfluss der Aufgabengröße war bei höherer Ambiguität stärker ausgeprägt. Betrachtet man wiederum die Aufgaben mit einer Ambiguität von 50% oder mehr, so zeigten sich hier maximale Ertragseinbußen von 15% [5%] (*Real*) und 10% [3%] (*Optimal* und *Normal*).

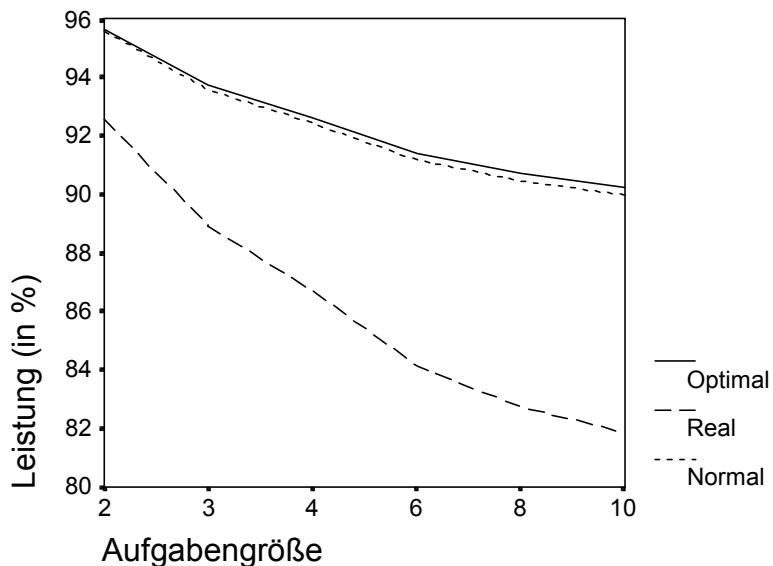


Abb. 50 Simulation VIII: Die Leistung in Abhängigkeit von der Zahl der Lotterien, die eine Aufgabe bilden (WahlAufgaben).

Der Einfluss von θ

Wie aus der Abb. 51 zu entnehmen ist, wirkten sich Veränderungen von θ sehr unterschiedlich auf die Leistung aus, in Abhängigkeit davon, welche Strategie gewählt wurde. Das θ -Optimum für *Optimal* lag bei 1; Abweichungen provozierten Leistungseinbußen von bis zu 5% [4%]. Für *Normal*

lag das Optimum bei 1,5 für *Real* sogar bei 2,5; Abweichungen führten hier zu Verschlechterungen von maximal 4% [2%] bzw. 8% [3%]. Dass das 0-Optimum für *Normal* über 1 liegt, dürfte daran liegen, dass die Schätzungen, die diese Strategie für *P* erbringt, bei steigender Ambiguität einer immer stärkeren Regression zur Mitte unterliegen. Ein θ größer eins arbeitet der damit einhergehenden Variabilitätseinschränkung der Schätzwerte entgegen. Warum das Optimum für *Real* noch höher lag, kann ich nicht erklären.

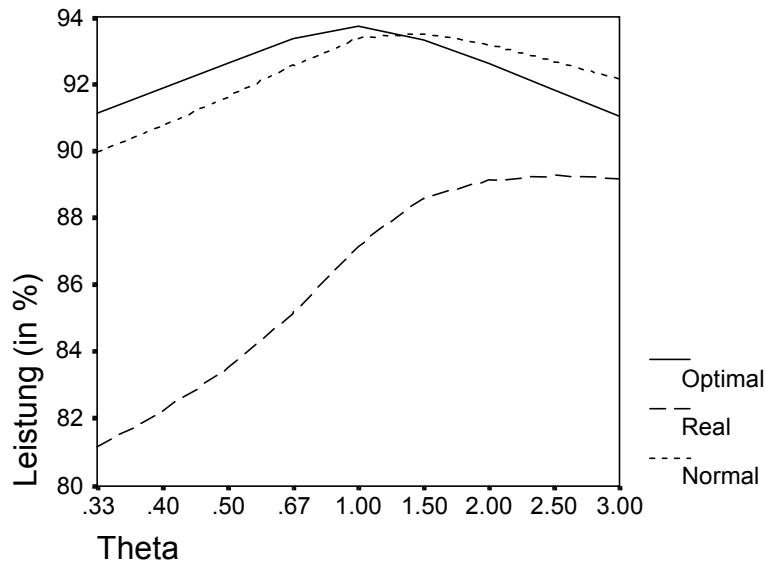


Abb. 51 Simulation VIII: Die Leistung in Abhängigkeit von θ (Wahlausgaben).

Betrachtet man wiederum nur Aufgaben mit mindestens 50% Ambiguität, so ist festzustellen, dass sich die θ -Optima für *Normal* und *Real* nach oben verschoben, auf 2 respektive 3. Für diese Strategien änderten sich die maximalen Leistungseinbußen praktisch nicht, bei *Optimal* sanken sie deutlich auf 3% [1%].

Die relative Wirkkraft der vermittelnden Variablen

Um ein einheitliches Maß dafür zu erhalten, wie stark der Einfluss der untersuchten Variablen war, wurde der Gesamtdatensatz nach den drei Strategien zur Ambiguitätsauflösung aufgesplittet. Für die Teildatensätze von *Optimal*, *Real* und *Normal* wurde dann mittels einfaktorieller Varianzanalysen für die verbleibenden Variablen ermittelt, wieviel Leistungsvarianz diese aufklärten. In Form von Produkt-Moment Korrelationen sind die Ergebnisse in der Abb. 52 wiedergegeben. Für die Strategie *Real*, welche am ehesten reales Entscheidungsverhalten widerspiegeln dürfte, ergab sich dabei, dass die Leistung mit Abstand am stärksten durch den Grad der Ambiguität und den Aufgabentyp beeinflusst wurde (jeweils $r \approx .6$). Mittlere Effekte zeigten die Aufgabengröße und θ (r jeweils $\approx .3$), während der Einfluss der Variabilität von € insgesamt minimal war.

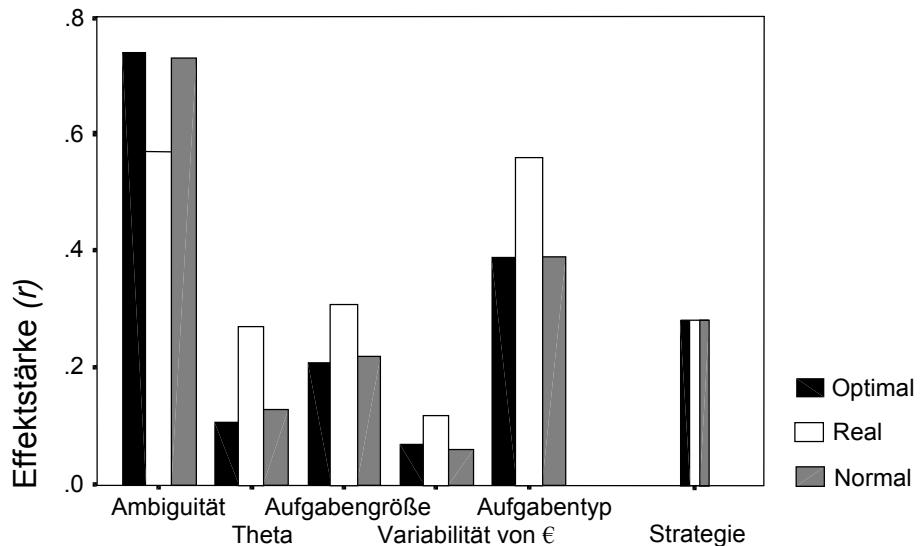


Abb. 52 Simulation VIII: Effektstärken der untersuchten Variablen (für die ersten fünf in Abhängigkeit von der benutzten Strategie), angegeben als Produkt-Moment Korrelation. Der letzte dreifarbiges Balken zeigt den Effekt der Strategie. Nähere Erläuterung im Text.

In gleicher Weise habe ich dann noch einmal den Gesamtdatensatz einer Varianzanalyse mit dem dreigestuften Faktor „Strategie“ unterzogen. Der Effekt der alternativen Verwendung von *Real*, *Optimal* oder *Normal* erwies sich dabei ebenfalls als mittelstark ($r = .28$; äußerster rechter Balken in Abb. 52).

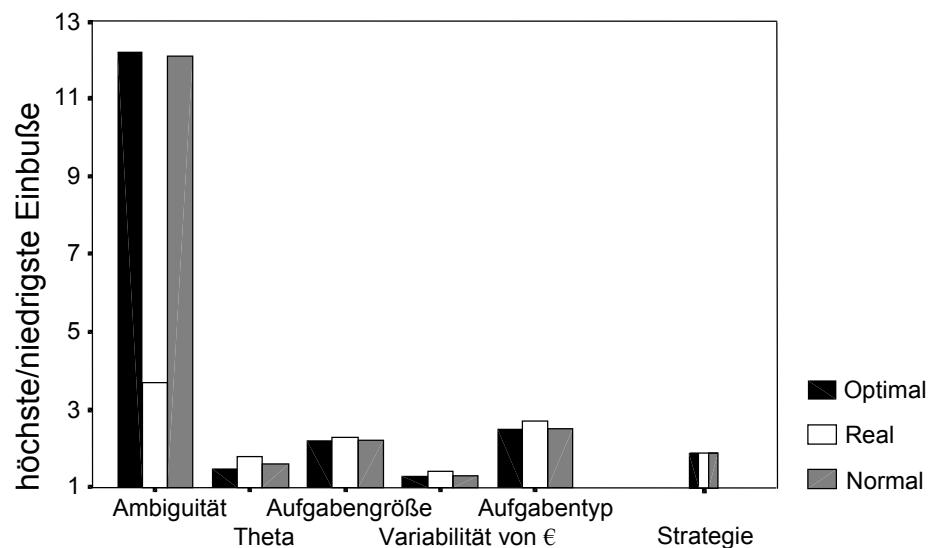


Abb. 53 Simulation VIII: Maximale mittlere Effekte der untersuchten Variablen, für die ersten fünf in Abhängigkeit von der benutzten Strategie. Der letzte dreifarbiges Balken zeigt den Effekt der Strategie. Nähere Erläuterung im Text.

Im Gegensatz zu den gerade diskutierten mittleren Effekten gibt Abb. 53 die „maximalen mittleren Effekte“ der untersuchten Variablen wieder. Maximaler mittlerer Effekt meint dabei Folgendes: Für jede Variable wurde hier der Leistungsabfall, der sich unter der ungünstigsten Ausprägung

der Variable ergab, zu demjenigen der günstigsten in Beziehung gesetzt. So ist etwa die ungünstigste Ausprägung von Ambiguität 90%. Für *Real* ergab sich hier eine mittlere Leistung von 82,3%, was 17,7% unter dem Optimum liegt. Die günstigste Ausprägung von Ambiguität ist 0% – hier betrug der Leistungsabfall lediglich 4,8%. Somit ergibt sich für die Variable Ambiguität bei Verwendung der Strategie *Real* ein maximaler mittlerer Effekt von 3,7 (17,7/4,8); ein Wert von eins würde bedeuten, dass gar kein Effekt vorliegt.

Wie zu sehen ist, ordnen sich die fünf Variablen hinsichtlich ihrer Wirkmächtigkeit unter allen drei Strategien in dieselbe Rangfolge wie zuvor. Die Bedeutung des Ambiguitätsgrades zeigte sich hier aber noch stärker als bei der Betrachtung der mittleren Effekte.

Analog zum oben geschilderten Vorgehen wurde auch der maximale mittlere Effekt der Strategie bestimmt; dieser erwies sich wiederum als mittelstark.

13.3.2 Andere Parameterwerte andere Ergebnisse?

Die hier erzielten Ergebnisse sind natürlich davon abhängig, welche Wertebereiche für die in die Simulation aufgenommenen Variablen untersucht wurden. Hätten prinzipiell ganz andere Ergebnisse resultieren können, wenn extremere Ausprägungen der betrachteten Variablen mit untersucht worden wären?

Für die Variable Ambiguität gibt es eine natürliche Grenze, 100%, die unwesentlich über dem untersuchten Ambiguitätsmaximum von 90% liegt. Mir ist keine Untersuchung bekannt, die systematisch der Frage nachgeht, mit was für Ambiguitätsgraden Menschen im Allgemeinen konfrontiert werden. Eine wichtige Quelle für Ambiguität stellt in jedem Fall die Vagheit sprachlicher Wahrscheinlichkeitsaussagen dar. In Experiment III führte die Verwendung solcher sprachlicher Ausdrücke zu einem Informationsverlust, der im Maximum dem einer Ambiguität von 45% entspricht (vgl. 11.1). Auch wenn ich das nicht belegen kann, erscheint es mir doch unwahrscheinlich, dass Menschen im Alltag öfter Wahrscheinlichkeitsinformationen ausgesetzt sind, deren Ambiguität 90% übersteigt. Somit beurteile ich die von der Simulation abgedeckte Ambiguitätsspannbreite als vollkommen ausreichend.

Für die Variabilität von ϵ gibt es eine natürliche Untergrenze, die darin besteht, dass keine Variabilität vorhanden ist, alle Optionen also den gleichen potenziellen Ertrag haben. Diese Untergrenze wurde in der Simulation berücksichtigt. Eine natürliche Obergrenze gibt es nicht, denn die Variabilität der Lotteriewerte ist beliebig steigerbar. Wie der Abb. 49 zu entnehmen ist, hat der Einfluss auf die Leistung aber eine klare Obergrenze, die in der Simulation auch erreicht wurde. Hätte die Simulation noch extremere Werte für die Variabilität von ϵ berücksichtigt, so wäre durch diese Variable also insgesamt weniger Leistungsvarianz aufgeklärt worden, am mittleren maximalen Effekt dieser Variable hätte sich aber praktisch nichts geändert. Auch die Variabilität von ϵ wurde also erschöpfend untersucht.

Auch für die Wirkung der Aufgabengröße gibt es offensichtlich einen Deckeneffekt. Betrachtet man Abb. 50, so liegt es nahe anzunehmen, dass der maximal erreichbare Effekt nicht wesentlich über demjenigen liegt, den die in der Simulation noch enthaltene Aufgabengröße von 10 Lotterien

bewirkte. Somit gilt hierfür dasselbe wie für die Variabilität von ϵ , dass nämlich die Betrachtung weiterer Ausprägungen dieser Variable dazu geführt hätte, dass durch sie weniger Leistungsvarianz aufgeklärt worden wäre, deutliche Veränderungen des mittleren maximalen Effektes hätte dies aber nicht nach sich gezogen.

Anders verhält es sich mit der letzten untersuchten Variable, θ . Es ist zu erwarten, dass extrem hohe θ Werte zu äußerst starken Effekten führen. Insbesondere bei hoher Ambiguität muss es verheerend wirken, Entscheidungen praktisch ausschließlich von der Wahrscheinlichkeitskomponente abhängig zu machen. Allerdings gibt es keinen Grund zu vermuten, dass Menschen ein solch extremes Verhalten zeigen; die Spannbreite der untersuchten θ Werte orientierte sich an den experimentellen Befunden der vorangegangenen Untersuchungen, deutlich über diese hinauszugehen erscheint mir nicht sinnvoll.

Insgesamt gilt also, dass extremere Ausprägungen der in der Simulation untersuchten Variablen wahlweise psychologisch wenig plausibel gewesen wären oder aber zu keinen stärkeren Effekten geführt hätten. Somit ist auszuschließen, dass die untersuchten Variablen wesentlich stärkere Effekte haben können, als hier gefunden wurden.

13.3.3 Die Beantwortung der Kernfragen und aus diesen resultierende Implikationen

Zu Eingang dieses Kapitels habe ich fünf spezifische Fragen formuliert, die die Simulation VIII beantworten sollte; diesen möchte ich mich jetzt zuwenden.

Die Adaption von θ

Welcher generelle Einfluss auf die Leistung ist zu erwarten, wenn (wie von den Vpn im Gewinnbereich gezeigt) θ mit zunehmender Ambiguität steigt? Poolt man die Daten von Experiment I und II (hier nur Gewinnbereich), so ergaben sich für θ die Mediane 1.4, 1.6, 1.7 und 2.5 (0% – 85% Ambiguität). Betrachtet man von den Simulationsergebnissen allein die Rankingaufgabe, so ergaben sich hier für *Real*, über sämtliche übrige Bedingungen gemittelt, als optimale θ -Werte 1,5 (0% Ambiguität), 2 (30% Ambiguität), 3 (60% Ambiguität) und 3 (90% Ambiguität)²¹. Völlig gegenintuitiv hatte es für die Vpn also einen adaptiven Wert, mit zunehmender Ambiguität die Wahrscheinlichkeitskomponente stärker zu gewichten. Unter einem normativen Gesichtspunkt wäre sogar eine stärkere Erhöhung von θ zu fordern. Das zunächst in normativer wie deskriptiver Hinsicht plausibel erscheinende Verhalten, mit zunehmender Ambiguität θ zu verringern, würde sich insgesamt kontraproduktiv auswirken. Da die Höhe von θ insgesamt aber nur einen sehr geringen Einfluss auf die Leistungen hatte (vgl. die Abbildungen 52 und 53), sind alle diese Effekte in ihrer Wirkung zu vernachlässigen. Für Wahlaufgaben ergab sich in allen hier angeführten Punkten ein quasi identisches Bild.

²¹ Die genauere Betrachtung der Daten weist darauf hin, dass bei 90% Ambiguität das θ -Optimum noch deutlich über 3 liegt. Die Simulation hat jedoch höhere θ -Werte nicht in Betracht gezogen.

Wie bereits erwähnt ergab sich für die Strategie *Optimal* ein etwas anderes Bild. Auf allen Ambiguitätsstufen erwies sich hier ein 0 von 1 als optimal; entgegen dem Augenschein war es also auch bei einer optimalen Auflösung der Ambiguität nicht angebracht, die Wahrscheinlichkeitsinformation bei steigender Ambiguität schwächer zu gewichten.

Die suboptimale Auflösung der Ambiguität

Die Strategie *Normal* unterscheidet sich von der optimalen Strategie dadurch, dass sie die Ambiguität auf eine Art auflöst, die geringe Kapazitätsanforderungen stellt und darüber hinaus beim ersten Augenschein leicht als die optimale Strategie erscheinen kann. Die Vereinfachung, die sie gegenüber *Optimal* vornimmt, kostete sie praktisch nichts: Es ergab sich praktisch kein Leistungsunterschied zu der ausgeklügelten Strategie (vgl. Abb. 47).

Ähnliche Befunde, dass sehr einfache Strategien zu ebenso guten Ergebnissen kommen wie wesentlich aufwendigere, mathematisch ausgefeilte Strategien, fanden sich auch in einem anderen Bereich der Entscheidungsforschung, nämlich der Vorhersage eines Kriteriums anhand einer Reihe von Prädiktoren (z. B. Dawes, 1979, Gigerenzer & Goldstein, 1996, Czerlinsky, Gigerenzer & Goldstein, 1999). Letztgenannte Autoren untersuchten anhand von 20 Variablen aus den verschiedensten Bereichen, wie gut sich diese jeweils aus einer Reihe von Prädiktoren vorhersagen ließen, entweder mittels multipler Regression oder anhand einer einfachen Strategie *Take the Best*, bei der sich das Urteil allein auf den stärksten Prädiktor stützte. Zu den vorherzusagenden Variablen gehörten so verschiedenartige wie die Ozonwerte an bestimmten Tagen in San Francisco, die Fertilität von verschiedenen Fischarten oder die Abbruchquoten an amerikanischen High Schools. Bei einer Kreuzvalidierung erwies sich dabei *Take the Best* der wesentlich rechenaufwendigeren und vom mathematischen Konzept anspruchsvolleren multiplen Regression überlegen.

In vergleichbarer Weise lohnte auch in der Simulation VIII der erheblich höhere Aufwand von *Optimal* nicht. Der wesentliche Unterschied dieser Simulation zu den zuvor erwähnten Untersuchungen besteht darin, dass hier die Prädiktoren (Wert- und Wahrscheinlichkeitsinformation) nicht linear sondern multiplikativ zu verknüpfen sind.

Der Einfluss von error

Einen erheblich größeren Einfluss als die suboptimale Auflösung der Ambiguität zeigte der Fehlerterm *error*. Gegenüber *Normal* verringerte er die Leistungen, weitgehend unabhängig vom Ambiguitätsniveau, um rund 2% im Rankingparadigma und um rund 5,5% bei den Wahlauflagen (vgl. Abb. 47). Wenn man jedoch bedenkt, wie groß *error* war – zur Erinnerung, dieser Term führte dazu, dass durchschnittlich um über 25% von der Attraktivitätsschätzung von *Normal* abgewichen wurde – so zeigt dies noch einmal, dass Entscheidungen unter Unsicherheit gegen Störeinflüsse recht robust sind. Die Größe des Fehlerterms, die ja anhand des aktuellen Verhaltens der Vpn festgelegt worden war, legt in meinen Augen die Vermutung nahe, dass dieser Term nicht allein Rechenfehler bei der multiplikativen Verknüpfung von Wert- und Wahrscheinlichkeitsinformation widerspiegelt. Mög-

licherweise beschreibt er eher eine Tendenz in Richtung Probability Matching (vgl. auch 9.3.6) oder eine allgemeine Strategieinkonsistenz. Wenn dem so ist, dann sollte ein entsprechendes Training für Entscheidungen unter Unsicherheit eher an diesem Punkt ansetzen, indem es Entscheidern die konsequente Beibehaltung einer einfachen Entscheidungsstrategie ans Herz legt, als dass es etwa die Regression zur Mitte von Wahrscheinlichkeitsschätzungen, die auf unsicherer Grundlage gewonnen wurden, thematisiert oder der Frage, inwieweit vagere Information anders gewichtet werden sollte als präzise, nachgeht. Auch Befunde von Dawes (1979) sowie Dawes und Corrigan (1974) weisen in dieselbe Richtung: Diese hatten die Vorhersageleistungen von Experten mit der Leistung äußerst simpler Vorhersagestrategien verglichen, die mit den Informationen gefüttert wurden, auf deren Grundlage auch die Experten ihre Urteile fällten. Eine solche einfache Regel sah etwa so aus, dass alle verfügbaren Informationen als gleich wichtig angesehen wurden und einfach die Zahl der Gründe, die in die eine Richtung wiesen, gegen die Zahl der Informationen, die in die entgegengesetzte Richtung wiesen, verrechnet wurden und das Urteil dann auf dieser Differenz basierte; eine solche Strategie ist also rechnerisch äußerst anspruchslos und kann von jedermann leicht erlernt und vollzogen werden. Die Leistungen dieser höchst einfachen Modelle war durchweg denjenigen der Experten überlegen.

Noch einmal: der Einfluss der Ambiguität

Der Einfluss der Ambiguität war schon näher behandelt worden (vgl. auch Abb. 47). An dieser Stelle möchte ich nur noch einmal darauf hinweisen, dass auch diese Befunde für die Robustheit von Entscheidungen unter Unsicherheit sprechen. Diese zeigten sich nicht nur relativ unempfindlich ge-

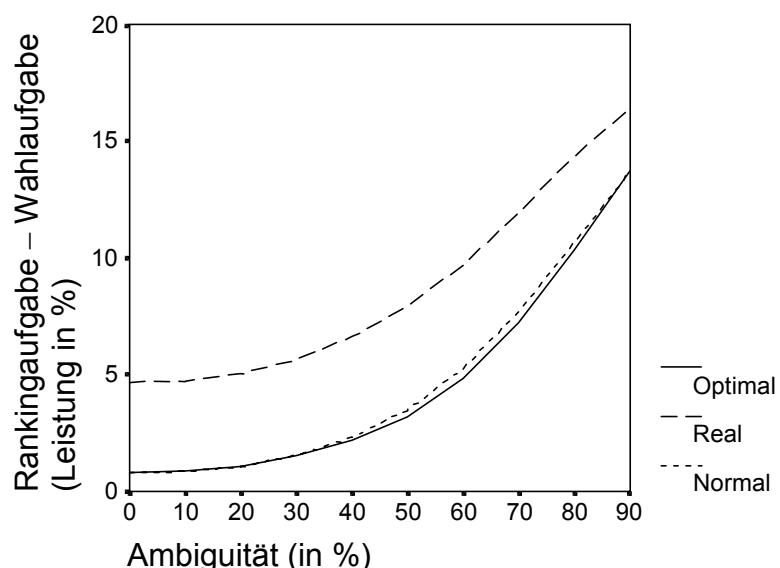


Abb. 54 Simulation VIII: Bei Wahlaufgaben wird die Leistung durch Ambiguität stärker beeinträchtigt als bei Rankingaufgaben. Mit zunehmender Ambiguität verstärkt sich dieser Unterschied.

gen Strategiedefizite (s. o.), sondern auch erheblicher Informationsverlust führte hier zu insgesamt erstaunlich geringen Leistungseinbußen. So erlaubte die konsequente Anwendung der äußerst simplen Strategie *Normal* im Rankingparadigma selbst bei einer Ambiguität von 90% noch eine Leistung von rund 93%.

Die Leistungen in den Wahlaufgaben erwiesen sich jedoch durch Ambiguität als deutlich störanfälliger, wobei die Leistungsunterschiede zwischen den beiden Aufgabentypen mit steigender Ambiguität exponentiell anwuchsen (vgl. Abb. 54).

Generell ist davon auszugehen, dass die Kosten der Verringerung von Ambiguität einen stark abfallenden Grenznutzen haben – d. h. bei hoher Ambiguität wird eine Anstrengung, diese zu vermindern, einen höheren Nutzen haben als dieselbe Anstrengung auf einem niedrigen Ambiguitätsniveau. Zwei Gründe sprechen dafür: Erstens bringt eine Ambiguitätsreduktion einer gewissen Stärke einen höheren Vorteil, wenn diese von einem hohen Ambiguitätsniveau aus erfolgt als wenn eine bereits recht eindeutige Wahrscheinlichkeitsinformation weiter präzisiert wird (vgl. Abb. 47). So brachte etwa, über sämtliche Bedingungen der Simulation VIII hinweg, eine Reduktion der Ambiguität von 90% auf 80% eine Leistungssteigerung von 2,5%; wurde aber die Ambiguität von 10% auf 0% verringert, so führte dies lediglich zu einem Leistungszugewinn von 0,1%. Zweitens dürfte zusätzlich eine Ambiguitätsminderung von 90% auf 80% mit weit weniger Aufwand zu erreichen sein als eine von 10% auf 0%. Anstrengungen zur Ambiguitätsminderung sollten sich also insbesondere dann bezahlt machen, wenn die Ambiguität sehr hoch ist.

Was kostet die verbale Beschreibung von Wahrscheinlichkeiten?

Um diese Frage zu beantworten, ist es zunächst notwendig herauszufinden, ob und wenn ja auf welche Weise sich die Unsicherheit, die in verbalen Wahrscheinlichkeitsaussagen steckt, auf die Simulation abbilden lässt. Denn schließlich gibt es eine Reihe von Unterschieden zwischen verbalen Beschreibungen und der Art, wie Ambiguität hier operationalisiert wurde: (i) In der Simulation war die Information, die dem Entscheider über P zur Verfügung stand, dergestalt, dass das ambiguitätsformende Intervall klar umrissene Grenzen (P_{lower} und P_{upper}) hatte. Für verbale Umschreibungen wie „eher wahrscheinlich“ gilt dies nicht; hier lassen sich keine präzisen Bedeutungsgrenzen angeben. (ii) In der Simulation war sichergestellt, dass P in dem angegebenen Intervall enthalten war. Es ist hingegen nicht zwangsläufig so, dass die wahre Gewinnwahrscheinlichkeit in dem Bedeutungsbereich liegt, das der Entscheider einer verbalen Wahrscheinlichkeitsbeschreibung beimisst. (iii) In der Simulation war klar definiert, wie die wahre Wahrscheinlichkeit P in die Wahrscheinlichkeitsinformation $[P_{\text{lower}}, P_{\text{upper}}]$ umgewandelt wurde, auf Grundlage derer der Entscheider handelte. Darüber, wie Menschen Wahrscheinlichkeiten in verbale Wahrscheinlichkeitsinformation umwandeln, wissen wir nichts. (iv) Nur ein Teil des bei sprachlicher Information zu erwartenden Leistungsverlustes dürfte auf die Ambiguität zurückgehen; ein wesentlicher Teil wird daher röhren, dass solche verbalen Äußerungen nicht von allen Menschen gleich verstanden werden (z. B. Mosteller & Youtz, 1990) und somit systematisch Missverständnisse zwischen Sender und Empfänger zu erwarten sind. Abb. 55 veranschaulicht diesen Unterschied noch einmal.

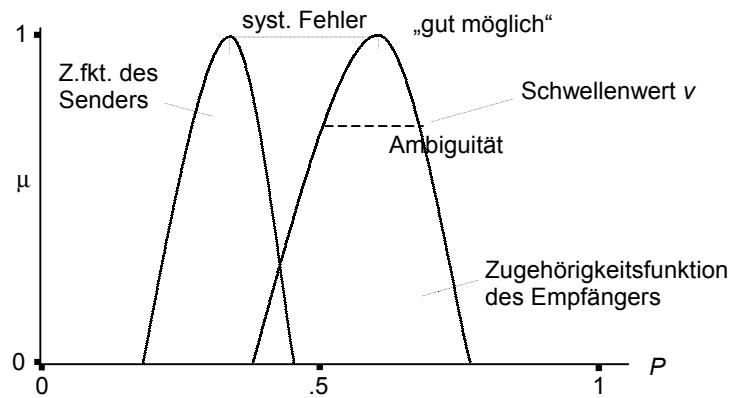


Abb. 55 Die verbale Beschreibung einer Wahrscheinlichkeit, hier beispielhaft „gut möglich“, führt beim Empfänger zu Ambiguität, die sich leistungsmindernd auswirkt. Zusätzlich kann aber auch ein systematischer Fehler wirksam werden, der daher röhrt, dass Sender und Empfänger ein unterschiedliches Verständnis von „gut möglich“ haben. Zur hier angenommenen Repräsentation der Bedeutung von „gut möglich“ vgl. Kap. 6.

Der letztgenannte Punkt sollte jedoch nicht großartig ins Gewicht fallen, denn die leistungsmindernde Wirkung der Ambiguität röhrt natürlich daher, dass aufgrund ihrer Wirkung die Wahrscheinlichkeitsschätzung des Entscheiders mit einem Zufallsfehler behaftet ist. Ob die Quelle eines solchen Fehlers nun in Ambiguität oder in einem systematischen Verständnisunterschied eines verbalen Ausdrucks liegt, ist dabei egal; beide Faktoren lassen sich in derselben „Währung“ beschreiben.

Inwieweit die Ergebnisse der Simulation auf die verbale Beschreibung von Wahrscheinlichkeiten übertragbar sind, ob also die oben aufgeführten Unterschiede (i) bis (iii) ins Gewicht fallen, lässt sich ansatzweise anhand der Ergebnisse von Experiment III überprüfen. Denn hier wurden dieselben Lotterien mit bekannten Wahrscheinlichkeiten einmal ohne Ambiguität präsentiert und einmal so, dass die Wahrscheinlichkeiten in Form verbaler Beschreibungen angegeben waren. Somit lässt sich überprüfen, ob die Vpn hier solche Leistungen erzielten, wie sie die Simulation VIII erwarten lässt. Für die Bedingung ohne Ambiguität lässt sich das Äquivalent der Simulation leicht finden: Ambiguität = 0%, Aufgabentyp = Rankingaufgabe, Variabilität von $\epsilon = 3,3$, Aufgabengröße = 6, $\theta = 2$ und Strategie = *Real* (vgl. Abb. 43). Hierfür ergab sich in der Simulation eine Leistung von 98,4%. Welches Ambiguitätsniveau in der Simulation sollte als Vergleichsmaßstab für die experimentelle Bedingung „verbale Wahrscheinlichkeitsinformation“ herangezogen werden? Die kritische Größe für einen zu erwartenden Leistungsabfall in dieser Bedingung ist, inwieweit die wahren Wahrscheinlichkeiten durch den Schleier der verbalen Beschreibungen akkurat wahrgenommen werden konnten. Ein Maß für diese Akkuratheit ist die Korrelation zwischen den von den Vpn geschätzten Wahrscheinlichkeiten, $P_{\text{Schätz}}$ (vgl. 11.1), und den wahren Wahrscheinlichkeiten, P . Daher wurde diese Korrelation für jede der 53 Vpn von Experiment III bestimmt. Mittels χ^2 -Transformation wurden diese dann gemittelt; der mittlere Zusammenhang betrug dabei $r = .945$. Im Paradigma der Simula-

tion wird ein Zusammenhang in dieser Höhe bei 31% Ambiguität erreicht²². Daher wurde als Äquivalent die entsprechende Bedingung mit 30% Ambiguität herangezogen. Die Simulation wies hier eine Leistung von 97,8% auf, also .6% weniger als in der Bedingung ohne Ambiguität.

Die empirisch gefundenen Leistungsdifferenzen zwischen den Bedingungen „0% Ambiguität“ und „verbale Wahrscheinlichkeitsinformation“ hatten ein Mittel von .9% und unterschieden sich nicht signifikant von dem prognostizierten Wert von .6% ($T = 1,1; df = 52; p = .27$). Zu beachten ist hierbei, dass *Real*, welches tatsächliches Entscheidungsverhalten nachzeichnen soll, ausschließlich auf der Grundlage der Ergebnisse der Experimente I und II entworfen wurde, so dass die Ergebnisse von Experiment III einen echten unabhängigen Test darstellen. Dass dieser positiv ausgefallen ist, legt nahe, dass die oben skizzierten möglichen Unterschiede (i) bis (iv) zwischen Situationen, in denen die relevanten Wahrscheinlichkeiten mittels Worten ausgedrückt werden, und dem Paradigma der Simulation nicht ins Gewicht fallen.

Der in Experiment III gefundene Leistungsunterschied zwischen den Bedingungen „verbale Wahrscheinlichkeitsinformation“ und „keine Ambiguität“ erwies sich wie von der Simulation vorhergesagt als äußerst gering. Können wir davon ausgehen, dass dies unter anderen Randbedingungen auch so aussähe?

Ehe dies näher beleuchtet wird, müssen wir uns vor Augen führen, dass in der experimentellen Aufgabe der Schaden, den der Ausdruck von Wahrscheinlichkeiten in Worten angerichtet hat, im Vergleich zu vielen Alltagssituationen womöglich eher gering ist. Denn im Experiment waren die wahren Wahrscheinlichkeiten den Sendern bekannt; somit war es möglich, diese alternativ in numerischer Form (nahezu) ohne Fehler zu kommunizieren, weswegen die Versuchsbedingung „0% Ambiguität“ hier einen adäquaten Vergleichsmaßstab abgibt. Es gibt Alltagssituationen, für die in gleicher Weise gilt, dass die wahren Wahrscheinlichkeiten (zumindest in guter Näherung) bekannt sind, aber nicht numerisch sondern verbal ausgedrückt werden. Dies ist etwa bei deutschen Beipackzetteln der Fall. Hier schlägt das Bundesgesundheitsamt vor, dass die Auftretensrisiken der einzelnen Medikamentennebenwirkungen durch drei verschiedene Formulierungen beschrieben werden, wobei jeder Beschreibung ein festes Risikointervall zugeordnet ist (Bundesgesundheitsamt, 1991). Auf der Grundlage der durchgeführten klinischen Studien dürften die Pharmahersteller hinlänglich genaue Kenntnisse über die Risiken der einzelnen Nebenwirkungen haben, deren Darstellung erfolgt jedoch nicht numerisch sondern verbal.

Anders als in diesem Beispiel dürfte aber den Sendern in den meisten Alltagssituationen eine wahre Wahrscheinlichkeit nicht bekannt sein; somit können sie, wenn sie auf einen verbalen zugunsten eines numerischen Wahrscheinlichkeitsausdrucks verzichten, lediglich den Fehler vermeiden, der daraus entspringt, dass der Empfänger möglicherweise ein systematisch anderes Verständnis von Wahrscheinlichkeitsausdrücken hat als die selbst (vgl. Abb. 55). Der Fehler, der daraus resultiert, dass die wahre Wahrscheinlichkeit nicht bekannt ist, lässt sich aber nicht beseitigen. In diesen Fällen wäre also die Simulationsbedingung „0% Ambiguität“ ein zu harter Vergleichsmaßstab. Schätzen wir die Leistungsminderung, die aufgrund verbaler Kommunikation zu erwarten ist, indem wir in der

²² Ergebnis einer Monte-Carlo Simulation.

Simulation die Leistung unter dem entsprechenden Ambiguitätspendant mit derjenigen bei fehlender Ambiguität vergleichen, so gelangen wir zu konservativen Schätzungen – die tatsächlichen Leistungseinbußen sollten oftmals niedriger ausfallen.

Welchen generellen Schaden lässt die Simulation nun nach diesen Vorüberlegungen vermuten? Da die drei Personen, die die verbalen Wahrscheinlichkeitsbeschreibungen für Experiment III beisteuerten, instruiert waren, diese so zu formulieren, dass andere Personen die wahren Wahrscheinlichkeiten aus diesen Beschreibungen möglichst gut rekonstruieren könnten, hatten die Beschreibungen möglicherweise eine höhere Qualität, als dies in der Alltagskommunikation oft der Fall ist. 31% Ambiguität erwies sich als Äquivalent für den Fehler, den diese Beschreibungen verursacht hatten (s. o.); aus gerade genannten Gründen betrachte ich deswegen an dieser Stelle 40% Ambiguität als Pendant für den Fehler, den Alltagskommunikation nach sich zieht. Gegenüber den Bedingungen ohne Ambiguität ergaben sich hier mittlere Leistungseinbußen für die Rankingaufgabe von 1,6% (*Real*) und 1,2% (*Normal*) und für die Wahlaufgabe von 3,5% (*Real*) und 2,7% (*Normal*); die maximalen Leistungseinbußen lagen jeweils rund doppelt so hoch.

Wenn man in Betracht zieht, dass diese Schätzungen eher pessimistisch sind (s. o.), so lässt sich der Schluss ziehen, dass der verbale Ausdruck von Wahrscheinlichkeiten an Stelle eines numerischen, die Leistungen entsprechend informierter Entscheider über eine Vielzahl von Randbedingungen hinweg nur geringfügig verschlechtert. Freilich kann ein geringer relativer Unterschied dennoch einem beachtlichen absoluten Unterschied entsprechen. Was folgt daraus? Der Vorschlag von Mosteller & Youtz (1990), Sprache, soweit sie sich auf den Ausdruck von Wahrscheinlichkeit bezieht, zu normieren, erscheint mir vollkommen illusorisch. Angesichts der Tatsache, dass auch ein oder zwei Prozent Leistungsminderung bei Entscheidungen unter Unsicherheit durchaus erheblich sein können, erscheint mir die Mahnung Beyth-Marom (1982), Experten sollten sich bei der Kommunikation subjektiver Sicherheit Zahlen bedienen, angemessen. Anders als sie halte ich es aber nicht für sinnvoll, dass Experten ihr Urteil in Form eines Intervalls angeben, um so gleichzeitig den Grad der Vagheit ihres eigenen Urteils auszudrücken. Denn welche Funktion könnte dies haben? Doch wohl nur die, dass die Entscheider die Information in Abhängigkeit von deren Vagheit gewichten. Wie die vorliegenden Ausführungen gezeigt haben, ist aber gerade das völlig irrelevant, und das Augenmerk der Entscheider würde somit auf einen Punkt gelenkt, dessen Beachtung kaum lohnt. Wie bereits ausgeführt erscheint es mir viel sinnvoller, wenn sich Entscheider statt dessen darauf konzentrieren, eine einfache Strategie konsequent zu befolgen.

13.4 Die Übertragbarkeit auf natürliche Entscheidungssituationen

Die Simulation VIII ist von ihrer Anlage dem Paradigma der Experimente I und II verpflichtet. Ein erster Test dahingehend, ob die Ergebnisse auch auf den Einfluss verbaler Wahrscheinlichkeitskommunikation generalisiert werden können, fiel ermutigend aus (vgl. 13.3.3). Dies ist um so wichtiger, als dass deren leistungsmindernde Wirkung nicht allein auf Ambiguität zurückgeht sondern auch auf die Wirkung eines von dieser unabhängigen Fehlers (vgl. dazu noch einmal Abb. 55). Somit erscheint es mir plausibel, dass die Simulation nicht allein sinnvolle Aussagen über die Wirkung von Ambiguität

machen kann sondern weiter gefasst solche über den Einfluss einer fehlerhaften Wahrnehmung der Gewinnwahrscheinlichkeiten; dazu ist es natürlich notwendig, einen solchen Fehler in die „Währung“ der Simulation, Ambiguität, umzurechnen, wie es in 13.3.3 beispielhaft geschah.

Eine fehlerhafte Wahrnehmung von Wahrscheinlichkeiten kann natürlich viele Quellen haben, nicht nur Ambiguität. So kann etwa ein Experte einem Entscheider eine präzise, nichtambige Wahrscheinlichkeit mitteilen, die schlicht falsch und von der wahren weit entfernt ist.

13.4.1 Fehlende „wahre Wahrscheinlichkeiten“

Alle bisherigen Überlegungen beruhten darauf, dass es eine wahre Wahrscheinlichkeit gibt und dass diese bekannt ist. Bei der Kalkulation der Kosten einer verbalen Kommunikation von Wahrscheinlichkeiten bin ich weiter davon ausgegangen, dass diese wahre Wahrscheinlichkeit dem Sender bekannt ist, wie dies in Experiment III tatsächlich der Fall war. In vielen, wahrscheinlich den allermeisten, natürlichen Situationen ist dies nicht der Fall. Stellen Sie sich vor, Sie hätten Streit mit Ihrem Vermieter, und es gelingt Ihnen nicht, mit diesem zu einer Einigung zu kommen. Sie suchen juristischen Beistand, und Ihre Anwältin sagt Ihnen, dass sie Ihre Chancen, falls es zu einem Prozess kommt, auf 40% schätzt. Wie sieht hier die „wahre“ Wahrscheinlichkeit aus? Kann es eine solche überhaupt geben? Und lässt sich die Einschätzung der Anwältin überhaupt überprüfen? Stellen wir uns vor, es kommt tatsächlich zum Prozess – egal ob Sie gewinnen oder verlieren, beide Ausgänge sind perfekt mit der Einschätzung Ihrer Beraterin zu vertreten. Die Güte ihrer Prognose lässt sich also nicht ohne weiteres überprüfen.

Zumindest dann, wenn man ein pragmatisches Verhältnis zu der Frage hat, was Wahrscheinlichkeiten sind, ist ein Ausweg aus dieser Situation möglich, und eine „wahre Wahrscheinlichkeit“ lässt sich konstruieren. Wie bereits erwähnt, lässt sich eine einzelne Wahrscheinlichkeitsschätzung der Anwältin nicht anhand der Beobachtung des tatsächlichen Prozessausgangs überprüfen (es sei denn, sie lautet 0% oder 100%). Eine Vielzahl solcher Angaben lässt sich aber als Gesamt entsprechend analysieren. Wie, soll im folgenden Abschnitt erklärt werden.

Wenn die Expertin zur oben genannten Aussage kommt, so bedeutet dies (i) sie rechnet damit, dass Sie den Prozess nicht gewinnen werden und (ii) beträgt die subjektive Sicherheit oder Konfidenz für diese Aussage 60%. Stellen wir uns vor, die Anwältin habe in der Vergangenheit schon einhundert Vorhersagen ähnlicher Art gemacht, und wir wüssten, welche sich davon als richtig erwiesen haben und welche als falsch. Wir können dann die Vorhersagen anhand ihrer Konfidenzen zu Gruppen zusammenfassen, also etwa alle Prognosen mit einer Konfidenz zwischen .5 und < .6, des Weiteren alle Prognosen mit einer Konfidenz von .6 bis < .7 und so weiter. In einem nächsten Schritt stellen wir dann dem Konfidenzmittelwert jeder Gruppe die relative Häufigkeit richtiger Vorhersagen gegenüber. Im Optimalfall würden diese Werte in guter Näherung übereinstimmen: Nehmen wir an, die Juristin hätte 20 Vorhersagen mit einer Konfidenz zwischen .5 und < .6 gemacht, und für diese ergäbe sich eine mittlere subjektive Sicherheit von .55, dann sollte dem entsprechend die Zahl der richtigen Vorhersagen in dieser Gruppe bei rund 11 liegen (55% von 20) – und nicht bei 3 oder 17. Und dann erscheint es auch vertretbar und sinnvoll davon auszugehen, dass die

„wahre Wahrscheinlichkeit“ dafür, dass Sie den Prozess verlieren werden, bei .6 liegt. Falls sich erwiese, dass die Anwältin nicht gut kalibriert ist, und in der diskutierten Konfidenzklasse etwa zu 80% richtigen Vorhersagen gekommen ist, so sollten wir davon ausgehen, dass auch in unserem Fall die „wahre Wahrscheinlichkeit“ etwa .8 beträgt, Ihre Chance, den Prozess zu gewinnen, also nur bei rund 20% liegt.

Die Abb. 56 erläutert noch einmal zusammenfassend, wie zu erwartende Leistungseinbußen auch dann bestimmt werden können, wenn keine „wahren“ Wahrscheinlichkeiten bekannt sind.

Wie können wir aber vorgehen, wenn, was äußerst wahrscheinlich ist, keine entsprechenden Erfahrungsdaten über die Prognosegüte der Anwältin vorliegen? Der einzige Ausweg besteht dann darin, auf allgemeinere Erkenntnisse, die die Konfidenzforschung zu Tage gefördert hat, auszuweichen. Diese ausführlich zu würdigen, würde den Rahmen dieser Arbeit sprengen. In aller Kürze lässt sich aber soviel sagen: Kaum bezweifelt wird, dass Konfidenzurteile einem Zufallsfehler unterliegen, der dazu führt, dass die angegebenen Sicherheitsurteile verglichen mit den wahren Werten einer Regression zur Mitte unterliegen (Erev, Wallsten & Budescu, 1994). Umstritten ist, ob es darüber hinaus eine generelle Tendenz dazu gibt, die Qualität des eigenen Wissens zu überschätzen, also insgesamt zu hohe Konfidenzurteile abzugeben (für eine eher pessimistische Sichtweise siehe z. B. Griffin & Tversky, 1992, und als Verfechter einer optimistischen Sicht Gigerenzer, Hoffrage & Kleinböltig, 1991, und Björkemann, 1994).

13.4.2 Faktoren, die nicht berücksichtigt wurden

Auch wenn Simulation VIII eine Vielzahl verschiedener Randbedingungen berücksichtigt hat, gibt es einige Faktoren, die möglicherweise relevant wären, hier aber nicht mit berücksichtigt worden sind. Der erste dieser Faktoren besteht in der Verteilung der Werte für P und ϵ . Beide Variablen wiesen in der Simulation eine uniforme Verteilung auf; diese scheint mir auch solange vertretbar, wie nichts über entsprechende natürliche Verteilungen bekannt ist. Nichts desto trotz ist es möglich, dass andere Verteilungen andere Ergebnisse nach sich gezogen hätten.

Weiter ging die Simulation davon aus, dass die Werte der beteiligten Dimensionen stets unverfälscht wahrgenommen werden. Diese Annahme stellt natürlich eine Vereinfachung natürlicher Verhältnisse dar, auch bei der Einschätzung von möglichen Erträgen oder Verlusten können Fehler auftreten. Dass ein zusätzlicher Fehler in der Wahrnehmung von ϵ die Effekte der hier untersuchten Variablen verändert, kann nicht mit Sicherheit ausgeschlossen werden.

Schließlich sei noch ein letzter Aspekt genannt, der möglicherweise wichtig wäre, hier aber nicht berücksichtigt wurde: In der Simulation VIII (wie auch in den drei Experimenten) waren alle Lotterien einer Aufgabe mit demselben Grad an Ambiguität behaftet. Wenn wir in einer Alltagssituation die Wahl zwischen verschiedenen Handlungsoptionen mit unsicheren Ausgängen haben, können diese natürlich unterschiedlich gut kalkulierbar sein. Die Kombination verschiedener Ambiguitätsstufen innerhalb einer Aufgabe kann insofern ins Gewicht fallen, als dass Menschen (im Gewinnbereich) weniger ambige Optionen sympathischer sind als solche, deren Chancen schwer zu bestimmen sind.

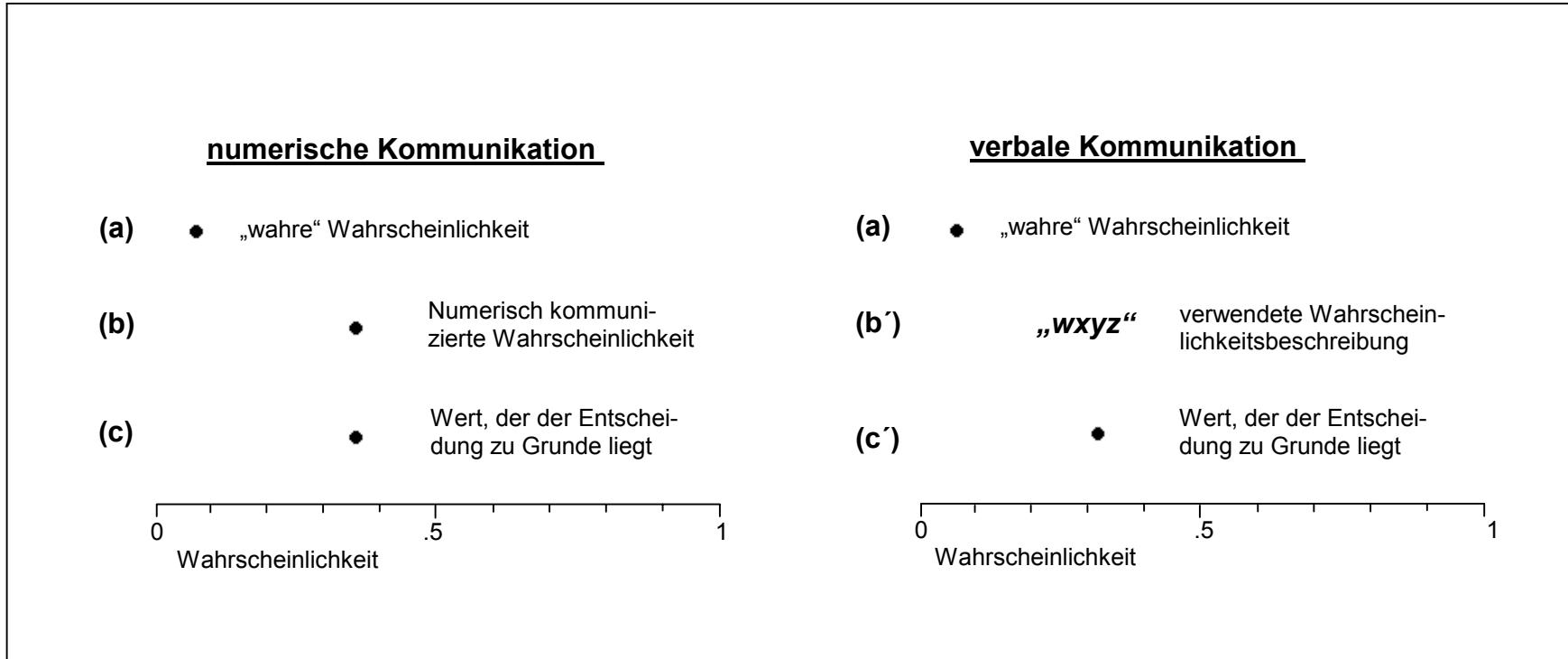


Abb. 56 Wie lassen sich die zu erwartenden Leistungseinbußen verbaler Wahrscheinlichkeitskommunikation bestimmen, wenn „wahre“ Wahrscheinlichkeiten nicht existieren? Vorausbedingung hierzu ist, die Verhältnisse von (a) zu (c) sowie von (a) zu (c') zu bestimmen. Aus (b) lässt sich über Kalibrationsstudien (a) sinnvoll konstruieren (s. Text). Das Verhältnis von (b) zu (c') lässt sich bestimmen, wenn man den Sender bittet, bekannte Wahrscheinlichkeiten in Worten zu beschreiben, und dann schaut, wie diese Beschreibungen von Empfängern in Zahlen rückübersetzt werden (dieses Vorgehen wurde in Experiment III erfolgreich angewendet).

14. FAZIT UND AUSBLICK

Ziel der vorliegenden Arbeit war, die plausibel erscheinende und von Wallsten, Budescu & Tsao (1997) postulierte Gewichtungshypothese anhand eines geeigneten, erst zu entwickelnden Instrumentariums zu überprüfen. Dies ist erfolgreich geschehen.

Die vorliegenden Befunde sprechen dafür, dass Menschen bei zunehmender Ambiguität ihre Entscheidungen *nicht* stärker an den Wertinfomationen ausrichten. Es zeichnete sich eher eine Tendenz ab, bei zunehmender Ambiguität vorsichtiger zu agieren und somit bei Gewinnaufgaben die Entscheidungen stärker von der Wahrscheinlichkeits-, und lediglich bei Verlustraufgaben verstärkt von der Wertinformation abhängig zu machen. Der gefundene Effekt ist schwach und er dürfte, wie Simulation VIII nahelegt, von geringer praktischer Relevanz sein. Bei steigender Ambiguität Entscheidungen stärker von der Wertinformation abhängig zu machen scheint per Augenschein vernünftig, ist es aber nicht. θ bei steigender Ambiguität zu senken hat im Gegenteil einen schwachen leistungsmindernden Effekt, insbesondere dann, wenn das Entscheidungsverhalten, wie bei den Vpn der Fall und bei Entscheidern sicher auch sonst zu erwarten, mit einer Fehlerkomponente behaftet ist.

Sowohl die Betrachtung der Leistungen der Vpn in den Experimenten als auch die Ergebnisse von Simulation VIII legen nahe, dass Entscheidungen unter Unsicherheit erstaunlich robust sind, sowohl gegenüber Rauschen in der Wahrscheinlichkeitsinformation als auch gegenüber suboptimalen Entscheidungsstrategien; dies scheint für eine große Vielzahl berücksichtigter Randbedingungen zu gelten. Daraus folgt unter anderem, dass eine Änderung der menschlichen Gewohnheit, Grade subjektiver Sicherheit in (interindividuell unterschiedlich interpretierten) Worten auszudrücken, generell nur äußerst geringe relative Leistungszuwächse nach sich ziehen dürfte. Den Grad der Präzision subjektiver Sicherheit in Form eines Vertrauensintervalls mit zu kommunizieren, wie es etwa Beyth-Marom (1982) vorschlägt, erscheint im Licht der vorliegenden Ergebnisse vergebliche Liebesmüh: Eine dadurch verursachte Veränderung der Entscheidungsstrategie des Empfängers dürfte – bestenfalls – keine Konsequenzen haben.

Die Ergebnisse der Simulation VIII legen nahe, dass Investitionen, die darauf abzielen Ambiguität zu reduzieren, generell einen stark abfallenden Grenznutzen haben.

Für eine Fülle von Randbedingungen überprüft zeigte sich, dass eine äußerst einfache Entscheidungsstrategie zu hervorragenden Ergebnissen kommt und in der Leistung einer ausgebufften (und rechnerisch sehr anspruchsvollen) Strategie praktisch ebenbürtig ist; vergleichbare Befunde liegen für andere Bereiche von Entscheidungen unter Unsicherheit vor (z. B. Czerlinsky, Gigerenzer & Goldstein, 1999). Für Trainingsmaßnahmen ergibt sich damit in meinen Augen ein vorrangiges Ziel: Entscheidern nahezubringen, dass es wenig Sinn macht, besonders komplexe Entscheidungsstrategien zu ersinnen, die einer Vielzahl von Randbedingungen Rechnung tragen, sondern dass vielmehr der *konsequente* Gebrauch durchdachter, einfacher Strategien vielversprechend erscheint.

Welche Perspektiven eröffnet die vorliegende Arbeit? Genauer zu verfolgen, wie Ambiguität die Gewichtung von Informationen beeinflusst erscheint mir wenig vielversprechend, da die vorliegen-

den Ergebnisse darauf hinweisen, dass entsprechende Effekte so klein sind, dass sie kaum praktische Relevanz haben. Vielversprechender scheint mir, an Simulation VIII anzuknüpfen. Ein erster Test dafür, dass die von ihr vorgenommenen Generalisierungen aussagekräftig sind, ist erfolgreich verlaufen (vgl. 13.3.3), weitere sollten also folgen. Nur ein Beispiel: Die Simulation sagt vorher, dass sich Ambiguität im Wahlparadigma wesentlich deutlicher auf die Leistungen niederschlägt als im hier verwendeten Rankingparadigma; eine solche Vorhersage ist einfach zu überprüfen. Weiter erscheint es mir lohnend, die Simulation auszubauen und um die in 13.4.2 angesprochenen Aspekte zu erweitern; so lässt sich einerseits prüfen, ob die bisher gefundenen Effekte in gleicher Weise auch unter weiteren Randbedingungen gültig sind; andererseits ergeben sich wahrscheinlich weitere neue Vorhersagen, die der experimentellen Forschung Impulse geben können.

ANHANG

A 1 Warum die in dieser Arbeit gewählte Modellierung der Attraktivität von Wahlmöglichkeiten nicht geeignet ist, Alternativen zu vergleichen, die sich in der Zahl möglicher Ausgänge unterscheiden.

Die hier gewählte Modellierung der Attraktivität von Wahloptionen folgt der Grundidee, dass Spiel i , in dem entweder ϵ_i mit der Wahrscheinlichkeit p_i gewonnen wird oder andernfalls nichts, gegenüber Spiel j , in welchem ϵ_j mit der Wahrscheinlichkeit p_j gewonnen wird oder andernfalls nichts, dann und nur dann vorgezogen wird wenn gilt, dass $\epsilon_i p_i^0 \geq \epsilon_j p_j^0$.

Diese Modellierung lässt sich nicht ohne weiteres auf die Wahl zwischen Optionen mit einer unterschiedlichen Zahl möglicher Ausgänge ausdehnen. Die folgende Überlegung wird dies deutlich machen: Nehmen wir an, wir hätten über eine Reihe von Wahlen vergleichbar denen in 7.3 ermittelt, dass sich das Verhalten einer Entscheiderin am besten modellieren lässt, wenn wir bei ihr einen 0-Wert von 2 annehmen. Diese stehe nun vor der Wahl zwischen den beiden folgenden Lotterien

- (A) Sie bekommen 10 DM mit einer Chance von .5, andernfalls nichts.
- (B) Sie bekommen 15 DM mit einer Chance von .25 oder 10 DM mit einer Chance von .25 oder andernfalls nichts.

Die oben angeführte Modellierung würde uns nun zu der unsinnigen Vorhersage zwingen, dass die Entscheiderin Lotterie A wählt, obwohl die Alternative B unbestreitbar die bessere ist (zumindest, solange man mehr Geld attraktiver findet als weniger). Denn für A ergäbe sich eine wahrgenommene Attraktivität von $10 \text{ DM} \cdot .5^2 = 2,5 \text{ DM}$. Für B ergäbe sich diese hingegen als $15 \text{ DM} \cdot .25^2 + 10 \text{ DM} \cdot .25^2 \approx 1,56 \text{ DM}$. Offensichtlich wäre eine solche Modellierung hier also fehl am Platze. Einen Ausweg aus diesem Dilemma bieten sogenannte RDU-Modelle (Rank Dependent Utility Models). Da die hier vorgestellt Modellierung für die Zwecke dieser Arbeit ausreicht, möchte ich es bei deren bloßer Erwähnung belassen und den interessierten Leser beispielhaft auf Lopes (1995) verweisen, wo die Grundidee der RDU-Modelle anschaulich beschrieben wird.

A 2 Instruktion von Experiment III

Die folgenden drei Seiten geben die Versuchsanleitung von Experiment III wieder. Die Instruktionen der Experimente I und II bestanden aus Ausschnitten aus dieser. Im Original waren die Abbildungen farbig. Die Instruktionen endeten mit einer Abbildung, die exemplarisch einen Bildschirm mit Aufgaben ohne Ambiguität zeigte (ähnlich Abb. 22). Aus Platzgründen ist diese hier nicht wiedergegeben.

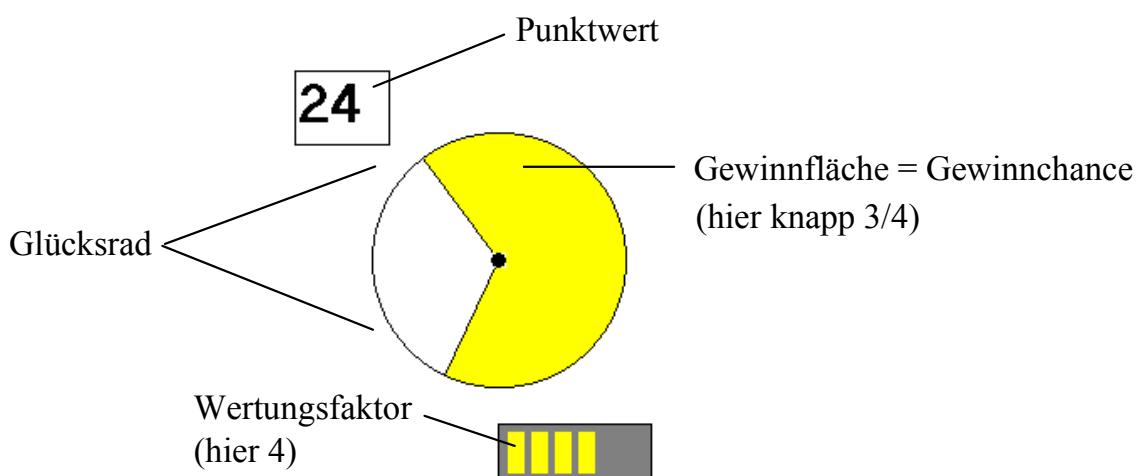
Liebe Teilnehmerin, lieber Teilnehmer,

herzlichen Dank, dass Sie sich bereit erklärt haben, an unserem Experiment teilzunehmen!

Wir interessieren uns dafür, auf welche Weise Menschen sich zwischen verschiedenen Handlungsmöglichkeiten entscheiden, wenn die Konsequenzen der verschiedenen Alternativen nicht genau zu bestimmen sind, weil der Zufall seine Hand mit im Spiel hat. Dieser Versuch soll uns zu einem tieferen Verständnis solcher Wahlen führen.

Zu diesem Zweck sollen Sie eine Reihe von Glücksspielen spielen, bei denen jedoch in hohem Maße auch Ihr **Können** gefordert ist. Bei diesen Spielen sammeln Sie Punkte – je mehr, desto besser, denn die drei Erstplatzierten dieses Versuches erhalten 80/60/40 DM. Ihnen werden dazu jeweils sechs Lotterien gleichzeitig präsentiert (wie dies aussieht, können Sie auf der letzten Seite sehen). Jede Lotterie besteht dabei aus zwei Komponenten: **Punktwert** und **Glücksrad**.

Der Punktwert der Lotterie wird Ihnen gutgeschrieben, wenn Sie in dieser Lotterie gewinnen. Er steht in dem eckigen Kästchen oberhalb des Glücksrades (s. nachfolgende Abb.). Stellen Sie sich oberhalb des Glücksrades eine Nadel vor. Wird das Rad gedreht und bleibt so stehen, dass die Nadel auf die gelbe Fläche zeigt, so haben Sie gewonnen, und die Punkte werden Ihrem Konto gutgeschrieben. Andernfalls gehen Sie bei dieser Lotterie leer aus. Die gelbe Fläche bezeichnet also Ihre **Gewinnchance** in dieser Lotterie – je größer die Fläche ist, desto größer ist Ihre Chance. Die Lotterien wurden vom Computer zufällig zusammengestellt. Als Punktwert wurde eine Zahl zwischen 3 und 97 ausgelost. Die Gewinnchance wurde ebenfalls zufällig zwischen 3% und 97% festgesetzt.



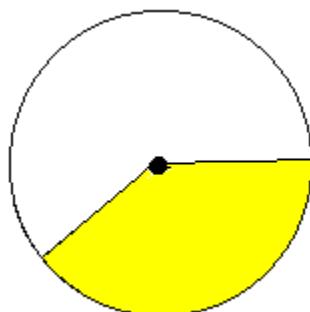
Ihre Aufgabe besteht nun darin, die sechs Lotterien zu **gewichten**. Dazu können Sie einer Lotterie den Faktor 6 geben, einer anderen den Faktor 5, einer weiteren den Faktor 4 usw. bis 1. Wenn Sie alle sechs Lotterien gewichtet haben, spielt der Computer jede Lotterie einmal für Sie. Gewinnt eine Lotterie, so bekommen Sie den Punktewert **mal dem Wertungsfaktor** auf ihrem Punktekonto gutgeschrieben. Im obigen Fall würden Sie also $24 \times 4 = 96$ Punkte erhalten.

Je günstiger eine Lotterie ist, desto höher sollten Sie diese also gewichten! D. h. 6 steht für die beste und 1 für die schlechteste Lotterie.

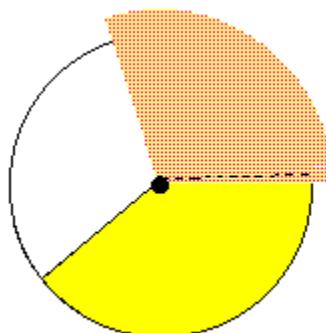
Hier ist Ihr Können gefragt. Wie Sie den Wertungsfaktor wählen, wird Ihnen nachher am Computer demonstriert.

Bei der Hälfte der Lotterien bekommen Sie jedoch Punkte **abgezogen**! In diesen Lotterien ist der Punktewert mit einem Minuszeichen versehen. Die gelbe Fläche bezeichnet hier also Ihr **Verlustrisiko**. Bei diesen Lotterien müssen Sie umdenken.

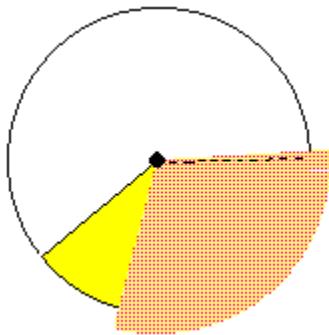
Bei vielen Lotterien können Sie das Glücksrad nicht vollständig sehen, sondern ein Teil desselben ist verdeckt. Bei diesen Lotterien wird wiederum durch Zufall bestimmt, wie die verdeckende Fläche zu liegen kommt. Wird also etwa die Gewinnfläche wie folgt ausgelost,



so kann die Verdeckung im einen Extremfall so darüber gelegt werden:



und im anderen Extremfall so:



Alle Zwischenpositionen sind natürlich genauso möglich.

Bei einem weiteren Teil der Lotterien ist gar kein Glücksrad zu sehen, weder eine verdecktes noch ein unverdecktes. Statt dessen ist die Wahrscheinlichkeit einer jeden Lotterie **mit Worten beschrieben**. So könnte etwa unter einem Punktwert „geringe Wahrscheinlichkeit“ stehen. Handelt es sich um Gewinnpunkte, so ist die Chance diese zu gewinnen also „gering“. Handelt es sich um Verlustpunkte, gilt das gleiche: Das Risiko, dass Ihnen diese Punkte abgezogen werden, ist „gering“.

Woher kommen diese Wahrscheinlichkeitsbeschreibungen? Diese Beschreibungen stammen von einer anderen Versuchsperson, die die exakten Chancen bzw. die exakten Risiken kannte. Diese Person hat die Lotteriewahrscheinlichkeiten so gut wie möglich beschrieben.

Wenn Sie irgendwelche Fragen zum Experiment haben oder Ihnen irgendetwas nicht völlig klar ist, haben Sie bitte keine Scheu, jederzeit zu fragen!!

Viel Erfolg!!

Tabelle A1: Aufgaben von Experiment III

P	ϵ								
.20	9	.17	87	.83	24	.11	11	.4	94
.10	33	.28	4	.77	32	.37	75	.21	34
.09	59	.88	7	.92	28	.44	70	.46	22
.19	54	.58	10	.91	49	.30	49	.67	24
.17	97	.70	95	.93	63	.62	27	.75	43
.92	28	.72	67	.74	93	.91	12	.83	74
$V(EV, \epsilon) = .667$		$V(EV, \epsilon) = .867$		$V(EV, \epsilon) = .933$		$V(EV, \epsilon) = .867$		$V(EV, \epsilon) = .533$	
$V(EV, P) = .600$		$V(EV, P) = .600$		$V(EV, P) = .533$		$V(EV, P) = .600$		$V(EV, P) = 1.00$	

P	ϵ								
.03	81	.47	92	.12	67	.11	7	.08	44
.30	97	.33	14	.33	33	.35	4	.87	9
.91	84	.70	76	.45	33	.29	29	.27	72
.86	74	.64	54	.21	20	.49	65	.88	28
.48	25	.83	79	.64	38	.82	42	.34	89
.94	92	.91	70	.79	73	.55	94	.95	62
$V(EV, \epsilon) = .667$		$V(EV, \epsilon) = .733$		$V(EV, \epsilon) = .733$		$V(EV, \epsilon) = .867$		$V(EV, \epsilon) = .667$	
$V(EV, P) = .933$		$V(EV, P) = .867$		$V(EV, P) = .933$		$V(EV, P) = .867$		$V(EV, P) = .800$	

A 3 Abfrage des subjektiven Verständnisses verbaler Wahrscheinlichkeitsbeschreibungen in Experiment III am Beispiel von Liste 1.

	beste Schätzung:	Intervall	
		von:	bis:
eher wahrscheinlich
sehr unwahrscheinlich
gering wahrscheinlich
sehr wahrscheinlich
weniger wahrscheinlich
eher unwahrscheinlich

Über die Wahrscheinlichkeiten habe ich mich besser informiert gefühlt durch [zutreffendes bitte ankreuzen]:

(a) die verdeckten Glücksräder (b) die Beschreibungen in Worten

Tabelle A2: Verbale Wahrscheinlichkeitsbeschreibungen in Experiment III.

P	Liste 1	Liste 2	Liste 3
.03	sehr unw.	fast ausgeschlossen	sehr kleine W.
.04	sehr unw.	sehr unw.	sehr geringe W.
.08	sehr unw.	sehr unw.	geringe W.
.09	sehr unw.	mit sehr geringer W.	sehr kleine W.
.10	sehr unw.	geringe W.	geringe W.
.11	sehr unw.	mit geringer W.	ziemlich geringe W.
.12	gering w.	unw.	kleine W.
.17	sehr unw.	eher geringe W.	geringe W.
.19	sehr unw.	eher geringe W.	geringe W.
.20	sehr unw.	mit geringer W.	geringe W.
.21	gering w.	eher unw.	kleine W.
.27	eher unw.	eher unw.	geringe W.
.28	eher unw.	mit geringer W.	geringe W.
.29	eher unw.	eher unw.	geringe W.
.30	eher unw.	eher unw.	geringe W.
.33	weniger w.	eher unw.	kleine W.
.34	eher unw.	eher unw.	kleine W.
.35	weniger w.	eher unw.	fast fifty-fifty
.37	eher unw.	eher unw.	eher geringe W.
.44	eher unw.	mit geringerer W.	circa fifty-fifty
.45	eher unw.	mit geringerer W.	fifty-fifty
.46	eher unw.	eher unw.	fifty-fifty
.47	eher unw.	mit ausgeglichener W.	fifty-fifty
.48	eher unw.	mit ausgeglichener W.	fifty-fifty
.49	eher unw.	mit ausgeglichener W.	fifty-fifty
.55	eher w.	mit ausgeglichener W.	fifty-fifty
.58	eher w.	mit größerer W.	circa fifty-fifty
.62	eher w.	eher w.	circa fifty-fifty
.64	eher w.	w.	recht hohe W.
.67	eher w.	mit hoher W.	recht hohe W.
.70	eher w.	w.	hohe W.
.72	eher w.	mit größerer W.	hohe W.
.74	eher w.	sehr w.	hohe W.
.75	eher w.	sehr w.	hohe W.
.77	sehr w.	w.	hohe W.
.79	sehr w.	sehr w.	hohe W.
.82	sehr w.	mit großer W.	hohe W.
.83	sehr w.	mit großer W.	hohe W.
.86	sehr w.	mit großer W.	hohe W.
.87	sehr w.	sehr w.	hohe W.
.88	sehr w.	sehr w.	sehr große W.
.91	sehr w.	fast sicher	sehr hohe W.
.92	sehr w.	höchst w.	sehr hohe W.
.93	sehr w.	fast sicher	sehr große W.
.94	sehr w.	fast sicher	sehr hohe W.
.95	sehr w.	höchst w.	sehr hohe W.

W. = Wahrscheinlichkeit, unw. = unwahrscheinlich, w. = wahrscheinlich.

A 4 Bestimmung des Punkteäquivalents einer *SFE* in der Bedingung „verbal“

Eine *SFE* entspicht dem Punkteverlust, der zu erwarten ist, wenn von einer Referenzstrategie in einer wohldefinierten Weise abgewichen wird (vgl. 9.3.2) Für die Bedingung „verbal“ wurde als Referenzstrategie gewählt, dass die Lotterien einer Aufgabe anhand ihrer wahrgenommenen Attraktivität geordnet werden. Dabei soll für die Attraktivität einer Lotterie gelten

$$\text{Attraktivität} = \epsilon \cdot P_{\text{Schätz.}}$$

$P_{\text{Schätz.}}$ entspricht dabei der Übersetzung in einen Prozentwert, die die Versuchsperson am Ende des Versuchs für die jeweilige verbale Wahrscheinlichkeitsbeschreibung vorgenommen hat. Da diese Schätzungen von Versuchsperson zu Versuchsperson verschieden waren, musste für jede individuell bestimmt werden, wieviele Punkte einer *SFE* entsprechen. Somit wurde für jede Versuchsperson einzeln bestimmt, wieviele Punkte sie in jeder Aufgabe gemacht hätte, wenn sie der Referenzstrategie gefolgt wäre. Sodann wurde berechnet, wieviele Punkte sie gemacht hätte, wenn sie von der Ordnung, die die Referenzstrategie ergibt, dergestalt abgewichen wäre, dass sie (i) die Ränge 1 und 2, (ii) die Ränge 2 und 3, (iii) die Ränge 3 und 4, (iv) die Ränge 4 und 5 und (v) die Ränge 5 und 6 vertauscht hätte. Für jede einzelne Aufgabe ergibt sich nun das *SFE*-Äquivalent als die Summe der Gewinnunterschiede zwischen der Referenzstrategie und den fünf Abwandlungen derselben, geteilt durch fünf. Das *SFE*-Äquivalent für alle zehn Aufgaben ergibt sich dann als Mittelwert der *SFE*-Äquivalente der Einzelaufgaben.

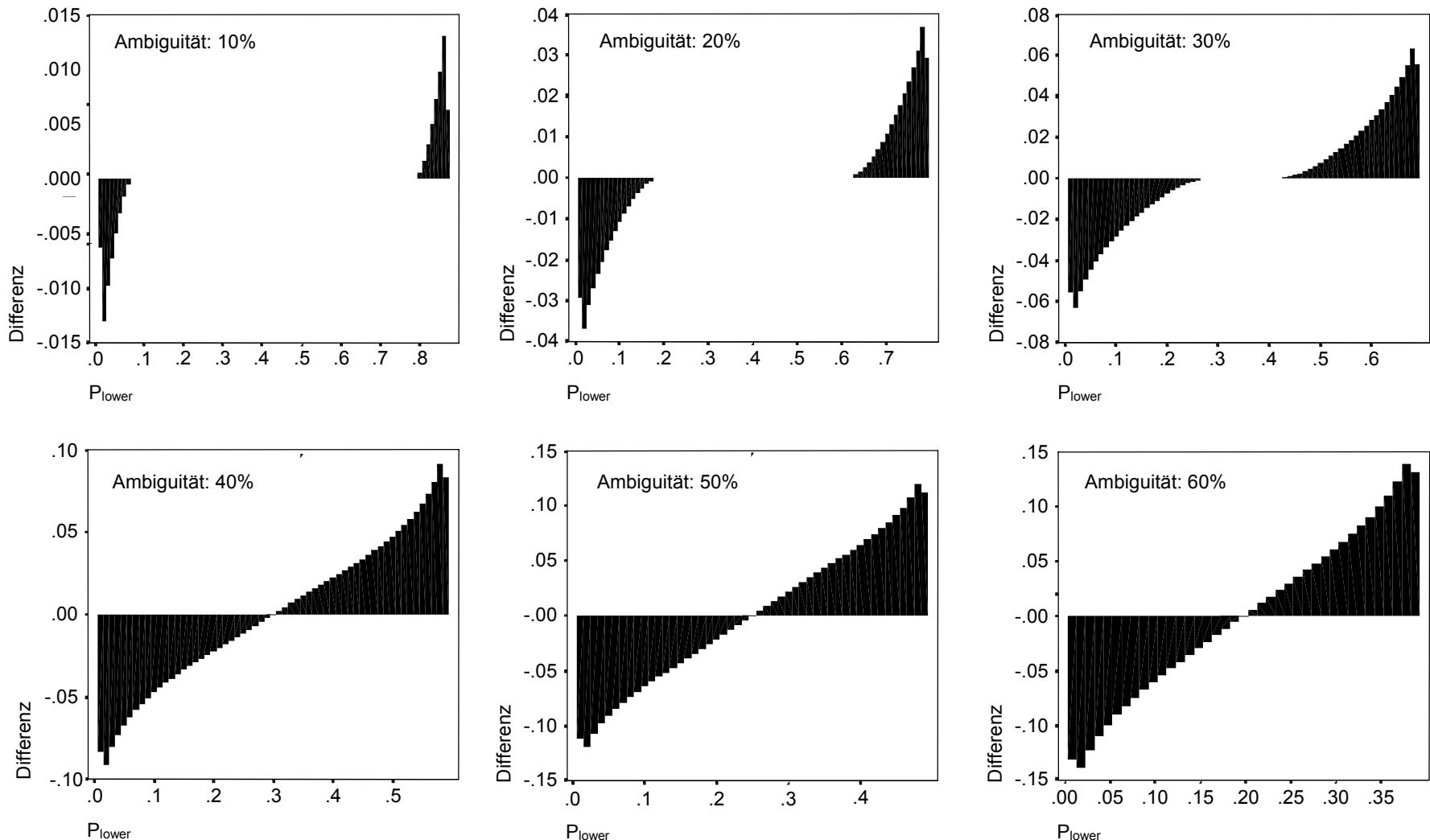


Abb. A1 Die Erwartungswerte für die Lotteriewahrscheinlichkeit P , in Abhängigkeit vom Grad der Ambiguität und der Untergrenze der Wahrscheinlichkeit P_{lower} . Da diese Werte nur minimal von der Winkelhalbierenden abweichen, ist hier die Differenz zum Schätzwert $P_{\text{lower}} + a/2$ angegeben. Bei einer Ambiguität von 20% und $P_{\text{lower}} = .01$ etwa liegt der Erwartungswert für P rund .03 niedriger als $P_{\text{lower}} + a/2$, beträgt also rund $.01 + .2/2 - .03 = .08$.

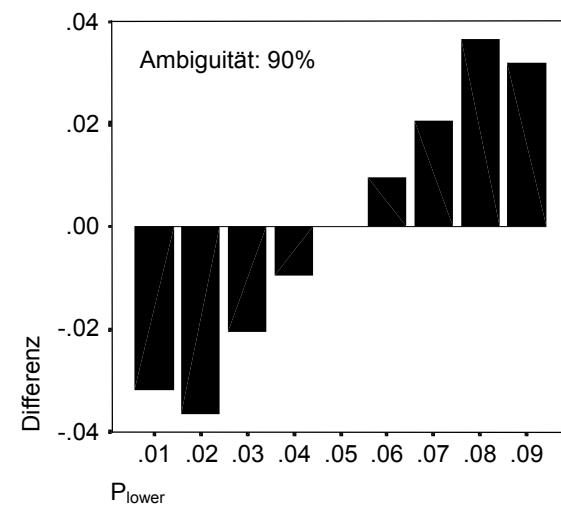
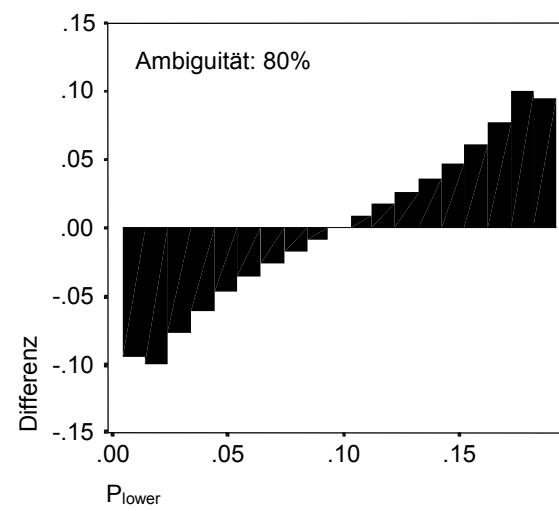
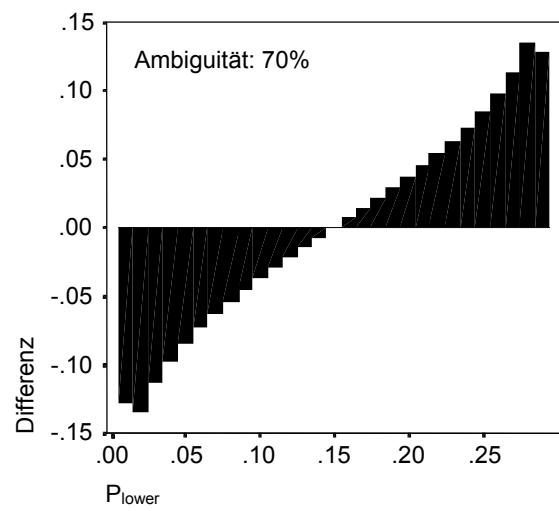


Abb. A1 (Fortsetzung)

GLOSSAR

ϵ	Wert/Punktzahl einer Lotterie
ϵ -Strategie	Rangordnung der Lotterien allein nach deren Werten.
Δ	Delta. θ wird so geschätzt, dass das Vp-Verhaltens möglichst gut nachgezeichnet wird. In der Regel kann das von der Vp gezeigte Verhalten jedoch nicht hundertprozentig genau repliziert werden, sondern die Replikation enthält einen Fehler. Δ ist das Maß für die Größe dieses Fehlers. Wenn aus dem Datensatz einer Vp eine beliebige Lotterie herausgegriffen wird, so gibt Δ an, welcher absolute Unterschied zwischen dem Rangplatz den die Vp dieser Lotterie gegeben hat und demjenigen, den das Modell repliziert, zu erwarten ist.
θ	Theta. Maß dafür, ob sich Wahlen stärker an der Wahrscheinlichkeits- oder stärker an der Wertdimension von Handlungsoptionen orientieren (3.2.2). Werte größer eins stehen für ersteres; ein Wert kleiner eins steht für eine umso stärkere Orientierung an der Wertdimension, je näher der Wert bei null liegt.
a	Stärke der Ambiguität in den Simulationen; ergibt sich als $P_{\text{upper}} - P_{\text{lower}}$. a ist von der Funktion her identisch mit $Ambig$, wird jedoch nicht als Prozent- sondern als Wahrscheinlichkeitswert angegeben.
$Ambig$	Größe der Verdeckungsfläche bei einer ambigen Lotterie mit graphisch dargestellter Wahrscheinlichkeit (vgl. Abb. 17). Die Größe von $Ambig$ wird in Prozentwerten angegeben; dabei bedeutet etwa 50%, dass die Hälfte des Glücksrades verdeckt ist.
Aufmerksamkeitshypothese	Vermutung darüber, warum es im Gewinnbereich bei hoher Ambiguität zu einem Anstieg von θ kommt: Hohe Ambiguität lenkt die Aufmerksamkeit der Entscheider auf die Wahrscheinlichkeitsinformation. In Folge dessen kommt ihr bei der Entscheidung ein höheres Gewicht zu. Vgl. 9.4.
Basismodell	Zur Schätzung von θ wird angenommen, dass Vpn die Lotterien gemäß der Größe $u(\epsilon) \cdot w(P_{\text{Schätz}})^{\theta}$ ordnen. Dieses Basismodell bedarf einer genaueren Spezifikation, d. h. es muss bestimmt werden, wie der Zusammenhang zwischen dem Lotteriewert ϵ und dessen Nutzen $u(\epsilon)$ ist, wie die gegebene Wahrscheinlichkeitsinformation in eine Schätzung $P_{\text{Schätz}}$ der Gewinnwahrscheinlichkeit überführt wird und wie diese gegebenenfalls wiederum in ein Entscheidungsgewicht $w(P_{\text{Schätz}})$ transformiert wird.

EU	Erwarteter Nutzen; von <i>expected utility</i> .
EV	Erwartungswert; von <i>expected value</i> .
Konservatismushypothese	Vermutung darüber, warum es im Gewinnbereich bei hoher Ambiguität zu einem Anstieg von θ kommt: Hohe Ambiguität führt zu vorsichtigerem, konservativerem Verhalten. Somit sollte θ im Gewinnbereich steigen und bei Verlustaufgaben sinken (vgl. 9.4).
Konservativ	Strategie, die Lotterien im Gewinnbereich anhand des Terms $\epsilon \cdot P_{\text{Sicht}}$ zu ordnen und im Verlustbereich anhand des Terms $\epsilon \cdot (P_{\text{Sicht}} + \text{Ambig})$. Die Höhe der Gewinn- bzw. Verlustwahrscheinlichkeit wird also immer maximal pessimistisch geschätzt.
Normal	Strategie, die Lotterien anhand des Terms $\epsilon \cdot (P_{\text{Sicht}} + \text{Ambig} / 2)$ zu ordnen. Diese Strategie geht also im Prinzip davon aus, dass unter der Hälfte der Verdeckungsfläche Gewinnfläche liegt (bei Verlustaufgaben: Verlustfläche).
P	<i>Probability</i> : Wahrscheinlichkeit, dass der Punktewert der Lotterie gewonnen (bzw. bei Verlustaufgaben verloren) wird.
P_a	Ankerwert, von dem aus gemäß der Venture Theory die mentale Generierung eines Entscheidungsgewichtes startet. In den <i>Venture</i> -Modellen wird als Ankerwert stets $P_{\text{Sicht}} + \text{Ambig} / 2$ angenommen.
P_{Sicht}	Der Teil der Gewinnfläche (bei Verlustaufgaben Verlustfläche), der sichtbar ist, also nicht von der Verdeckungsfläche <i>Ambig</i> verborgen wird. Bei fehlender Ambiguität sind P und P_{Sicht} identisch.
Rankingaufgabe	Hier sind die Lotterien einer Aufgabe in eine Rangordnung zu bringen. Alle Lotterien der Aufgabe werden einmal gespielt. Für jede Lotterie gilt dabei: Falls sie gewinnt (bzw. bei Verlustaufgaben: verliert), erbringt sie $\epsilon \cdot \text{Rangplatz}$ Punkte.
SEUT	Subjective Expected Utility Theory.
SFE	Standard-Fehlereinheit. Maß für die Leistung von Vpn bzw. Antwortstrategien. Eine <i>SFE</i> entspricht dem erwarteten Punkteverlust, der sich gegenüber <i>Normal</i> ergibt, wenn zwei zufällig ausgewählte benachbarte Ränge vertauscht werden (vgl. 9.3.2).

$u(\epsilon)$	Nutzen (<i>utility</i>) von ϵ .
$V(x, y)$	Korrelation zwischen zwei Rangreihen. Die eine Rangreihe entsteht dabei dadurch, dass die Elemente einer Menge anhand des Attributes x geordnet werden; die andere Rangreihe entsteht dadurch, dass dieselben Elemente anhand des Attributes y geordnet werden. Der Wert 0 zeigt eine perfekte negative Korrelation zwischen den so erzeugten Rangreihen an, 1 steht für einen perfekten gleichsinnigen Zusammenhang. Entsprechend steht .5 für einen fehlenden Zusammenhang.
	<ul style="list-style-type: none"> - ϵ: Rangreihe, die sich bei der Ordnung anhand der Lotteriewerte ergibt. - C: Rangreihe, die die V_p vorgenommen hat (<i>Choice</i>). - EV: Rangreihe, die sich aus der Ordnung der Lotterien anhand ihrer Erwartungswerte ergibt. - P: Rangreihe, die sich bei der Ordnung anhand der Lotteriewahrscheinlichkeiten ergibt.
$V_p(n)$	Versuchsperson(en).
$w(P)$	Entscheidungsgewicht w in das die Wahrscheinlichkeit P überführt wird. w regelt das Attraktivitätsverhältnis verschiedener Wahrscheinlichkeiten zueinander. So wird in der Regel eine Gewinnwahrscheinlichkeit von .002 weniger als doppelt so attraktiv sein wie eine von .001 (vgl. 3.3).

Wahlaufgabe

Bei diesem Aufgabentyp gilt es, eine der Lotterien, die gemeinsam eine Aufgabe bilden, auszuwählen. Nur diese wird dann gespielt, und falls sie gewinnt (bzw. bei Verlustaufgaben: verliert), wird ihr Punktwert ϵ gutgeschrieben.

LITERATUR

- Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting*, 1, 257-269.
- Birnbaum, M. H. (1983). Base rates in Bayesian inference: signal detection analysis of the cab problem. *American Journal of Psychology*, 96, 85-94.
- Björkman, M. (1994). Internal cue theory: calibration and resolution of confidence in general knowledge. *Organizational Behavior and Human Decision Processes*, 58, 386-405.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Berlin: Springer.
- Bortz, J., Lienert, G. A. & Boehnke, K. (1990). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.
- Budescu, D. V. & Wallsten, T. S. (1990). Dyadic decisions with numerical and verbal probabilities. *Organizational Behavior and Human Decision Processes*, 46, 240-263.
- Budescu, D. V., Weinberg, S. & Wallsten, T. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 281-294.
- Budescu, D. V. & Weiss, W. (1987). Reflection of transitive and intransitive preferences: a test of prospect theory. *Organizational Behavior and Human Decision Processes*, 39, 184-202.
- Bundesgesundheitsamt (1991). *Verbesserung der Beipackzettel. Eine Initiative des Bundesgesundheitsamtes*. Berlin: Bundesgesundheitsamt.
- Camerer, C. & Weber, M. (1992). Recent developments in modeling preferences: uncertainty and ambiguity. *Journal of Risk and Uncertainty*, 5, 325-370.
- Czerlinski, J., Gigerenzer, G. & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd und die ABC Research Group (Hrg.) *Simple Heuristics that Make us Smart*. New York: Oxford University Press.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571-582.
- Dawes, R. M. (1988). *Rational Choice in an uncertain world*. San Diego: Harcourt Brace Jovanovich.
- Dawes, R. M. & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95-107.
- Dröscher, C. (1993). *Fuzzy Logic. Methodische Einführung in krauses Denken*. Reinbeck: Rowohlt.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, 51, 380-417.
- Einhorn, H. J. & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review*, 92, 433-461.
- Eisenführ, F. & Weber, M. (1993). *Rationales Entscheiden*. Berlin: Springer.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75, 643-669.

Erev, I., Bornstein, G. & Wallsten, T. S. (1993). The negative effect of probability assessment on decision quality. *Organizational Behavior and Human Decision Processes*, 55, 78-94.

Erev, I. & Cohen, B. L. (1990). Verbal versus numerical probabilities: efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes*, 45, 1-18.

Erev, I., Wallsten, T. S. & Budescu, D. V. (1994). Simultaneous over- and underconfidence: the role of error in judgment processes. *Psychological Review*, 101, 519-527.

Estes, W. (1964). Probability learning. In A. W. Melton (Hrg.) *Categories of Human Learning*. New York: Academic Press.

Fillenbaum, S., Wallsten, T. S., Cohen, B. L. & Cox, J. A. (1991). Some effects of vocabulary and communication task on the understanding and use of vague probability expressions. *American Journal of Psychology*, 104, 55-60.

Frisch, D. & Baron, J. (1988). Ambiguity and rationality. *Journal of Behavioral Decision Making*, 1, 149-157.

Gigerenzer, G. (1996). Why social context matters to rationality. In P. B. Baltes & U. Staudinger (Hrg.) *Interactive Minds. Life-Span Perspectives on the Social Foundation of Cognition*. Cambridge: Cambridge University Press.

Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, 103, 650-669.

Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, 98, 506-528.

González-Vallejo, C. G., Erev, I. & Wallsten, T. S. (1994). Do decision quality and preference order depend on whether probabilities are verbal or numerical? *American Journal of Psychology*, 107, 157-172.

González-Vallejo, C. & Wallsten, T. S. (1992). Effects of probability mode on preference reversal. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 855-864.

Good, I. J. (1959). Kinds of probability. *Science*, 129, 443-447.

Griffin, D. & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, 24, 411-435.

Hogarth, R. M. & Einhorn, H. J. (1990). Venture theory: a model of decision weights. *Management Science*, 36, 780-803.

Jungermann, H., Pfister, H.-R. & Fischer, K. (1998). *Die Psychologie der Entscheidung: Eine Einführung*. Heidelberg, Berlin: Spektrum.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.

- Kahnemann, D. & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11, 143-157.
- Kühberger, A. (1994). Risiko und Unsicherheit: Zum Nutzen des Subjective Expected Utility-Modells. *Psychologische Rundschau*, 45, 3-23.
- Lichtenstein, S. & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89, 46-55.
- Lienert, G. A. & Raatz, U. (1994). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lopes, L. L. (1995). On modelling risky choice: why reasons matter. In J.-P. Caverni, Bar Hillel, F. H. Barron & H. Jungermann (Hrg.) *Contributions to Decision Making*. Amsterdam: Elsevier.
- Machiavelli, N. (1513/1986). *Der Fürst / Il Principe*. Ditzingen: Reclam.
- MacCrimmon, K. R. (1968). Descriptive and normative implications of the decision-theory postulates. In K. Borch & J. Mossin (Hrg.) *Risk and Uncertainty*. London: MacMillan. [Zit. nach Camerer & Weber (1992)]
- McClelland, A. G. R. & Bolger, F. (1994). The calibration of subjective probabilities: theories and models 1980-1993. In G. Wright & P. Ayton (Hrg.) *Subjective Probability*. Chichester: John Wiley.
- Mosteller, F. & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science*, 5, 2-34.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109-133.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Pfister, H.-R. (1994). Noch immer von Nutzen - das SEU-Modell. *Psychologische Rundschau*, 45, 157-160.
- Raiffa, H. (1961). Risk, ambiguity, and the Savage axioms: comment. *Quarterly Journal of Economics*, 75, 690-694.
- Reagan, T. R., Mosteller, F. & Youtz, C. (1989). Quantitative meanings of verbal probability expression. *Journal of Applied Psychology*, 74, 433-442.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Stigler, G. J. (1950). The development of utility theory. *The Journal of Political Economy*, 58, 307-327 (Teil I) und 373-396 (Teil II).
- Tune, G. S. (1964). Response preferences: a review of some relevant literature. *Psychological Bulletin*, 61, 286-302.
- Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, 79, 281-299.
- Tversky, A. & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102, 269-283.
- Tversky, A. & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105-110.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124-1131.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453-458.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk under Uncertainty*, 5, 297-323.

Tversky, A., Sattath, S. & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95, 371-384.

von Neumann, J. & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Wallsten, T. S. (1971). Subjectively expected utility theory and subjects' probability estimates: use of measurement-free techniques. *Journal of Experimental Psychology*, 88, 31-40.

Wallsten, T. S., Budescu, D.V. & Erev, I. (1988). Understanding and using linguistic uncertainties. *Acta Psychologica*, 68, 39-52.

Wallsten, T. S., Budescu, D. V. & Tsao, C. J. (1997). Combining linguistic probabilities. In: Scholz, R. W. & Zimmer, A. (Hrg.), *Qualitative Aspects of Decision Making*. Lengerich, Pabst.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R. & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, 115, 348-365.

Wallsten, T. S., Budescu, D., Zwick, R. & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, 31, 135-138.

Wu, G. (1994). An empirical test of ordinal independance. *Journal of Risk and Uncertainty*, 9, 39-60.

Zeigarnik, B. (1927). Über das Behalten von erledigten und unerledigten Handlungen. *Psychologische Forschung*, 9, 1-85.

Zimmer, A. (1983). Verbal vs. numerical processing of subjective probabilities. In: Scholz, R. W. (Hrg.), *Decision Making under Uncertainty*. North-Holland: Elsevier, 159-182.