

ÜBER DIE UNABHÄNGIGKEIT ZWISCHEN
SUBJEKTIVEN UND OBJEKTIVEN
WAHRSCHEINLICHKEITEN

von
Jürgen KRIZ

mit einem Kommentar von
Anatol RAPOPORT

Forschungsbericht No. 17
Juli 1968

Allen, die zum Gelingen dieser Arbeit beigetragen haben, sei hiermit herzlich gedankt. Insbesondere danke ich der Schulbehörde Hamburg, sowie den Schulleitungen der Caspar-Voght-Schule, der Hansa-Schule, der Kirchenpauerschule und der Wichernschule für die großzügige Unterstützung bei der Datenerhebung. Dem Institut für Höhere Studien und Wissenschaftliche Forschung - besonders meinem Abteilungsleiter Herrn Dr. STÖCKELLE - danke ich für die ideelle und materielle Hilfe bei der Durchführung dieses Projektes. Ganz besonders herzlich danke ich Herrn Prof. Dr. WRIGHT (Adelaide) und Herrn Prof. Dr. RAPOPORT (Michigan) für zahlreiche Hinweise und die Möglichkeit, die vorliegende Arbeit ausführlich diskutieren zu können.

Wien, im Juli 1968

Jürgen Kriz

Zur Geschichte der Entscheidungstheorie

1955 publizierte EDWARDS (3) eine Arbeit, in der er vier Modelle für ein Entscheidungsverhalten unter Risiko diskutierte. Diese Modelle spiegeln gleichzeitig die geschichtliche Entwicklung der Entscheidungstheorie bis zu diesem Zeitpunkt wider.

Allen vier Modellen liegt derselbe Gedankengang zugrunde: Eine Entscheidung habe N Alternativen ($A_1, A_2, \dots, A_i, \dots, A_n$). Jede mögliche Konsequenz einer Alternative A_i wird durch ein Produkt zweier Zahlen repräsentiert. Die eine Zahl stellt den "Wert" dieser Konsequenz dar, die andere die Wahrscheinlichkeit, mit der diese Konsequenz auftritt. Die Produkte werden über alle Konsequenzen einer Alternative summiert. Die Modelle postulieren nun, daß man bei einer Entscheidung diese Produktsummen miteinander vergleicht, und die Alternative mit der maximalen Produktsumme gewählt wird.

Die Modelle unterscheiden sich darin, daß die Wahrscheinlichkeit subjektiv oder objektiv sein kann. Ebenso läßt sich der Wert einer Konsequenz entweder objektiv (als Geldwert) oder subjektiv (als Utility) ausdrücken. Jedes der vier Modelle basiert auf einer der vier Kombinationsmöglichkeiten.

Im traditionellen Modell wird die Produktsumme aus objektiven Wahrscheinlichkeiten und Geldwert maximiert. Dies Modell ist streng normativ, d.h. es schreibt vor, wie man sich entscheiden sollte. Schon vor zwei Jahrhunderten argumentierte BERNOULLI, daß bei Entscheidungen eher die erwartete Utility maximiert wird. Auch VON NEUMANN & MORGENSTERN (21) vertraten das EU-Modell (expected utility). Ihr großes Verdienst war es, die

Entscheidungstheorie in das Blickfeld der Sozialwissenschaften zu rücken, Methoden für eine Utility-Messung vorzuschlagen und einen speziellen Zweig der Entscheidungstheorie, die Spieltheorie zu begründen.

Das dritte Modell - es kombiniert subjektive Wahrscheinlichkeiten mit objektivem Geldwert - ist praktisch bedeutungslos. Es wurde von PRESTON & BARATTA (13) zur Messung der SW¹ benutzt. Das vierte Modell wurde von EDWARDS (3) als SEU-Modell (subjectively expected utility) eingeführt. Es ist bis heute das meist angewandte, getestete und diskutierte Modell.

Viele Experimente auf dem Gebiet der Entscheidungstheorie befassen sich mit der Erprobung und Testung dieser Modelle. Übersichten über die Literatur geben EDWARDS (4,5,6), PETERSON & BEACH (11) und VLEK (18). Die meisten Arbeiten befassen sich mit der Messung der SW und (bzw. oder) der Utility, obwohl das SEU-Modell auch ohne die Messung dieser beiden Variablen getestet werden kann (z.B. COOMBS, BEZEMBINDER & GOODE (2)).²

Zur Messung der subjektiven Wahrscheinlichkeit

Eine Beschäftigung mit der SW wirft bereits bei oberflächlicher Betrachtung zwei Probleme auf:³

(1) Objektive Wahrscheinlichkeit ist nur in dem abgeschlossenen Intervall $[0,1]$ definiert; gilt diese Beschränkung auch für die subjektive Wahrscheinlichkeit?

-
- 1) Im Text wird gelegentlich "subjektive Wahrscheinlichkeit" mit "SW" und "objektive Wahrscheinlichkeit" mit "OW" abgekürzt werden.
 - 2) genauer gesagt: einige Konsequenzen des SEU-Modells können ohne SW- bzw. Utility-Messung getestet werden.
 - 3) Der Gedankengang erfolgt in Anlehnung an EDWARDS (4) und VLEK (18)

(2a) Objektiv ist die Wahrscheinlichkeit, daß ein Ereignis eintritt, plus der Wahrscheinlichkeit, daß dies Ereignis nicht eintritt, 1.

(2b) Wenn von drei Ereignissen A, B, und C eines sicher auftreten muß, gilt objektiv : $P(A) + P(B) + P(C) = 1$. Sofern (1) und (2b) auch subjektiv gelten, folgt daraus, daß dann die subjektiven Wahrscheinlichkeiten identisch mit den objektiven sein müssen. Aus (1) und (2a) folgt dies nicht, aber die Beziehung zwischen SW und OW muß ähnlich aussehen wie in Abb 1.

Die Experimente zur Messung der SW sehen fast alle gleichartig aus. Das Prinzip beruht meist auf der Annahme, daß es je zwei Werte x und y so gibt, daß eine Vpn indifferent ist, ob sie x mit einer bestimmten Wahrscheinlichkeit zu bekommen y sicher zu bekommen vorziehen soll oder nicht. X und y können dabei sowohl objektive Werte sein (vergl. S 2), wie auch Utilities. Letzteres setzt aber voraus, daß man die Utility-Funktion der betreffenden Vp schon kennt.¹

Das erste Experiment dieser Art wurde 1948 von PRESTON & BARATTA (13) durchgeführt. Ein bestimmter Betrag (y) konnte von jeder Vp geboten werden, um damit das Anrecht auf ein Spiel zu haben, in dem der Betrag x mit der Wahrscheinlichkeit P (oder nichts mit der Wahrscheinlichkeit 1-P) gewonnen werden konnte. Die Ergebnisse zeigten eine relativ gute Übereinstimmung zwischen SW und OW, mit einer Überschätzung niedriger und einer Unterschätzung hoher Wahrscheinlichkeiten. Ähnliche Ergebnisse bei ähnlichem Versuchsaufbau² fanden u.a. GRIFFITH (7), ATTNEAVE (1), SPROWLS (15), MOSTELLER & NOGEE (10) und TVERSKY (17) (vergl. Abb.2).

1) näheres siehe EDWARDS (4, 5, 6)

2) wobei die SW z.T. nur ein Nebenergebnis anderer Messungen war.

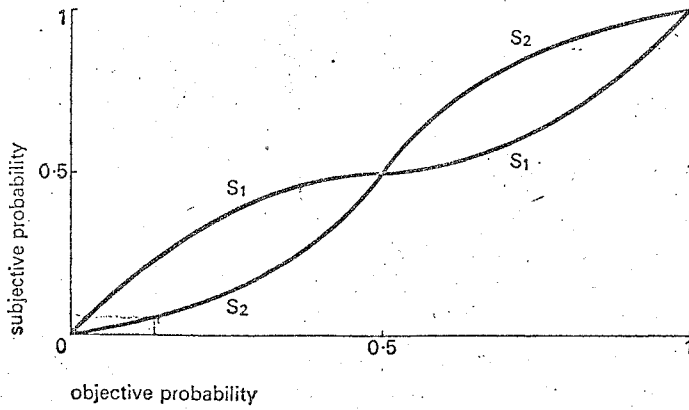


Abb. 1
Hypothetische SW-Kurven
aus EDWARDS & TVERSKY
(6)

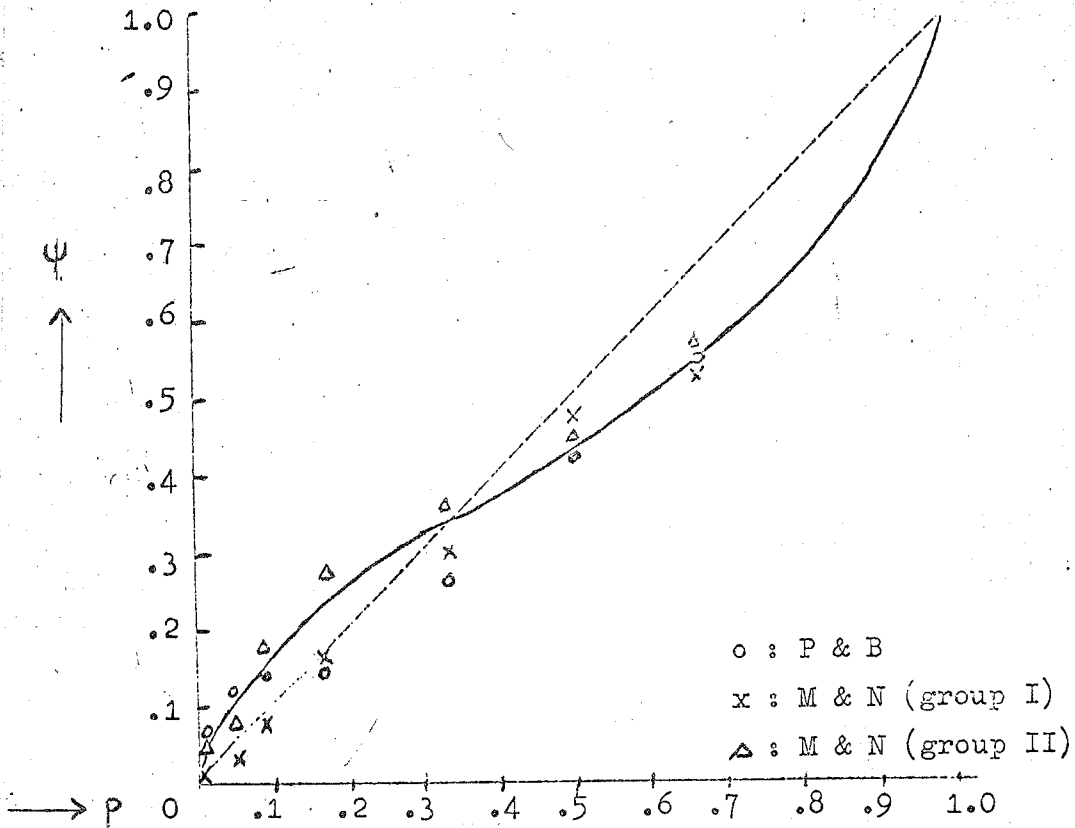


Abb. 2 (aus VLEK (18))

Zusammenhang zwischen SW und OW, wie er in den Arbeiten von PRESTON & BARATTA (P & B) und MOSTELLER & NOGEE (M & N) gefunden wurde.

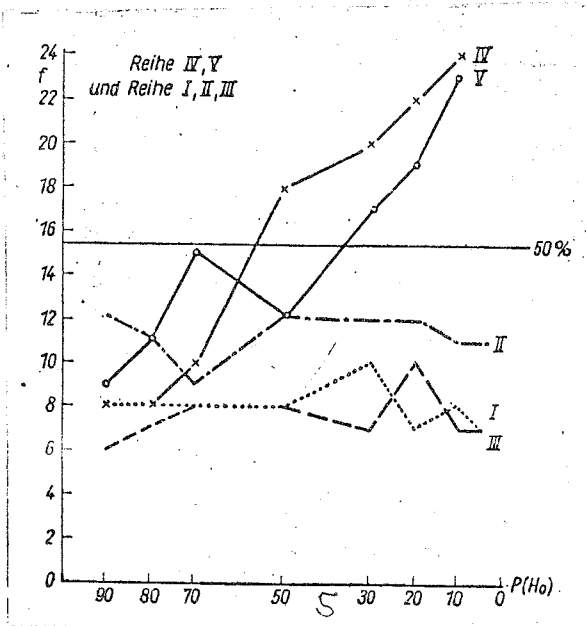


Abb. 3 (aus KRIZ(9))

Zusammenhang zwischen OW und der Häufigkeit der "NEIN"-Urteile.

Bei Kurve I bis III variiert die Stichprobengröße;
bei Kurve IV und V variiert das Gesamtverhältnis.

Eine andere Art von Experimenten, die auch unter SW-Messung geführt werden, findet man bei SHUFORD (14), STEVENS & GALANTER (16), PHILIP (12) und EDWARDS (6). Dabei muß eine Vp den relativen Anteil von Elementen (z.B. "rote Punkte") in einer Menge (z.B. rote und blaue Punkte) schätzen. Der Begriff der subjektiven Wahrscheinlichkeit scheint für solche Experimente aber falsch gewählt zu sein.

Weniger gute Übereinstimmung zwischen SW und DW wurde in letzter Zeit von zahlreichen Autoren im Zusammenhang mit BAYES' Theorem gefunden. (vergl. S. 40)

Sieht man von den Ergebnissen in diesem Bereich aber ab, so läßt sich zusammenfassend sagen, daß die meisten Experimente eine gute Übereinstimmung zwischen SW und DW gezeigt haben.

Entscheidungen im Alltag und der zwei-Stichproben-Vergleich

Es kommt selten vor, daß wir in unserem Alltagsleben Entscheidungen von der Art gegenüberstehen, wie sie bei den berichteten Experimenten getroffen werden mußten. Viel häufiger müssen Diskriminierungs-Entscheidungen gefällt werden. Es steht dabei eine bestimmte Anzahl Alternativen zur Wahl, wobei man über die Alternativen selbst nur eine beschränkte Information zur Verfügung hat. Ein selbstverständlich sehr vereinfachtes - Modell solcher Entscheidungssituationen ist der zwei-Stichproben-Vergleich. Dabei muß für zwei Stichproben S_a und S_b entschieden werden, ob sie aus einer Grundgesamtheit stammen, oder aber die Wahrscheinlichkeit dafür so gering ist, daß sie aus zwei verschiedenen Grundgesamtheiten A und B stammen. Z.B. liegt eine solche Entscheidungssituation vor, wenn eine Hausfrau neu in einen Ort zieht und zuerst in zwei Läden abwechselnd ihre Eier kauft, bis sie auf Grund der Stichproben entscheidet, daß der Laden 1 die besseren Eier hat.

Ein ähnlicher Entscheidungsprozess liegt vielleicht vor, Wenn wir uns für den Kauf einer bestimmten Autotype, für ein Lehrbuch unter vielen gleichartigen, für eine bestimmte Tageszeitung usw. entscheiden. Dabei kann man sich den gesamten Entscheidungsprozeß aus vielen solchen zwei-Stichproben-Vergleichen zusammengesetzt denken. Die Gesamtmenge der Alternativen wird in zwei Untermengen geteilt, diese verglichen, die gewählte Untermenge wieder geteilt usw. Wenn man z.B. beobachtet, wie Vpn aus einer Reihe von Photos das schönste heraussuchen, findet man oft ein solches Vorgehen. Es ist durchaus denkbar, daß auch andere Entscheidungsvorgänge ähnlich verlaufen. Obwohl der zwei-Stichproben-Vergleich - wie betont - eine äußerst vereinfachte Darstellung solcher Entscheidungen sein soll, wollen wir noch weiter vereinfachen, indem nur Alternativmerkmale betrachtet werden. Die Elemente jeder Stichprobe haben nur zwei Eigenschaften E_1 und \bar{E}_1 ("nicht E_1 "). Alternativmerkmale sind z.B. gut-schlecht, groß-klein, hell-dunkel usw., d.h. eine Nominalskala mit zwei Punkten. Die Eigenschaften müssen durchaus nicht nur nominal meßbar sein, es ist aber erforderlich, daß die Zuordnung der Elemente zu den beiden Alternativen eindeutig erfolgt.

Subjektive Wahrscheinlichkeit im zwei-Stichprobenvergleich für Alternativmerkmale

Der zwei-Stichprobenvergleich für Alternativmerkmale läßt sich am besten in einer Vierfeldertafel darstellen (vergl. KRIZ (9)). Auf Grund irgendwelcher Dimensionen unterscheidet man zwei Grundgesamtheiten A und B, aus denen eine Stichprobe S_a von n_a Elementen bzw. S_b von n_b Elementen entnommen wird. Die Stichprobe besteht aus a Elementen der Eigenschaft E_1 und b Elementen der Eigenschaft E_1 (bzw. c Elementen der Eigenschaft \bar{E}_1 und d Elementen der Eigenschaft E_1).

A → S _a	E ₁	E ₂	Σ
B → S _b	a	b	n _a
	c	d	n _b

Die Fragestellung lautet: (H₁) Unterscheiden sich A und B hinsichtlich ihrer Verteilung von E₁ und E₂, oder (H₀) bilden A und B hinsichtlich dieser Variablen eine Grundgesamtheit (und die beobachteten Unterschiede sind rein zufällig).

Diese Frage ist nicht mit Sicherheit zu beantworten. Daher ist es üblich, ein Signifikanzniveau anzugeben, auf dem die Entscheidung gefällt wird. Das Signifikanzniveau sagt aus, wieviel Prozent Irrtumswahrscheinlichkeit man bei einer Entscheidung für eine der Hypothesen (meist H₁) maximal eingehen will.

Die objektive Wahrscheinlichkeit dafür, daß die beiden Stichproben aus einer Grundgesamtheit stammen, berechnet sich nach dem bekannten Vier-Felder-Chi-Quadrat. Für die Beschreibung des Entscheidungsverhaltens im zwei-Stichproben-Vergleich ist daher folgende Frage interessant: Wie groß ist die subjektive Wahrscheinlichkeit, daß S_a und S_b aus derselben Grundgesamtheit stammen? Um diese Frage zu beantworten ist es nötig, die subjektive Wahrscheinlichkeit irgendwie zu messen.

Kritik an der SW-Messung

Ein typisches Beispiel dafür, wie in den meisten Experimenten das Problem der SW-Messung angegeben wurde, findet man bei VLEK:¹

1) VLEK (18), S 5 (vom Autor übersetzt)

"Sofern man festsetzt, daß die subjektiven Wahrscheinlichkeiten einer gesonderten Berechnung bedürfen, weil sie unterschiedlich von den zugeordneten objektiven Wahrscheinlichkeiten sind, dann ist der nächste Schritt, herauszufinden, welche Art einer Funktion die erste mit der letzteren verknüpft."

In dieser Argumentation wird etwas stillschweigend vorausgesetzt, was zumindest einer expliziten Formulierung bedürfte; nämlich daß es überhaupt eine Funktion gibt, die SW und OW miteinander verknüpft. Aber diese wichtige Grundannahme findet man in keiner Arbeit explizit formuliert oder diskutiert, d.h. es scheint gar nicht bewußt zu sein, daß es sich nur um eine Annahme handelt. Dafür gibt es mehrere Erklärungen, die i.W. in der Entwicklung der Entscheidungstheorie liegen:

- 1.) Die ersten Entscheidungsmodelle verwendeten nur die OW. Bei der Konzipierung des SEU-Modells wurde diese durch die SW ersetzt. Ein funktionaler Zusammenhang war evident.
- 2.) Die berichteten Experimente zeigen eine relativ gute Übereinstimmung zwischen SW und OW.
- 3.) Bei diesen Experimenten wäre die Annahme der Unabhängigkeit zwischen SW und OW nicht sinnvoll gewesen, denn sie wurden unternommen, um den Zusammenhang beider zu messen.

Es zeigt sich also, daß bei den bisherigen Experimenten die Grundannahme einer Abhängigkeit der SW von der OW sinnvoll war und mit den erhaltenen Daten in Einklang stand.

In gewisser Weise waren aber alle diese Experimente trivial. Entweder wurde die objektive Wahrscheinlichkeit direkt vorgegeben (PRESTON & BARATTA), oder es ging um

die Wahrscheinlichkeit, daß eine bestimmte Zahl gewürfelt wurde (MOSTELLER & NOGEE), eine Kugel in ein bestimmtes Fach fällt (EDWARDS), ein Glücksrad auf weißen oder schwarzen Feldern stehen bleibt (TVERSY) usw. Überall war die OW auch für Laien praktisch evident - und wovon sonst sollten die SW einer solchen Entscheidungssituation abhängen, wenn nicht von dieser so evidenten OW ?

Man kann nun aber nicht erwarten, daß irgendein in solchen Experimenten gefundener Zusammenhang auch dann gilt, wenn sich die OW selbst aus mehreren Komponenten zusammensetzt. Zumindest muß genau an dieser Stelle die Frage nach einem solchen Zusammenhang gestellt werden. Dabei zeigt sich, daß die Argumentation auf S 4 nicht ganz einwandfrei war. Es liegt ihr nämlich ebenfalls die Annahme zugrunde, daß die SW von der OW abhängen. Ohne diese Annahme können sich die SW von A, B und C (vergl. S 4) durchaus zu 1 addieren, ohne das OW und SW identisch sein müssen.

Viele neuere Experimente im Zusammenhang mit BAYES'-Theorem¹ fanden größere Abweichungen zwischen SW und OW. Die OW in solchen Experimenten ist aber sehr komplex und keineswegs subjektiv evident. Vielleicht hängt die SW hier gar nicht mehr von der OW ab. Vielleicht sind nur einige Parameter dieser komplexen OW subjektiv relevant und andere irrelevant. Vielleicht ist der so übereinstimmend gefundene "conservatism-effect" gar kein "Effekt", der kompliziert erklärt werden muß, (sondern ein Artefakt der postulierten SW-OW-Beziehung und dem ähnlichen Aufbau fast aller dieser Experimente). Vielleicht ist es müßig über die Funktion zwischen OW und SW

1) Es würde über den Rahmen dieser Arbeit hinausgehen, auf BAYES'-Theorem näher einzugehen. Es sei daher auf EDWARDS & TVERSKY (6) verwiesen.

zu streiten, weil es gar keine Funktion gibt.

Dies sind bewußt provozierende Fragen. Aber es scheint mir, daß sie erörtert werden sollten - und manche vielleicht, bevor man Hypothesen über die Funktion zwischen SW und OW diskutiert.

Das Vier-Felder-Chi-Quadrat in Abhängigkeit von verschiedenen Parametern.

Nach dem Schema von S 8 ergibt sich:

$$(1) \quad \chi^2 = \frac{(a+b+c+d)(/ad-bc/ - \sqrt{2}(a+b+c+d))^2}{(a+b)(a+c)(b+d)(c+d)}$$

Wie man zeigen kann, ist dieses χ^2 von 4 Variablen (oder Parametern) abhängig. Aus Interpretationsgründen - und in Übereinstimmung mit einer früheren Arbeit - wurde folgendes unabhängige Quadrupel gewählt:

- V_1 : n_a (die Stichprobengröße von S_a)
- V_2 : n_a/n_b (das Verhältnis der Stichprobengr. zueinander)
- V_3 : a/b (das Verhältnis in Stichprobe S_a)
- V_4 : $(a/b):(c/d)$ (das Gesamtverhältnis)

Betrachtet man jeweils den Einfluß nur einer Variablen auf das χ^2 und hält die übrigen drei Variablen konstant, so zeigt sich für V_1 und V_2 eine interessante Gesetzmäßigkeit.¹⁾ Wird die Stichprobengröße von S_a (und damit auch von S_b) k -mal so groß, ergibt sich:

$$(2) \quad \bar{\chi}^2 = \frac{k(a+b+c+d)(k^2 ad - k^2 bc)^2}{k(a+b)k(a+c)k(b+d)k(c+d)}$$

1) Bei den folgenden Betrachtungen wird die Kontinuitätskorrektur $(a+b+c+d)/2$ nicht berücksichtigt.

$$(3) \quad \bar{\chi}^2 = \frac{k^5}{k^4} \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} = k \cdot \chi^2$$

Das $\bar{\chi}^2$ wird also ebenfalls k-mal so groß.

Variiert man das Stichprobenverhältnis (V_2), so wächst das χ^2 nicht linear mit diesem Verhältnis. Das χ^2 kann über einen bestimmten Wert nicht hinauswachsen (5).

$$(4) \quad \chi^{*2} = \frac{(ak+bk+c+d)(akd-bkc)^2}{(ak+bk)(ak+c)(bk+d)(c+d)}$$

$$(5) \quad \lim_{k \rightarrow \infty} \chi^2 = \frac{(ad-bc)^2}{ab(c+d)}$$

Der Zusammenhang zwischen dem χ^2 und V_3 (dem Verhältnis in S_a) ist sehr kompliziert. Man kann aber sagen, daß ein extremeres Verhältnis die größere Zufallswahrscheinlichkeit hat. Der Zusammenhang zwischen dem χ^2 und V_4 (dem Gesamtverhältnis) ist derart, daß das χ^2 um so größer wird, je mehr das Gesamtverhältnis von 1 abweicht.

Es zeigt sich also, daß die Abhängigkeit des χ^2 von den vier Parametern recht kompliziert ist. Man kann nicht erwarten, daß der Einfluß aller vier Parameter von naiven Vpn richtig erkannt wird; ganz abgesehen davon, daß die objektive Wahrscheinlichkeit wiederum selbst eine komplizierte - wenn auch monotone - Funktion des χ^2 ist. In einer früheren Arbeit (KRIZ (9)) wurde eine Anzahl vier-Felder-Tafeln dargeboten. Den Vpn wurde erklärt, es handelte sich dabei um Stichproben, die aus der gesamten Jahresernte Orangen der Länder A und B gezogen worden seien. Die Vpn mußten angeben, ob sie die Behauptung glaubten, daß der Unterschied zwischen den Stichproben

rein zufällig sei, d.h. die Jahresernten von A und B gleich gut wären. Es zeigte sich, daß für die Häufigkeit der "Nein"-Urteile die Stichprobengröße (V_1) irrelevant ist (vergl. Abb 3, S 5).

Der Autor kam zu der Hypothese, daß ebenfalls V_2 und V_3 subjektiv irrelevant sind. Sofern man annimmt, daß die Häufigkeit der "Nein"-Urteile eine Funktion der SW ist, würde das bedeuten, daß die SW nur vom Gesamtverhältnis abhängt. Um diese und weitere Hypothesen an einer größeren Stichprobe und mit umfangreicherem Material zu überprüfen, wurde folgendes Experiment unternommen:

Material

Das Versuchsmaterial bestand aus 90 Aufgaben (vergl. Instruktion). Die 90 Aufgaben lassen sich untergliedern in:

- 5 Reihen, in denen die Stichprobengröße variiert
 - 5 Reihen, in denen das Verhältnis der Stichproben zueinander variiert
 - 2 Reihen, in denen das Verhältnis in S variiert
 - 2 Reihen, in denen das Gesamtverhältnis^a variiert
 - 1 Reihe, in der mit zunehmendem Gesamtverhältnis die Stichprobengröße abnimmt
- 17 Aufgaben, die anfangs nur den Zweck hatten, das Versuchsmaterial so aufzufüllen, daß die Reihen nicht mehr als solche erkannt werden konnten. Die 90 Aufgaben wurden in einer Zufallsverteilung dargeboten. Es wurden 35 verschiedene Reihenfolgen hergestellt, sodaß nur jeweils 4 Vpn mit derselben Aufgabe begannen.

Instruktion

In Vorversuchen mit dem Versuchsmaterial erwies es sich als ungünstig, die vier-Felder-Tafeln darzubieten. Nach den Aussagen der Vpn wurde meist die Anzahl defekter Birnen (siehe unten) zur Stichprobengröße ins Verhältnis gesetzt. Die Vpn mußten sich somit die für sie relevante Information zuerst berechnen. Um die Ergebnisse von dem Einfluß der Schwierigkeit und Fehlerhaftigkeit freizuhalten, wurde die Instruktion wie folgt vorgegeben:

"Aus der gesamten Jahresproduktion Glühbirnen der Firmen A und B wird je eine Stichprobe entnommen und einem Kontrolltest unterzogen. Dabei ergibt sich:

	Größe der Stichprobe	Anzahl der defekten Birnen	(in %)
<u>Firma A</u>	(...%)
<u>Firma B</u>	(...%)

Firma A behauptet nun, der Unterschied sei rein zufällig, und ihre gesamte Jahresproduktion, aus der die Stichprobe stammte, sei genau so gut, wie die der Firma B. Glauben Sie, daß sich angesichts des obigen Testergebnisses diese Behauptung halten läßt?

(JA / NEIN)

Bitte geben Sie an, wie groß Ihnen die Wahrscheinlichkeit dafür erscheint, daß bei zwei wirklich gleich guten Jahresproduktionen ein derartiger Unterschied zwischen den Stichproben auftreten kann.



Jede Vp erhielt diese Instruktion schriftlich. Zusätzlich wurde mündlich folgendes erklärt:

"Markieren Sie bitte Ihre Wahrscheinlichkeit durch ein Kreuz auf dieser 10cm langen Linie. Der linke Endpunkt - 0% - bedeutet, daß der Unterschied zwischen den Stichproben so groß ist, daß sie völlig unmöglich aus gleich guten Jahresproduktionen stammen können. Der Endpunkt rechts - 100% - bedeutet, daß der Unterschied mit absoluter Sicherheit bei gleich guten Jahresproduktionen auftreten kann. Ein Kreuz 3 cm vom linken Endpunkt würde also bedeuten, daß 30% Wahrscheinlichkeit besteht, daß ein solcher Unterschied bei gleich guten Jahresproduktionen auftreten kann. Haben Sie noch Fragen?"¹⁾
 Für die 90 Aufgaben standen 50 Minuten zur Verfügung, was allgemein als "reichlich" bezeichnet wurde. Wenn eine Vp mit den Aufgaben fertig war, wurde sie vom V1 aufgefordert, auf der Rückseite der Antwortblätter die Kriterien anzugeben, die für ihre Entscheidung ausschlaggebend waren.

Versuchsdurchführung und Vorauswertung

Als Versuchspersonen dienten 136 Schüler und Schülerinnen der 12. Klasse an vier Hamburger Gymnasien. Die Versuche wurden in Gruppen durchgeführt. Von den 136 Vpn hatten 17 Vpn die Aufgaben unvollständig gelöst. Sie wurden in der weiteren Auswertung nicht berücksichtigt. Zu jeder Aufgabe wurden von den restlichen 119 Vpn die "JA"- und "NEIN"-Antworten ausgezählt. Die Auswertung der Wahr-

1) Die Fragestellung und Erklärung ist statistisch nicht ganz einwandfrei. Es geht - exakt gesagt - nicht um die Wahrscheinlichkeit, daß "ein so großer Unterschied auftreten kann", sondern um die Wahrscheinlichkeit, daß "ein so großer oder größerer Unterschied auftreten kann". Diese exakte Formulierung hätte die Vpn aber nur verwirrt.

scheinlichkeitsangaben wurde nur an 103 Vpn vorgenommen, da bei 16 Vpn die Markierungen nicht eindeutig waren (Auslassungen, Doppelmarkierungen, usw.).

Auswertung

Die Auswertung wurde i.W. nach folgenden Gesichtspunkten vorgenommen:

- 1.) Ein Parameter (V_1, V_2, V_3 oder V_4) wird für die subjektive Wahrscheinlichkeit als irrelevant bezeichnet, wenn bei Aufgabenreihen, in denen dieser Parameter variiert, die subjektive Wahrscheinlichkeit konstant ist.
- 2.) Ein Parameter (V_1, V_2, V_3 oder V_4) wird für die Entscheidung als subjektiv irrelevant bezeichnet, wenn bei Aufgabenreihen, in denen dieser Parameter variiert, die Häufigkeit der "JA"-Entscheidungen (oder "NEIN"-Entscheidungen) konstant ist.

Die weiteren Gesichtspunkte seien nur noch für die SW explizit formuliert:

- 3.) Gilt (1) für mindestens einen Parameter, so hängt die subjektive Wahrscheinlichkeit nicht von der objektiven ab.
- 4.) Ist die SW von der OW unabhängig, so ist jede funktionale Beziehung zwischen beiden ein Artefakt der Aufgaben. Das bedeutet: sofern man die Parameter kennt, von denen die SW abhängt, muß es möglich sein, zu jeder beliebigen Pseudo-Funktion zwischen SW und OW Aufgaben zu entwickeln.
- 5a) Hypothese: Die SW hängt i.W. vom Gesamtverhältnis (V_4) ab.
- 5b) Gelten (3) und (5a), so folgt daraus: Aufgaben, bei denen mit zunehmender Stichprobengröße das Gesamtverhältnis abnimmt, führen zu einer Paradox-Kurve; d.h. sofern man die SW als Funktion der OW darstellt, nimmt die SW mit zunehmender OW ab.

Ergebnisse

Die verbalen Aussagen der Vpn (nach welchen Kriterien sie entschieden hatten) konnten und sollten nicht mehr als einige Anhaltspunkte für die Auswertung geben. Die meisten Vpn gaben etwa folgendes an: Für sie sei die Differenz zwischen dem Prozentsatz defekter Birnen in S_a und in dem S_b ausschlaggebend gewesen.¹⁾ Allerdings dürfen die Stichproben nicht allzu klein sein, und der Unterschied zwischen den Stichprobengrößen dürfe nicht allzu groß sein. Einige Vpn gaben aber auch an, daß sie bei allen Aufgaben die Stichprobengröße und das Stichprobenverhältnis berücksichtigt hätten.

Um einen Aufschluß über interindividuelle Unterschiede hinsichtlich der Entscheidungskriterien zu bekommen, wurden von jeder Vp folgende Korrelationen berechnet:²⁾

R_1 : Korrelation zwischen SW und DW

R_2 : Korrelation zwischen 1-SW und dem Gesamtverhältnis (V_4)

R_3 : Korrelation zwischen 1-SW und der %-Differenz.

Die Korrelationen sind in Tab. 1 wiedergegeben.

Es zeigt sich, daß bei den meisten Vpn R_3 und R_2 hoch sind, R_1 dagegen niedrig ist. Allerdings gibt es auch Vpn, bei denen R_1 hoch ist, aber R_2 und R_3 niedrig sind. Nur bei sehr wenigen Vpn sind die drei Korrelationskoeffizienten gleich groß.

Interessant ist, daß von den verbalen Aussagen der Vpn nicht auf die Korrelation geschlossen werden kann: Nicht bei allen Vpn, die angeblich Stichprobengröße und -verhältnis berücksichtigt hatten, ist R_1 höher als R_2

1) Diese Differenz wird im Text mit "%-Differenz" bezeichnet.

2) Es handelt sich um Rangkorrelationen (nach Spearman) über alle 90 Aufgaben.

Tab. 1

Vp-Nr.	R ₁	R ₂	R ₃	Vp-Nr.	R ₁	R ₂	R ₃
1	.49	.87	.92	2	.29	.55	.54
3	.41	.69	.69	4	.31	.63	.65
5	.38	.00	.01	6	.16	.28	.25
7	.49	.72	.75	8	.45	.79	.89
9	.32	.77	.82	10	.40	.59	.50
11	.45	.81	.90	12	.35	.78	.77
13	.50	.18	.08	14	.52	.75	.62
15	.32	.79	.87	17	.66	.78	.76
18	.37	.72	.78	19	.53	.58	.55
22	.36	.50	.67	23	.47	.87	.94
24	.58	-.15	-.26	25	.57	.87	.72
26	.46	.89	.94	27	.41	.68	.80
28	.48	.86	.90	30	.14	.43	.47
31	.48	.83	.93	32	.20	.32	.37
33	.42	.70	.70	34	.44	.74	.82
35	.45	.86	.93	36	.34	.77	.81
38	.59	.72	.72	40	.34	.83	.90
41	.20	.62	.71	42	.32	.81	.90
43	.34	.56	.63	44	.54	.69	.61
45	.72	.13	.18	46	.53	.90	.88
47	.24	.63	.68	50	.46	.84	.90
51	.18	.63	.70	52	.77	.23	.10
53	.54	.77	.50	54	-.04	.41	.44
55	.58	.75	.67	56	.32	.47	.45
57	.66	.77	.79	58	.43	.88	.94
60	.54	.50	.53	61	.60	.17	.07
62	.41	.85	.91	63	.67	.72	.53
64	.68	.48	.49	65	.60	.50	.86
66	.58	.82	.67	67	.48	.52	.83
68	.48	.61	.69	69	.45	.83	.88
70	.41	.83	.87	72	.67	.68	.62
73	.35	.69	.79	74	.66	.16	.09
76	.36	.74	.78	77	.41	.76	.85
78	.46	.83	.89	79	.42	.68	.67
80	.47	.57	.51	81	.54	.65	.68
82	.38	.74	.73	83	.25	.72	.67
84	.48	.28	.36	85	.30	.71	.65
86	.36	.45	.52	87	.39	.77	.81
88	.44	.73	.77	89	.41	.54	.60
90	.47	.75	.81	91	.20	.74	.85
92	.42	.65	.66	93	.56	.69	.60
94	.36	.52	.55	95	.38	.54	.50
96	.32	.62	.73	97	.45	.78	.81
98	.40	.62	.66	99	.49	.83	.90
100	.50	.83	.87	101	.52	.70	.76
102	.56	.61	.65	103	.74	.23	.16
104	.66	.45	.34	105	.39	.65	.33
106	.50	.76	.73	107	.50	.81	.50
108	.46	.86	.93	109	.40	.71	.76
110	.54	.86	.83	111	.40	.70	.76
112	.58	.56	.66	113	.71	.14	.01
114	.32	.20	.34				

und R_3 . Umgekehrt haben von den Vpn, deren SW hoch mit der OW korreliert, einige ausgesagt, sie hätten sich nach der %-Differenz entschieden. Es besteht somit ein Unterschied zwischen dem, was bei den Vpn für die SW (und die Entscheidung) wirklich ausschlaggebend ist, und dem, was die Vpn introspektiv glauben, daß es für sie ausschlaggebend sei.

Auf Grund der Ergebnisse von Tab. 1 wurden die Vpn in zwei Gruppen unterteilt. Die erste Gruppe besteht aus 15 Vpn, bei denen R_1 größer als R_2 und R_3 ist, oder aber R_2 und R_3 sehr gering sind. Die restlichen 88 Vpn bilden die zweite Gruppe. Da bei der zweiten Gruppe die Variation der SW sehr gut durch die %-Differenz bzw. durch das Gesamtverhältnis beschrieben werden kann, wird diese Gruppe im folgenden auch "deskriptive Gruppe" genannt. Bei den 15 Vpn der ersten Gruppe ist die SW eher normativ; diese Gruppe wird daher mit "normative Gruppe" bezeichnet werden. Die "Gesamtgruppe" besteht bei der Betrachtung der SW aus diesen beiden Gruppen, bei der Betrachtung von $f(x)$ außerdem noch aus den 16 Vpn, für die keine Angaben über die SW zur Verfügung stand.

In Abb. 4 bis Abb. 16 ist die Gesamtgruppe (119 bzw. 103 Vpn) durch schwarze Punkte und durchgezogene Linien, die deskriptive Gruppe (88 Vpn) durch Kreuze und gestrichelte Linien und die normative Gruppe (15 Vpn) durch Kreise und gepunktete Linien dargestellt. Die obere Abbildung ("a") zeigt jeweils den Zusammenhang zwischen OW (Abszisse) und der relativen Häufigkeit der "NEIN"-Antworten¹⁾ (Ordinate). Die untere Abbildung ("b") zeigt den Zusammenhang zwischen OW (Abszisse) und SW (Ordinate).

1) Die relative Häufigkeit der "NEIN"-Antworten wird im Text und in den Abb. mit " $f(x)$ " abgekürzt.

Die Tabelle darunter enthält die wichtigsten Angaben über die Aufgaben der jeweiligen Reihe.

Variation der Stichprobengröße (V_1): Bei Reihe I (Abb.4), II (Abb.5) und IV (Abb.6(2)) zeigt sich, daß die Häufigkeit der "NEIN"-Urteile unabhängig von der OW ist. Bei I und II steigt $f(x)$ der normativen Gruppe mit fallender OW signifikant an ($\alpha < 0.01$). Für die SW - OW Beziehung ergeben sich ebenfalls bei der Gesamtgruppe und der deskriptiven Gruppe Abszissenparallelen, bei der normativen Gruppe dagegen nimmt die SW mit der OW ab.

Bei Reihe III (Abb.6(1)) steigen die Kurven zwar signifikant an ($\alpha < 0.01$), aber dieser Anstieg läßt sich noch durch den geringen Stichprobenumfang erklären¹⁾. Diese Erklärung muß dagegen für Reihe V (Abb.7) entfallen. Hier steigt $f(x)$ für die Gesamtgruppe mit fallender OW signifikant an ($\alpha < 0.01$). Dies Ergebnis steht zu den Ergebnissen der Reihen I, II und IV in Widerspruch. Es ist dem Autor nicht gelungen eine Erklärung dafür zu finden.

Variation des Stichprobenverhältnisses (V_2): Für Reihe VI (Abb.8) und VIII (Abb.10) ergeben sich auch hier bei der Gesamtgruppe und der deskriptiven Gruppe Abszissenparallelen für $f(x)$ und die SW. Nur für die 15 normativen Vpn ergibt sich ein Zusammenhang zwischen OW und $f(x)$ ($\alpha < 0.01$) bzw. zwischen SW und OW. Bei Reihe VII zeigt sich für die normative Gruppe kein Zusammenhang. Bei der Gesamtgruppe nimmt $f(x)$ hier sogar mit der OW ab ($\alpha < 0.05$). Dies wäre zwar durch die größer werdenden Unterschiede zwischen den Stichprobenumfängen erklärbar¹⁾; da aber bei Reihe VI und VIII ein solcher Effekt nicht aufgetreten ist, liegt ein Aufgabenartefakt näher.

1) Viele Vpn gaben an, daß für "NEIN"-Entscheidungen die Stichproben nicht allzu klein und die Unterschiede zwischen den Stichprobengrößen nicht allzu groß sein dürfen.

Reihe I

Abb.4a

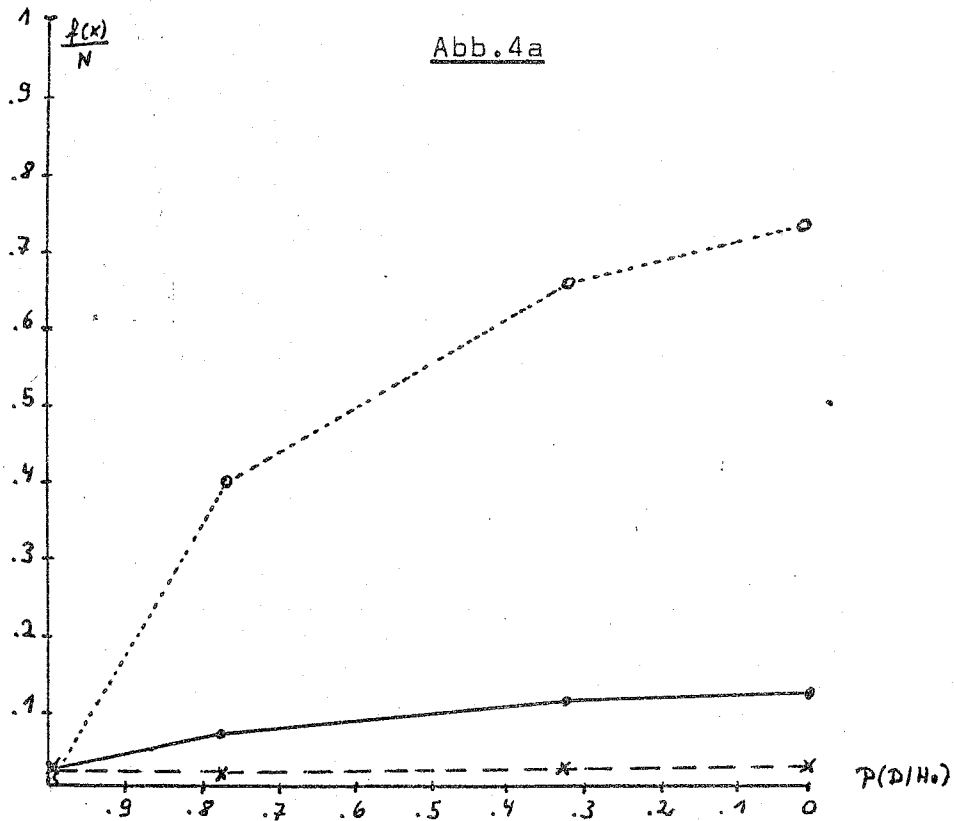
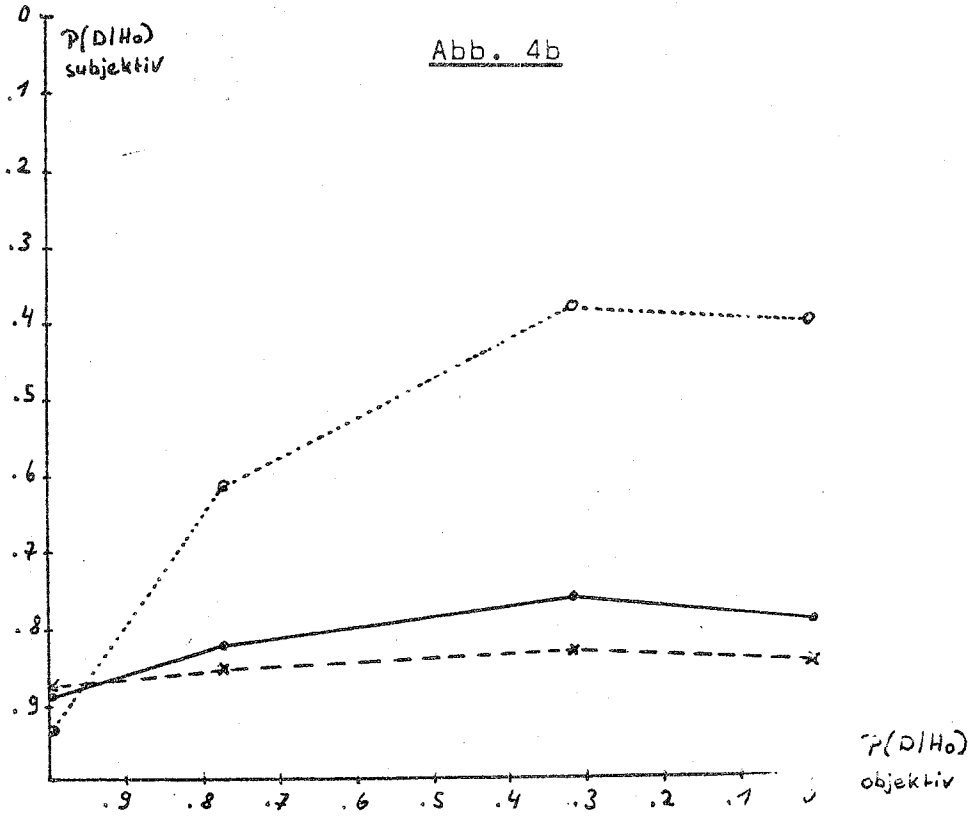


Abb. 4b



Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
n_a	n_b	I (in S_a)	II (in S_b)	I-II		
200	200	50,5	50,0	0,5	1,0202	<0,0001
2000	2000	"	"	"	"	0,0810
20000	20000	"	"	"	"	0,9801
200000	200000	"	"	"	"	9,9803

Reihe II

Abb. 5a

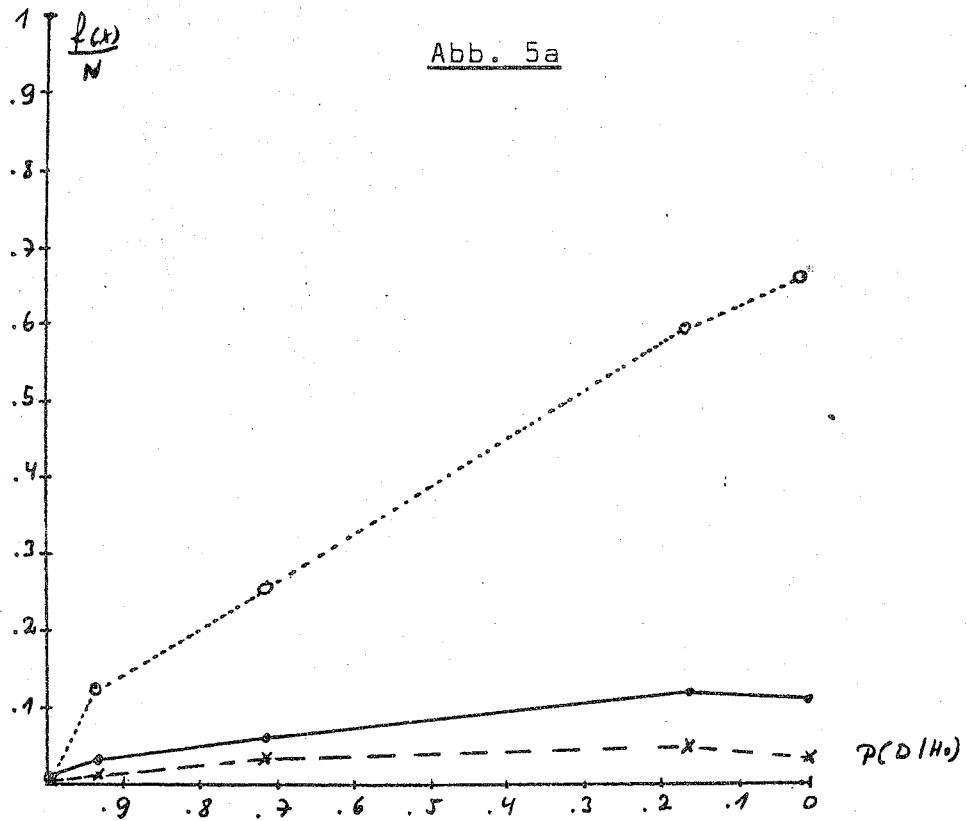
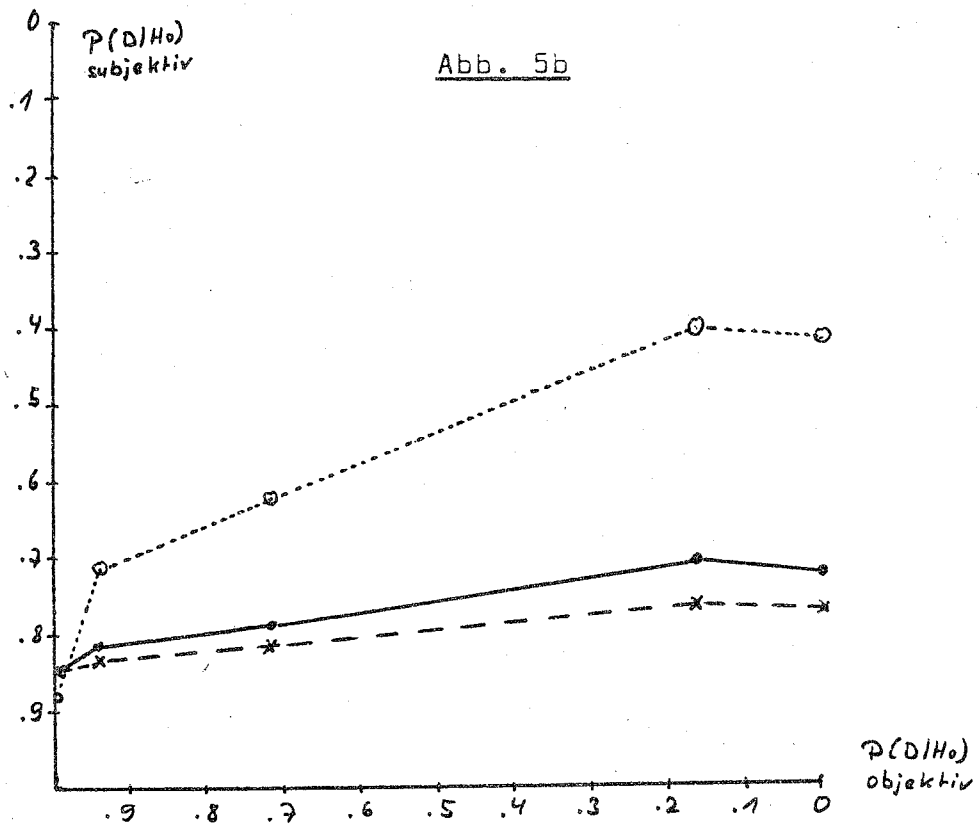
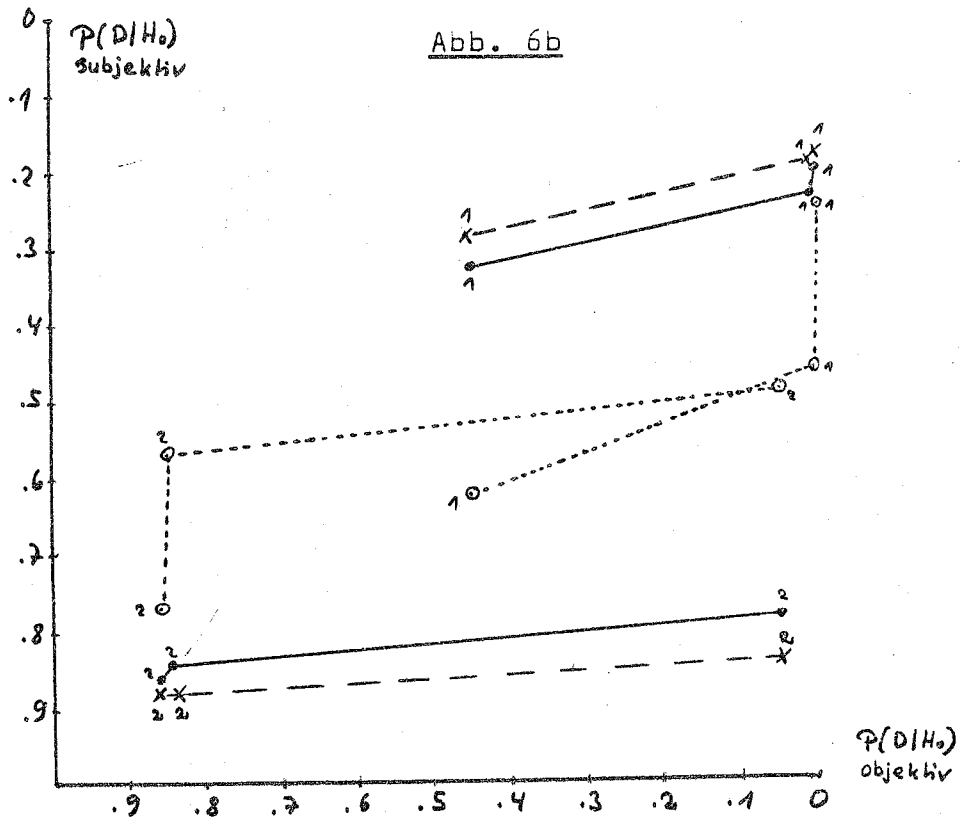
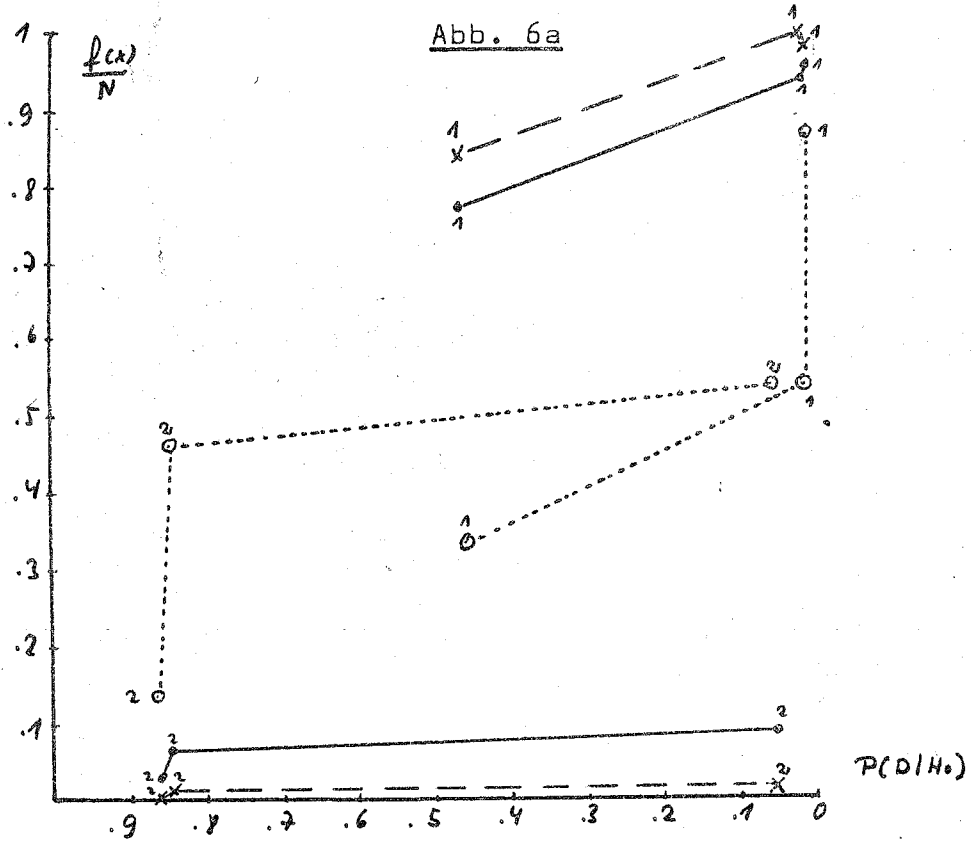


Abb. 5b



Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
n _a	n _b	I (in S _a)	II (in S _b)	I-II		
100	100	60,0	59,0	1,0	1,0424	<0,0001
1000	1000	"	"	"	"	0,1681
10000	10000	"	"	"	"	2,0336
50000	50000	"	"	"	"	10,3331
2. Wert (100)	100	26,0	25,0	1,0	1,0541	0,0060



	Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
	n _a	n _b	I (in S _a)	II (in S _b)	I-II		
1)	15	15	60,0	40,0	20,0	2,2500	0,5333
	150	150	"	"	"	"	11,2133
	1500	1500	"	"	"	"	19,2011
2)	3000	3000	66,9	66,7	0,2	1,0106	0,0270
	30000	30000	"	"	"	"	0,3577
	300000	300000	"	"	"	"	3,6710

Reihe V

Abb. 7a

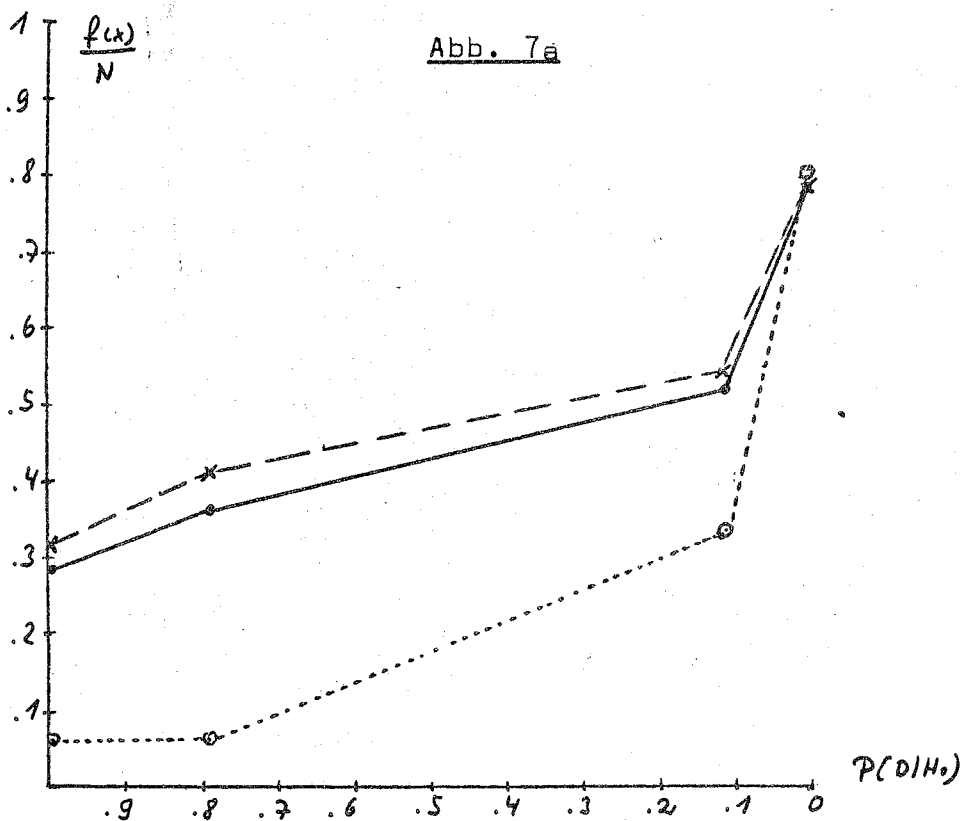
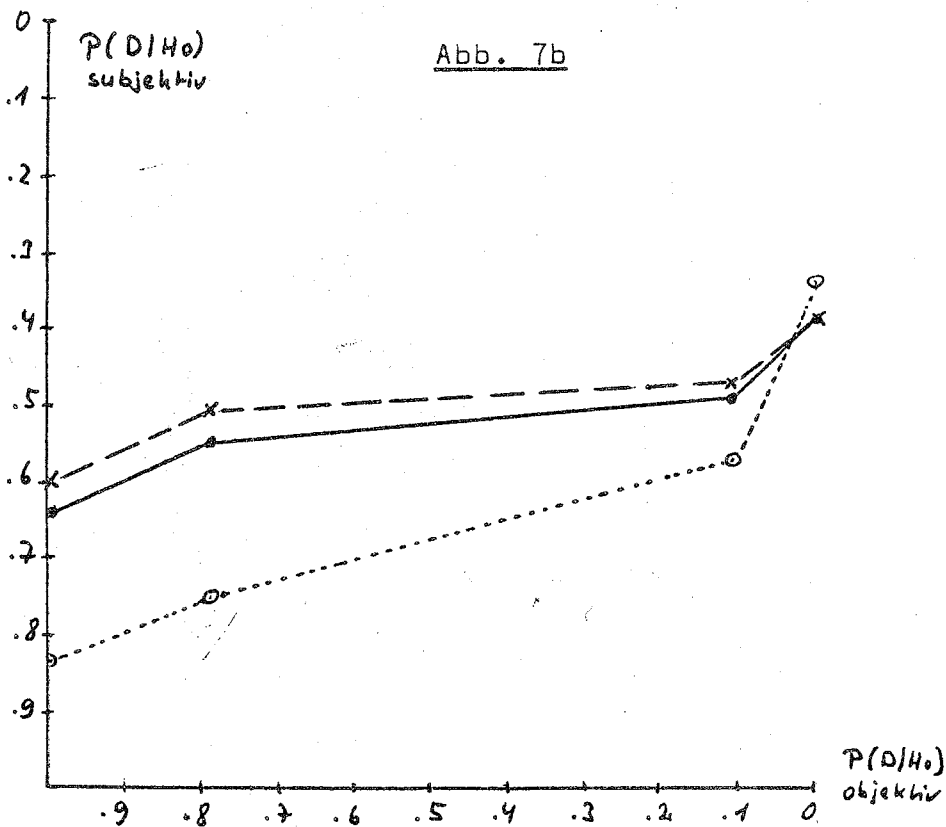


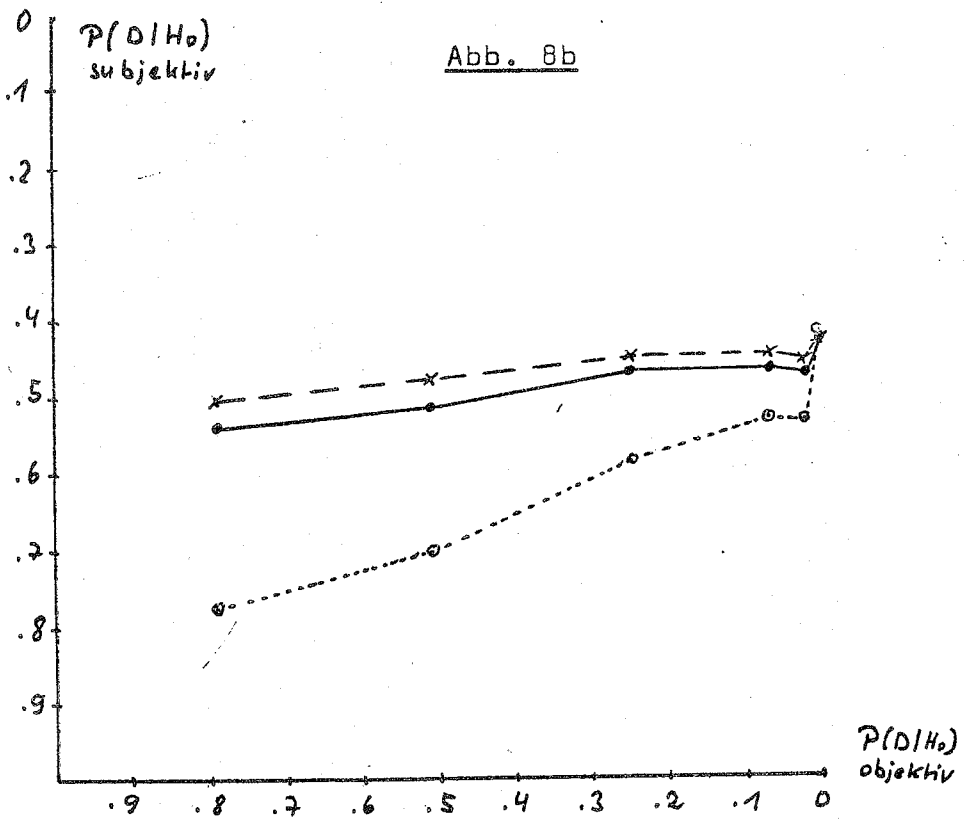
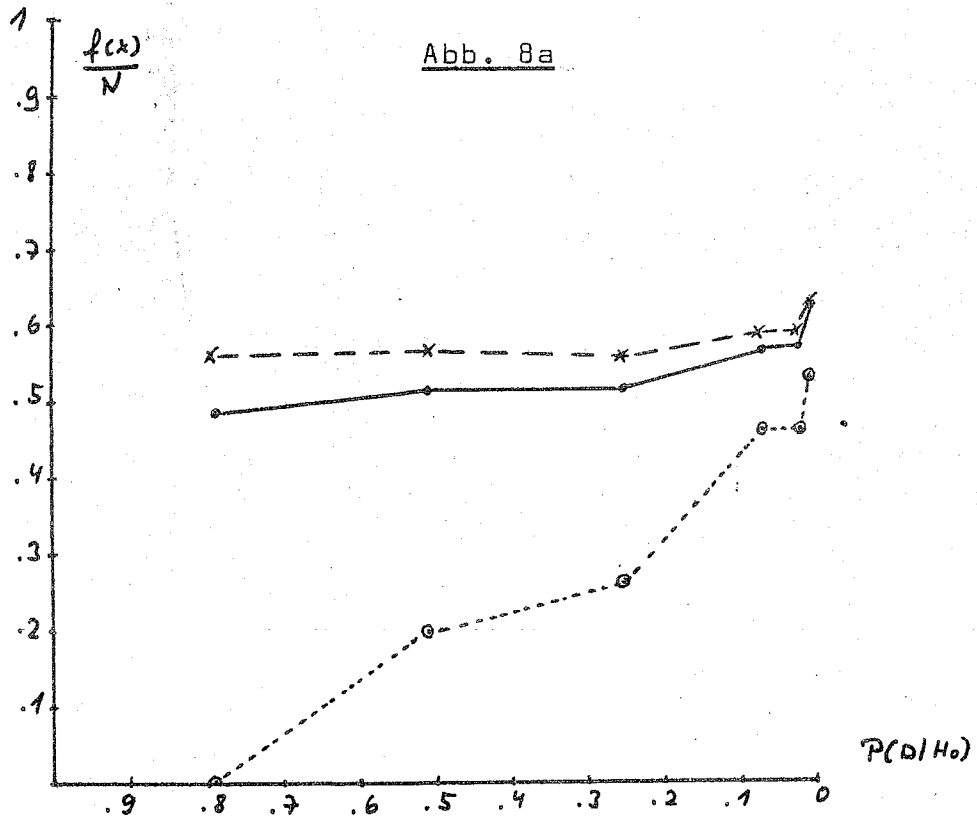
Abb. 7b



Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
n_a	n_b	I (in S_a)	II (in S_b)	I-II		
15	15	40,0	33,3	6,7	1,3333	<0,0001
30	30	"	"	"	"	0,0718
300	300	"	"	"	"	2,5909
3000	3000	"	"	"	"	28,4217

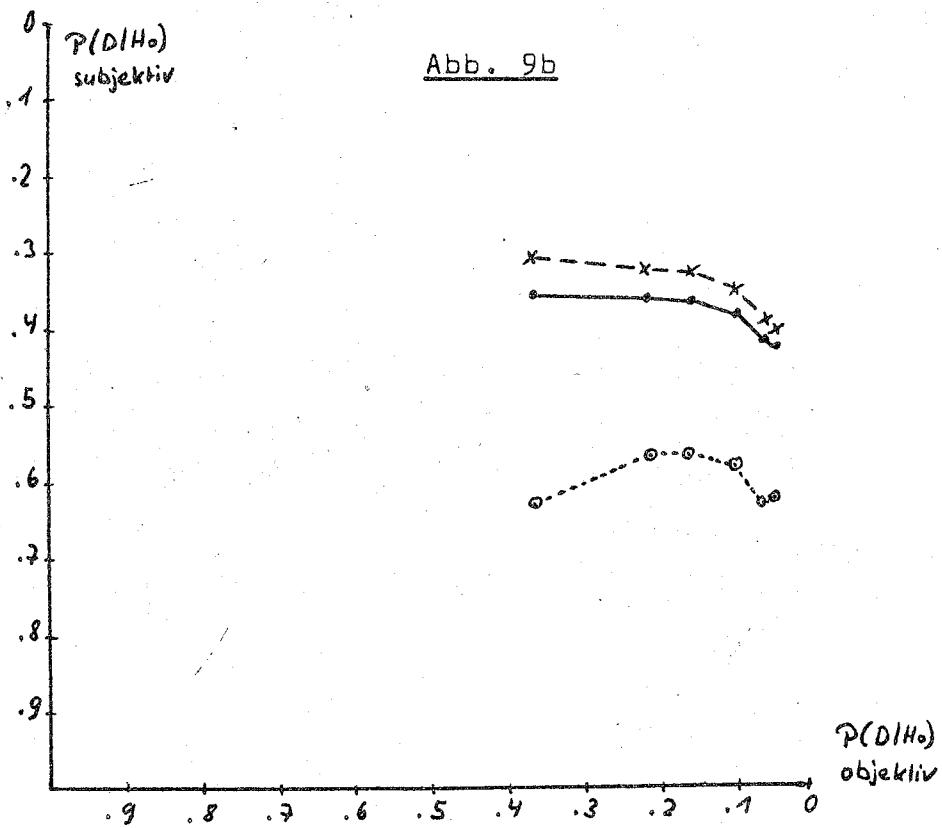
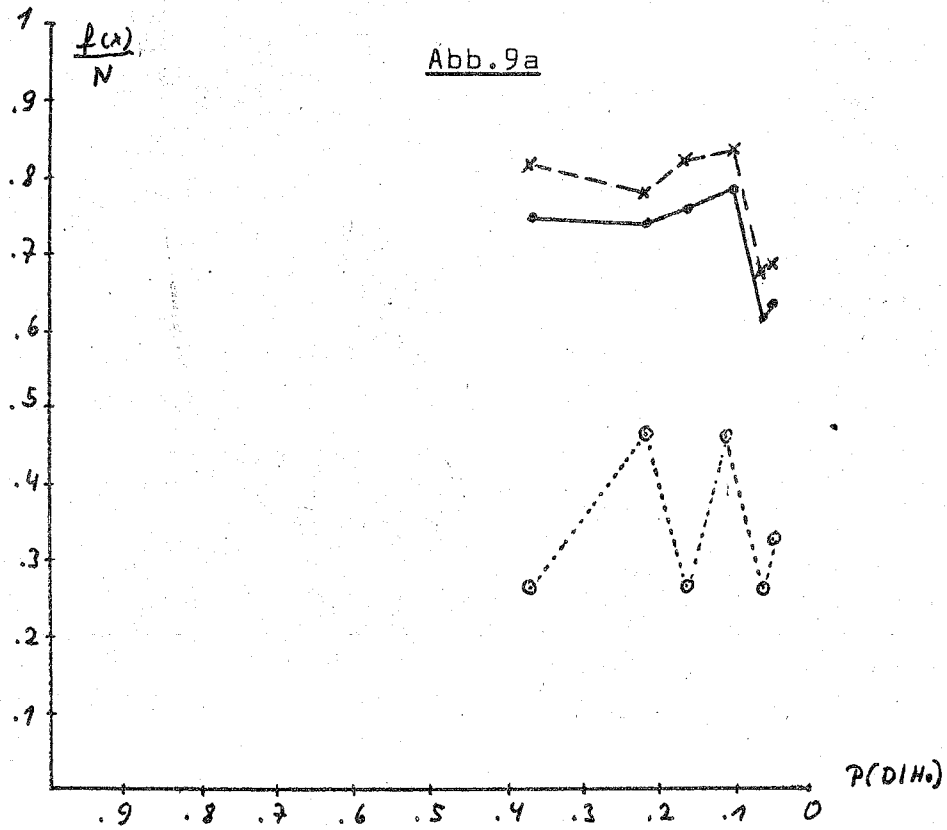
Reihe VI

Abb. 8a

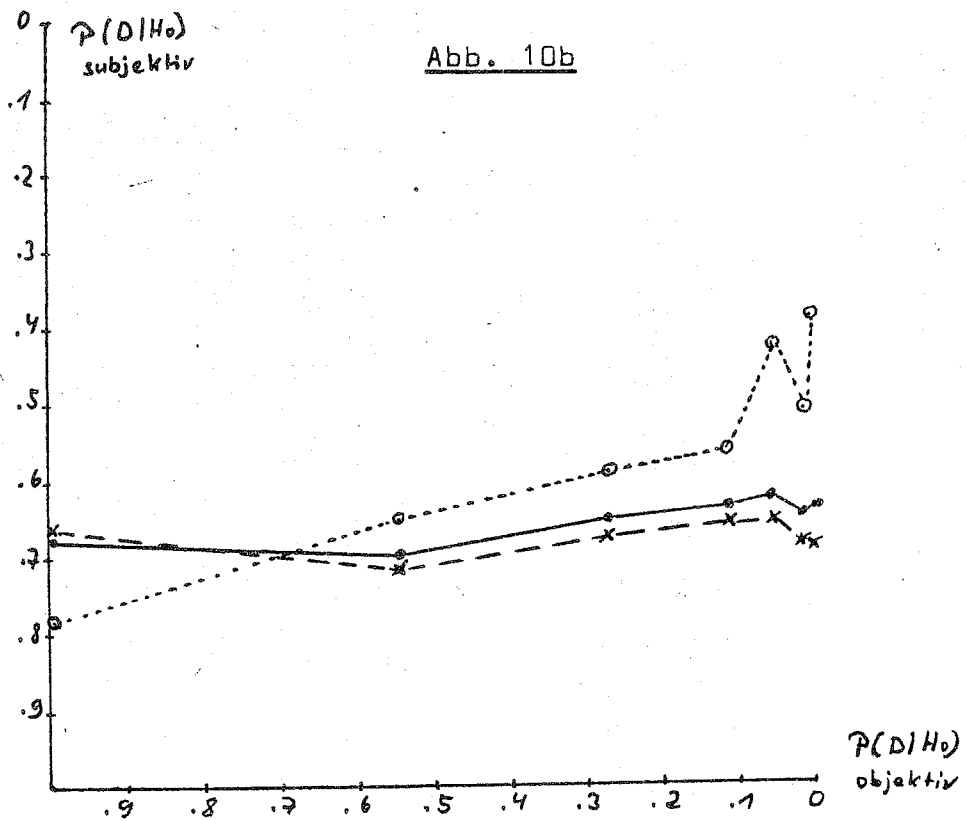
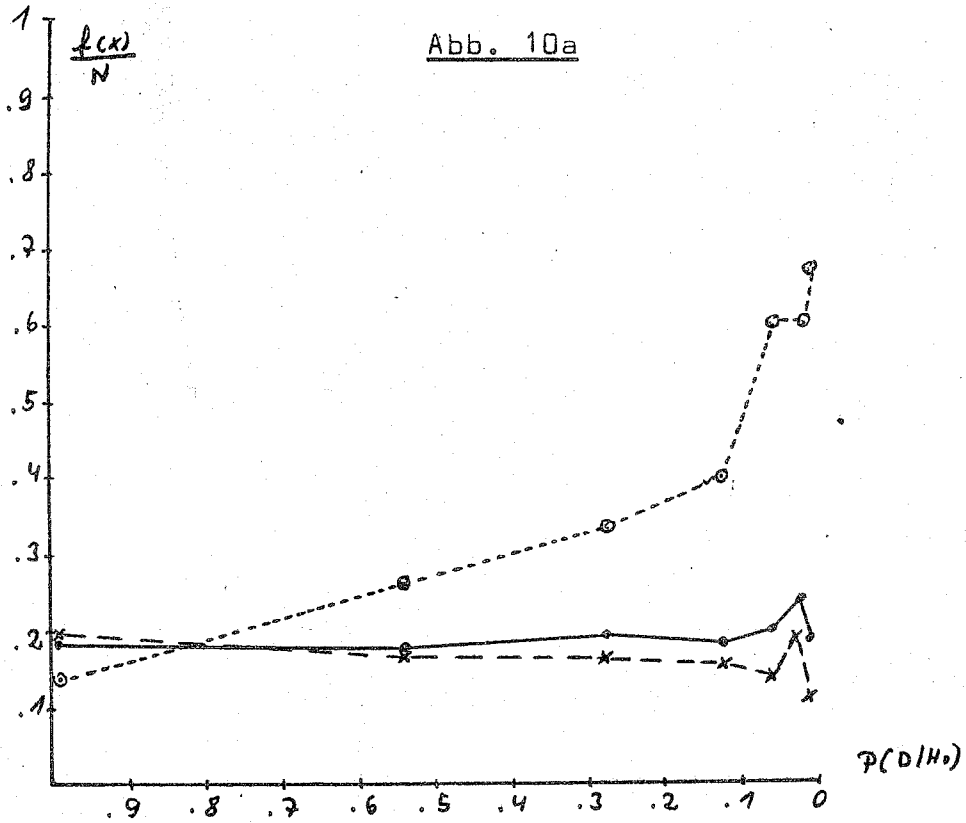


Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
n_a	n_b	I (in S_a)	II (in S_b)	i-II		
14	1000	57,1	50,0	7,1	1,3333	0,0685
35	"	"	"	"	"	0,4340
84	"	"	"	"	"	1,3088
210	"	"	"	"	"	3,2640
420	"	"	"	"	"	5,7642
840	"	"	"	"	"	9,0718

Reihe VII



Stichprobengröße		defekte Birnen (%)			Gesamt- verhältnis	Chi ²
n _a	n _b	I (in S _a)	II (in S _b)	I-II		
18	40	66,7	50,0	16,7	2,0000	0,8018
36	"	"	"	"	"	1,5295
54	"	"	"	"	"	2,0035
108	"	"	"	"	"	2,7750
270	"	"	"	"	"	3,5306
540	"	"	"	"	"	3,8709



Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
n _a	n _b	I (in S _a)	II (in S _b)	I-II		
21	6000	52,4	50,0	2,4	1,1	<0,0001
210	"	"	"	"	"	0,3700
630	"	"	"	"	"	1,1994
1470	"	"	"	"	"	2,5832
2520	"	"	"	"	"	3,9302
6300	"	"	"	"	"	6,8778
21000	"	"	"	"	"	10,5014

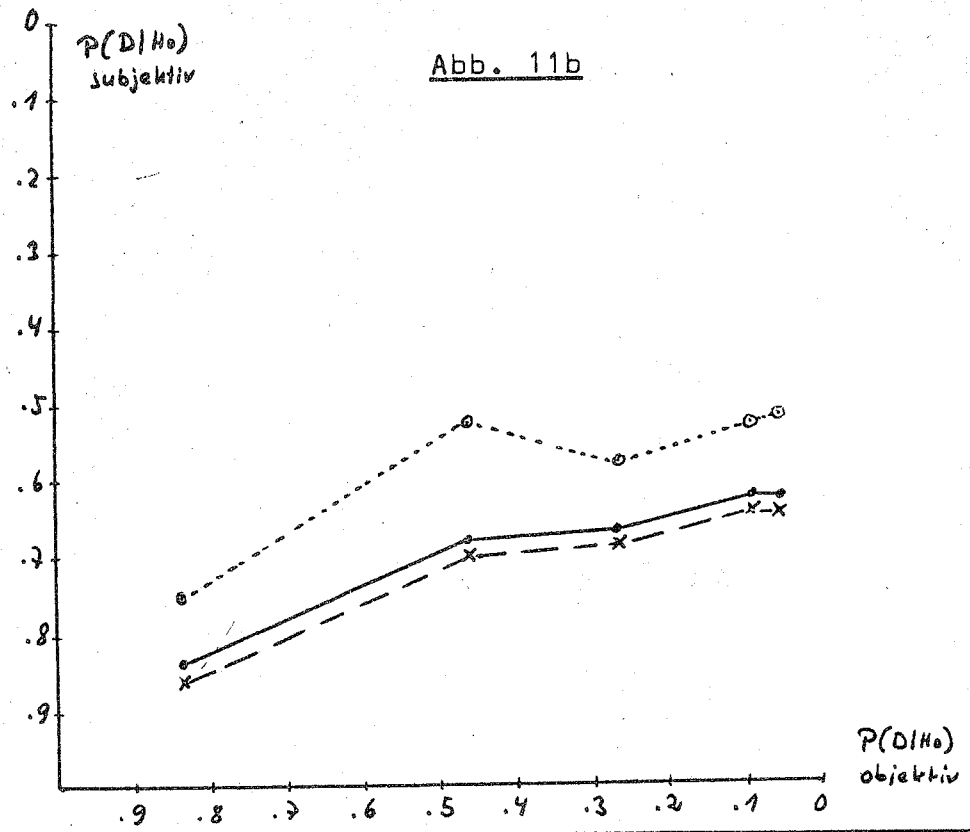
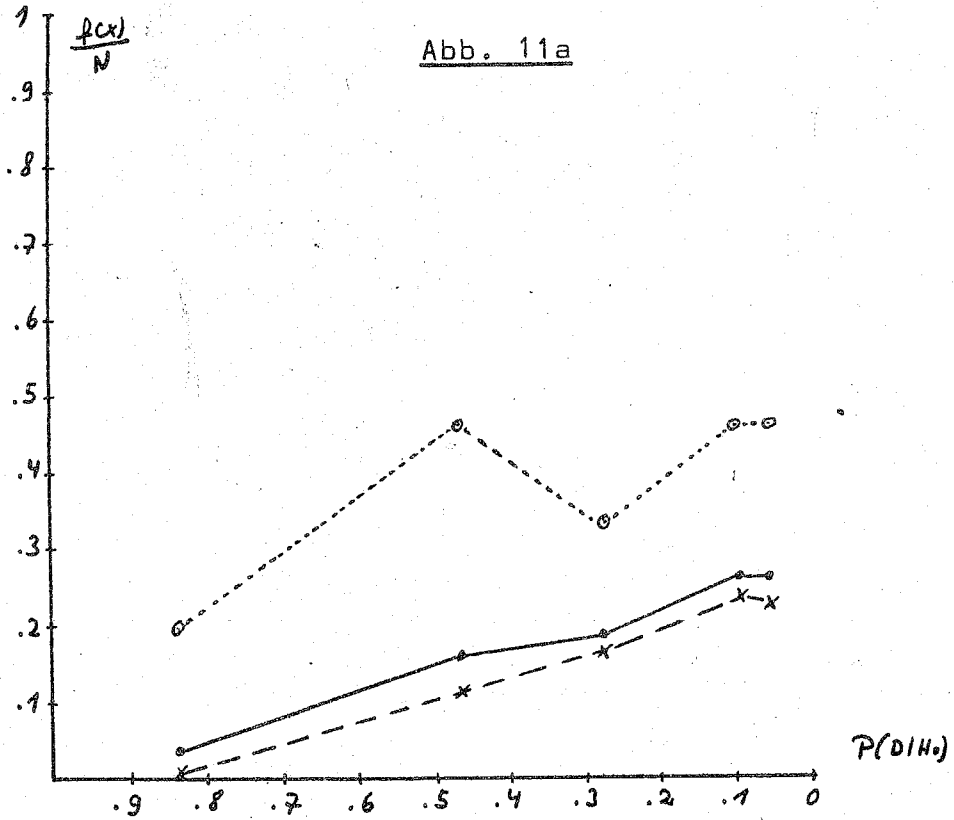
Bei Reihe IX und X (ohne Abb.) variiert das Stichprobenverhältnis, ohne daß sich dabei die OW wesentlich verändert. In keiner der drei Gruppen weicht $f(x)$ von einer Gleichverteilung über die Aufgaben ab.

Variation des Verhältnisses in S_a : Aus Reihe XI (Abb.11) und XII (Abb.12) ergibt sich, daß die Variation des Verhältnisses in S_a subjektiv relevant ist. Dies ist zwar nicht im Sinne der ursprünglichen Hypothese (nach der nur V_4 subjektiv relevant sein sollte), aber nach den bisherigen Ergebnissen hat sich gezeigt, daß besonders die %-Differenz subjektiv relevant ist - und genau diese %-Differenz variiert gemeinsam mit V_3 .

Variation des Gesamtverhältnisses (V_4): Die Variation des Gesamtverhältnisses ist - wie erwartet - subjektiv relevant. In Reihe XIII (Abb.14) und XIV (Abb.14) ist der Zusammenhang zwischen $f(x)$ und OW signifikant ($\alpha < 0.01$). Die SW ist fast identisch mit der OW.

Paradox-Kurve: Bei Reihe XV (Abb.15) zeigt sich am deutlichsten, daß ein Zusammenhang zwischen OW und SW bzw. zwischen OW und $f(x)$ ein Artefakt des Versuchsmaterials ist. Bei geeigneter Auswahl des Materials läßt sich jeder Zusammenhang signifikant nachweisen (-und damit keiner).

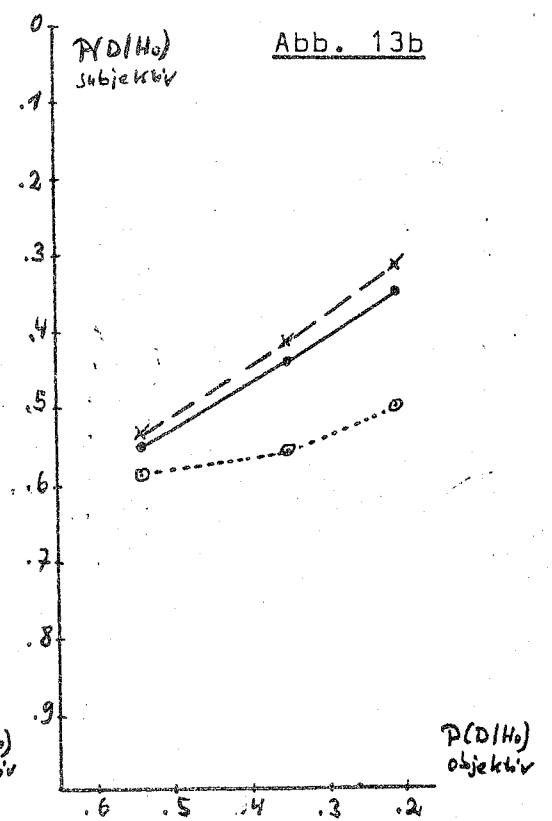
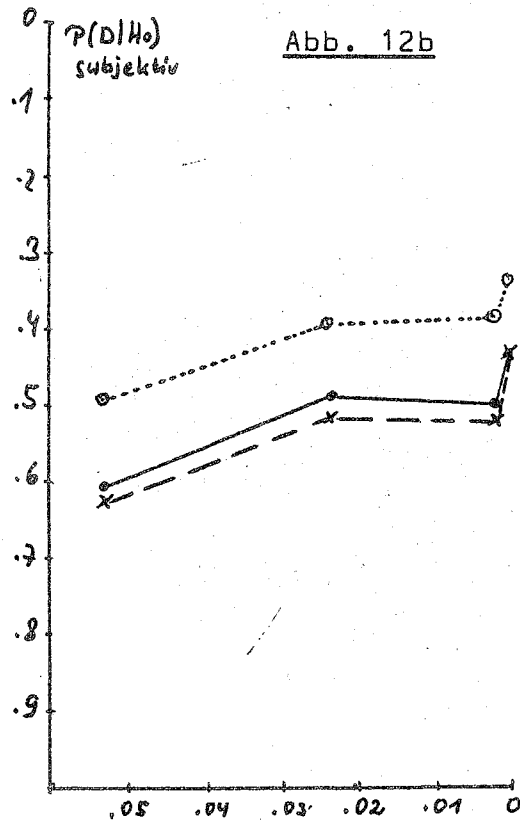
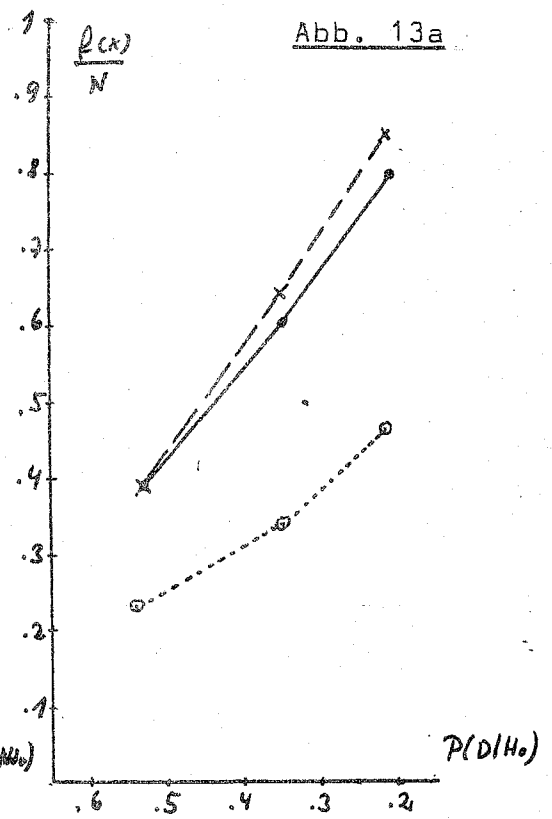
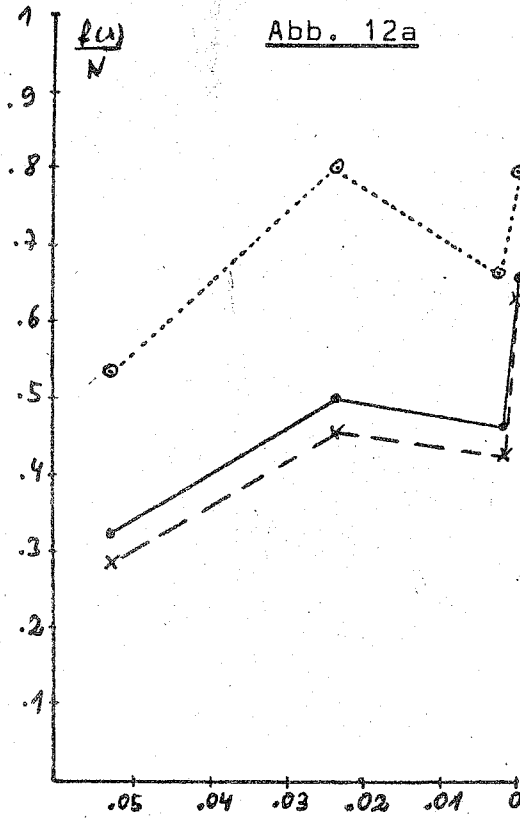
%-Differenz gegen Gesamtverhältnis: Aus Reihe XI und XII ergab sich, daß $f(x)$ mit zunehmender %-Differenz signifikant anwächst, auch wenn das Gesamtverhältnis konstant bleibt. Da bei der Konstruktion der Aufgaben die Variable "%-Differenz" nicht in Betracht gezogen wurde, ist keine Aufgabenreihe für diese Variable erstellt worden. Aus dem Aufgabenmaterial wurden erst nachträglich Aufgaben mit gleicher %-Differenz herausgesucht und zu Reihe XVI zusammengestellt. Es ergibt sich aber nur ein Zusammenhang zwischen $f(x)$ und OW (Abb.16), der aber auf den 1. Wert



Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
n _a	n _b	I (in S _a)	II (in S _b)	I-II		
1000	1000	1,2	1,0	0,2	1,2024	0,0460
"	"	5,0	4,2	0,8	1,2005	0,5583
"	"	9,7	8,2	1,5	1,2026	1,2026
"	"	25,1	21,8	3,3	1,2021	2,8522
"	"	37,2	33,0	4,2	1,2027	3,6900

Reihe XII

Reihe XIII



	Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
	n _a	n _b	I (in S _a)	II (in S _b)	I-II		
Zu Abb. 12	1000	1000	6,2	4,2	2,0	1,5077	3,6616
	"	"	8,5	5,8	2,7	1,5088	5,0913
	"	"	15,7	11,0	4,7	1,5068	9,1461
	"	"	29,1	21,4	7,7	1,5075	15,3011
Zu Abb. 13	100	100	30,0	25,0	5,0	1,2857	0,4013
	"	"	32,0	"	7,0	1,4118	0,8833
	"	"	34,0	"	9,0	1,5455	1,5386

Reihe XIV

Abb. 14a

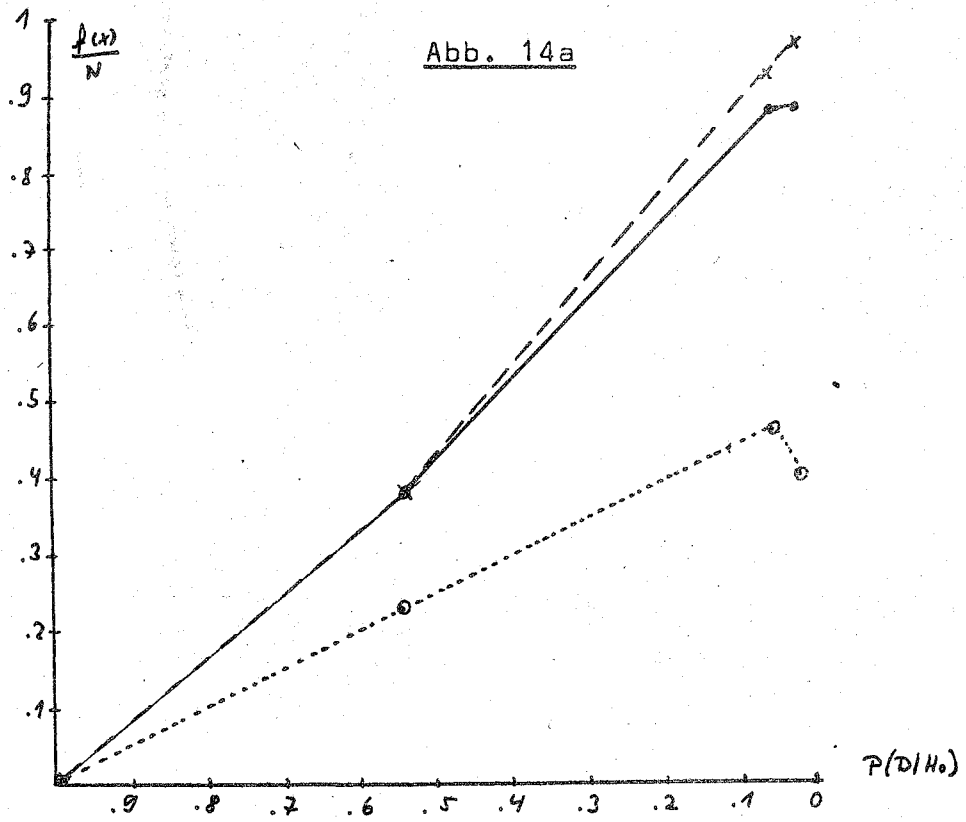
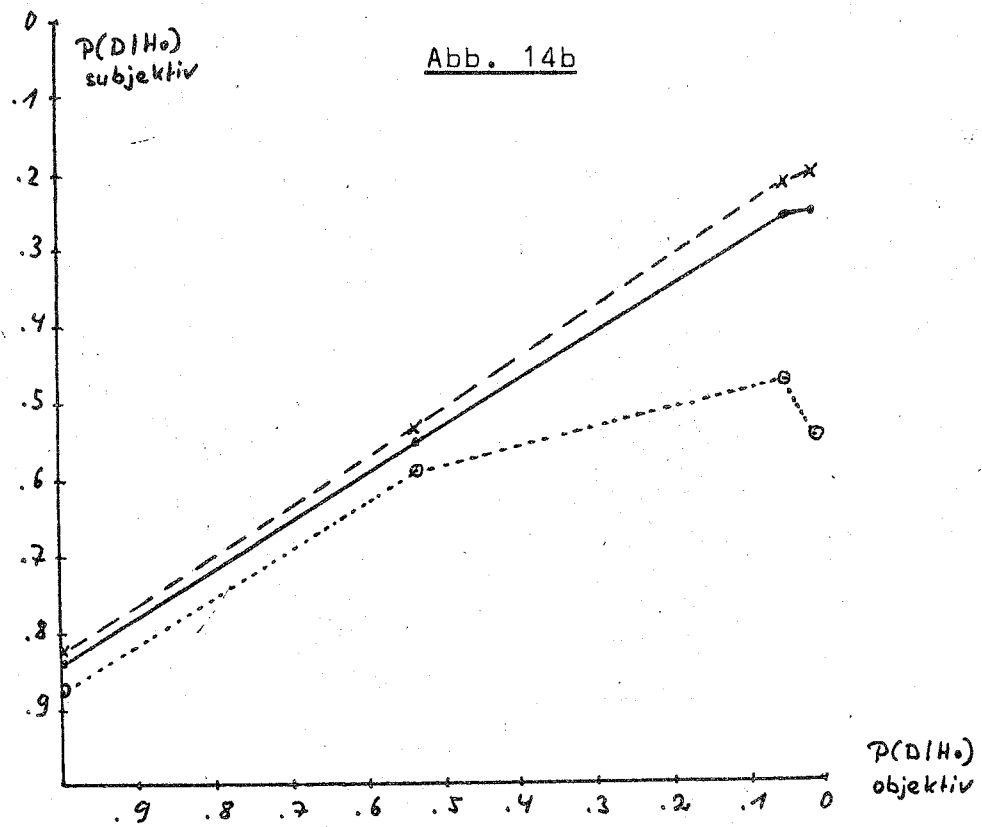


Abb. 14b



Stichprobengröße		defekte Birnen (%)			Gesamt- verhältnis	Chi ²
n _a	n _b	I (in S _a)	II (in S _b)	I-II		
100	100	30,0	29,0	11,0	1,0493	<0,0001
"	"	"	25,0	55,0	1,2857	0,4013
"	"	"	17,0	13,0	2,0924	4,0050
"	"	"	15,0	15,0	2,4286	5,6201

Reihe XV

Abb. 15a

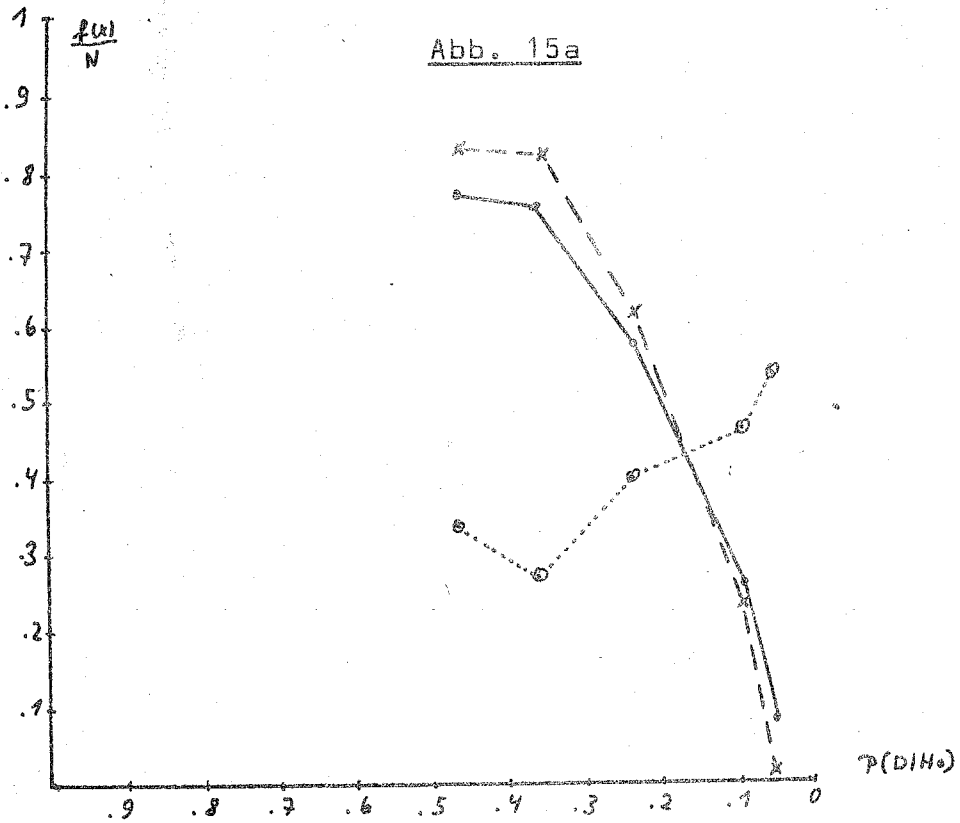
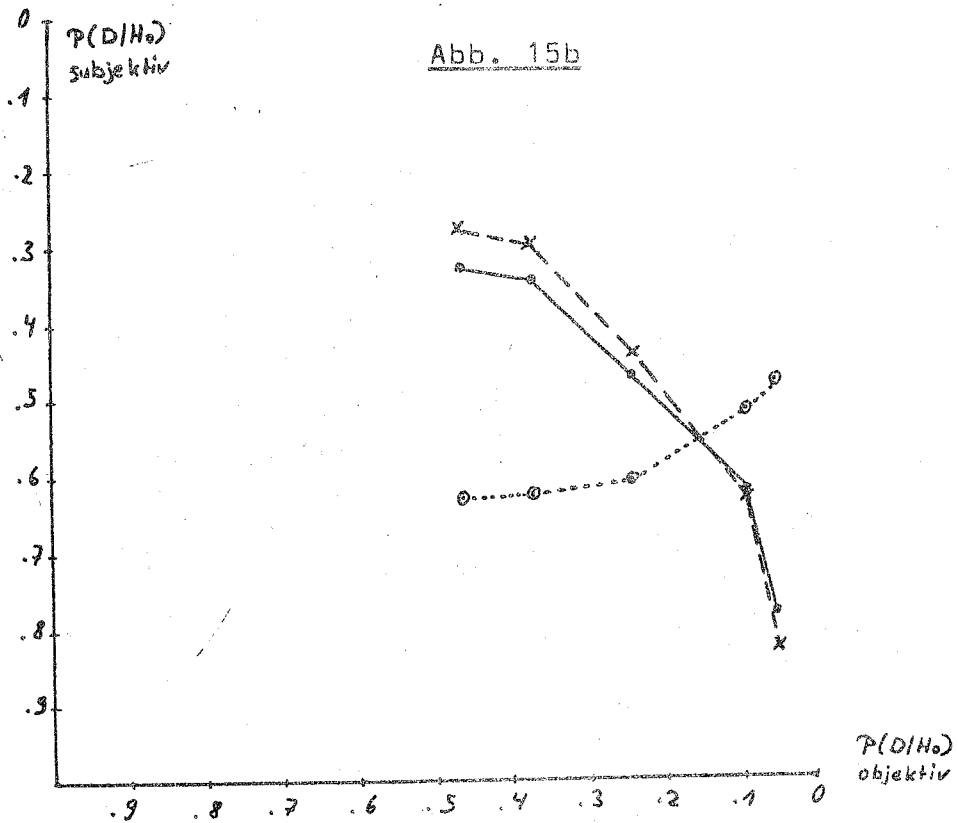


Abb. 15b



Stichprobengröße		defekte Birnen (%)			Gesamtverhältnis	Chi ²
n _a	n _b	I (in S _a)	II (in S _b)	I-II		
15	15	60,0	40,0	20,0	2,2500	0,5333
18	40	66,7	50,0	16,7	2,0000	0,8018
100	100	28,0	20,0	8,0	1,5556	1,3432
1000	1000	25,1	21,8	3,3	1,2021	2,8522
300000	300000	66,9	66,7	0,2	1,0106	3,6800

Reihe XVI

Abb. 16a

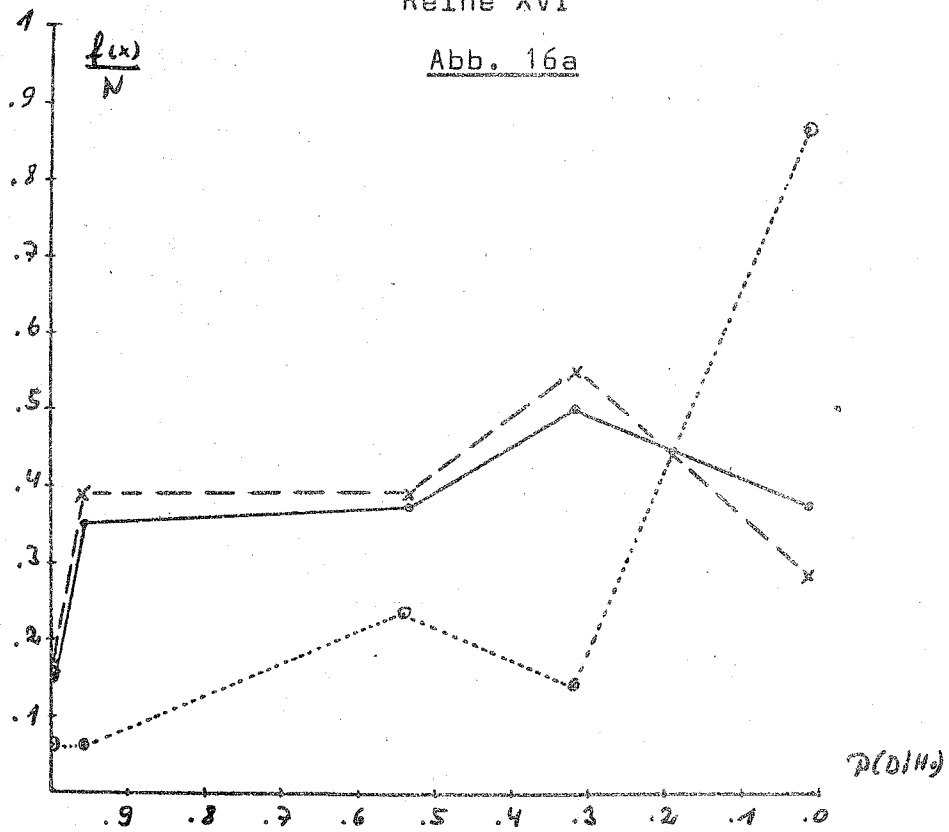
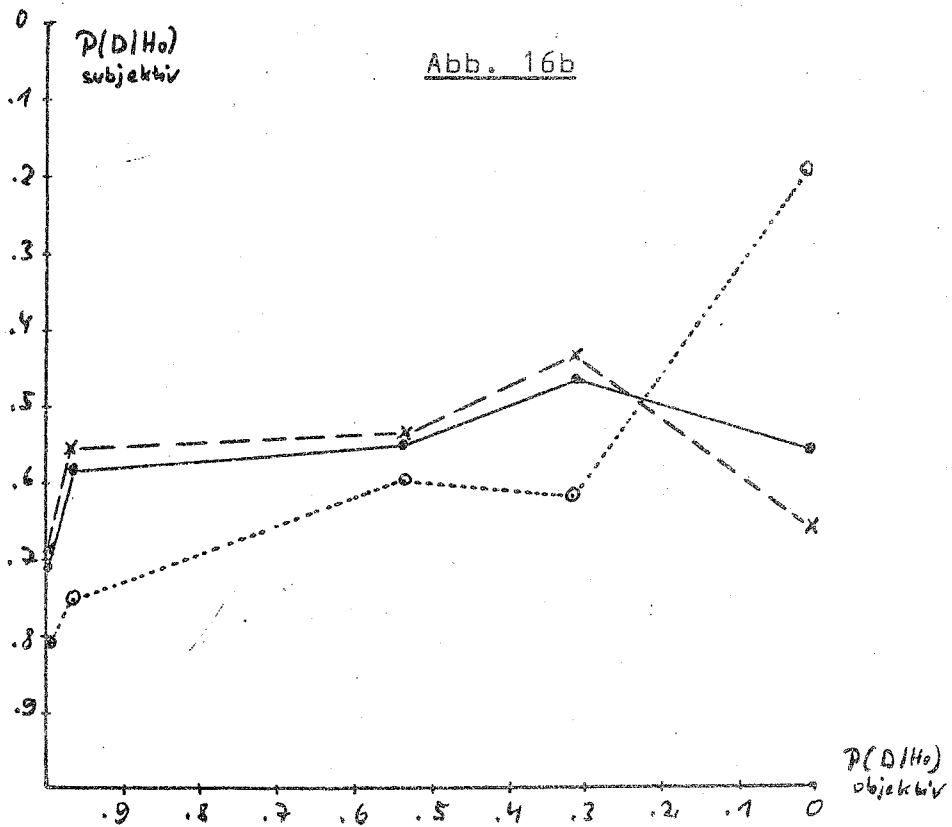


Abb. 16b



Stichprobengröße		defekte Birnen (%)			Gesamt- verhältnis	Chi ²
n _a	n _b	I (in S _a)	II (in S _b)	I-II		
20	20	40,0	35,0	5,0	1,2381	<0,0001
20	25	45,0	40,0	"	1,2273	0,0013
100	100	30,0	25,0	"	1,2857	0,4013
100	100	11,0	6,0	"	1,9363	1,0286
200000	400000	55,0	50,0	"	1,2222	1334,6156

zurückgeht, was wiederum auf die sehr kleine Stichprobe in dieser Aufgabe zurückzuführen ist. Ein Zusammenhang zwischen $f(x)$ und dem Gesamtverhältnis zeigt sich nicht.

An Hand der Aufgabenreihen zeigt sich, daß die Häufigkeit der "NEIN"-Entscheidungen hauptsächlich von der %-Differenz abhängt. Diese Abhängigkeit soll nun am gesamten Aufgabenmaterial untersucht werden. Dazu wurden folgende Korrelationen berechnet:

- R_1 : Korrelation zwischen $f(x)$ und $1-Ow$
- R_2 : Korrelation zwischen $f(x)$ und dem Gesamtverhältnis
- R_3 : Korrelation zwischen $f(x)$ und der %-Differenz

Tab. 2 zeigt diese Korrelationen für die Gesamtgruppe und einzelne Teilgruppen:

<u>Tab. 2</u>			Gruppe
R_1	R_2	R_3	
.65	.92	.93	Gesamtgruppe (119 Vpn)
.56	.94	.96	deskriptive Gruppe (88 Vpn)
.87	.29	.22	normative Gruppe (15 Vpn)
.77	.84	.79	nach verbalen Aussagen wahrscheinlich normativ (26 Vpn)
.80	.77	.71	nach verbalen Aussagen sicher normativ (14 Vpn)

Es zeigt sich, daß für die Gesamtgruppe eine außerordentlich hohe Korrelation zwischen %-Differenz und $f(x)$ besteht. Die Korrelation zwischen $f(x)$ und dem Gesamtverhältnis ist genau so hoch. Dagegen ist die Korrelation zwischen der Ow und $f(x)$ geringer (0.01), wenn auch ebenfalls noch signifikant (0.01). Zwar korreliert bei den Aufgaben die %-Differenz und das Gesamtverhältnis mit der Ow (.46 bzw. .50), aber die Höhe von R_2 und R_3 ist nicht darauf zurückzuführen - die Partialkorrelationen unter Ausschaltung der Variation der Ow sind für R_2 und R_3 unverändert hoch.

Ebenso läßt sich aber auch die Korrelation zwischen der OW und $f(x)$ nicht durch die Korrelation zwischen der OW mit der %-Differenz einerseits und der %-Differenz mit $f(x)$ andererseits erklären: die Partialkorrelation ist hier genau so hoch wie R_1 .

Bei der deskriptiven Gruppe ist R_3 höher ($\alpha < 0.05$), R_1 ist niedriger (> 0.05) als bei der Gesamtgruppe (sofern man bei R_1 den Einfluß der %-Differenz auspartialisiert).

Bei der normativen Gruppe sind R_2 und R_3 nicht signifikant von 0 verschieden. R_1 ist bedeutend höher als bei der Gesamtgruppe ($\alpha < 0.01$).

Zum Vergleich wurden Vpn auf Grund ihrer verbalen Aussagen zu "normativen" Gruppen zusammengestellt. Die erste Gruppe (26 Vpn) enthält die Vpn aus deren Aussagen es wahrscheinlich ist, daß sie versuchten, normativ zu handeln. Die Aussagen der Vpn sind aber oft nicht sehr klar. Daher wurden aus der ersten Gruppe 14 Vpn ausgelesen, bei denen die Aussage sicher lautet, daß sie sich nach normativen Kriterien entschieden hatten. Es zeigt sich aber, daß bei diesen Gruppen keine wesentlichen Unterschiede zwischen R_1 , R_2 und R_3 auftreten. Die Entscheidungen dieser Vpn gehen genau so auf normative wie auf nicht normative Variablen zurück.

Es zeigt sich somit, daß die Untergliederung der Vpn in eine normative und eine deskriptive Gruppe auf Grund der Tab. 1 wesentlich besser ist, als auf Grund der verbalen Aussagen der Vpn.

Insgesamt läßt sich sagen, daß bei der Gesamtgruppe 87% der Variabilität der Entscheidungen durch die %-Differenz determiniert ist. Nur 42% der Variabilität ist dagegen durch die OW determiniert.

Zusammenfassung und Diskussion

Mit einem kurzen geschichtlichen Überblick über wesentliche Experimente zur Entscheidungstheorie wurde zu zeigen versucht, daß diesen Experimenten eine Annahme zugrunde liegt: die Abhängigkeit zwischen subjektiver und objektiver Wahrscheinlichkeit.

Es wurde argumentiert, daß diese Grundannahme sinnvoll (aber trivial) ist, sofern in den Entscheidungssituationen die OW recht einfach zu berechnen ist. Es ergibt sich aber die Frage, ob die Ergebnisse solcher Experimente auf Entscheidungssituationen übertragbar sind, bei denen die OW selbst eine komplizierte Funktion mehrerer Parameter ist.

Ein Experiment wurde an 119 Vpn durchgeführt, das SW und Entscheidungsverhalten im zwei-Stichproben-Vergleich für Alternativmerkmale untersucht. Dabei zeigten sich folgende Hauptergebnisse:

- 1) Man kann zwei Gruppen von Vpn unterscheiden. Eine Gruppe (ca. 15% der Vpn) verhält sich eher normativ, d.h. ihre SW und Entscheidungen sind durch die OW gut erklärbar. Die zweite Gruppe (ca. 85% der Vpn) verhält sich nicht normativ; ihr Verhalten ist durch andere Parameter besser zu beschreiben, als durch die OW.
- 2) Die Einteilung in diese zwei Gruppen läßt sich nicht auf Grund der introspektiven Aussagen der Vpn vornehmen. Es besteht ein Unterschied zwischen dem, wonach entschieden wird, und dem, wonach man glaubt, daß man sich entscheidet. Da die normative Gruppe nur aus einem sehr kleinen Teil der gesamten Vpn besteht, lassen sich für die Gesamtgruppe noch folgende Aussagen machen:
- 3) Für die subjektive Wahrscheinlichkeit und für Entscheidungen ist die Stichprobengröße irrelevant. (Bei allzu kleinen Stichproben werden Unterschiede allerdings

als wahrscheinlicher angesehen).

4) Das Verhältnis der Stichproben zueinander ist subjektiv ebenfalls irrelevant.

5) Subjektiv relevant sind die Verhältnisse in den Stichproben und das Gesamtverhältnis. Die Entscheidungen und die SW lassen sich sehr gut durch die Differenz, die ein Alternativmerkmal prozentual an der einen und an der anderen Stichprobe hat, beschreiben. 87% der Variabilität der Entscheidungen ist durch diese Variable determiniert.

6) Eine Abhängigkeit zwischen SW und OW (und zwischen OW und Entscheidungen) ist ein Artefakt der experimentellen Bedingungen.

Ein Artefakt ist an sich nichts Negatives, sofern man sich des Artefaktes bewußt ist. Es sei daher nochmals betont, daß die berichteten Experimente auch unter dem Gesichtspunkt der Unabhängigkeit von OW und SW durchaus sinnvoll waren. Allerdings wird man für die weitere Forschung die Ergebnisse dieser Experimente nicht verallgemeinern dürfen. Wenn man die Abhängigkeit zwischen OW und SW nicht mehr postuliert, werden damit die Möglichkeiten, Entscheidungsverhalten zu beschreiben und vorherzusagen, nicht beschränkt, sondern eher erweitert.

Der Versuch, subjektive Wahrscheinlichkeiten und Entscheidungen durch andere Variable als die objektive Wahrscheinlichkeit zu beschreiben, scheint für einen kleinen Teilbereich innerhalb der Vielfalt von Entscheidungsarten gelungen zu sein. Ob es möglich sein wird, aus den Ergebnissen dieser Arbeit auch für weitere Bereiche Nutzen zu ziehen, wird sich erst zeigen müssen.

Comments on "Über die Unabhängigkeit zwischen subjektiven
und objektiven Wahrscheinlichkeiten" by J. Kriz

by Anatol RAPOPORT

Two philosophical questions immediately come to mind in connection with Kriz's experiments and his conclusions.

1. What does it mean to say that one has demonstrated empirically that one quantity is not a function of another?
2. What is an objective probability?

We say that one quantity is a function of another (in the mathematical sense) if to every value of the second in a given range there corresponds exactly one value of the first. Since all empirical determinations of quantities are subject to errors one could argue that an empirical demonstration of a functional relation or of the absence thereof is strictly speaking not possible. In practice, this difficulty is circumvented by assuming that errors of a given magnitude occur with certain probabilities. As a consequence, one obtains a range of discrepancies between theoretical and observed quantities which can be attributed to chance fluctuations and thus a degree of "credibility" that an assumed functional relation exists or does not exist.

The second question is connected with difficulties of a more fundamental sort. These arise if one tactly assumes that an "objective probability" is an attribute of an event in quite the same way as a physical variable, e.g., volume, mass, or temperature is an attribute of an object. We can always make the latter assumption, because there exist operations (measurements) generally acceptable, which lead to the determination of the quantity in question. In this way it makes sense to compare experimentally the "objectively determined" quantity with a corresponding

"subjectively determined" one. The comparison is actually a comparison of the results of two measurements. Typically a subject is given a body to hold and is asked to estimate its weight (or to compare the weights of two bodies); or he is given a light to see and is asked to estimate its intensity. These estimates are his "subjectively determined" values of the quantities in question. They can be compared with the generally accepted "objectively determined" quantities.

Let us see whether this procedure can be applied in the case of probabilities. To my way of thinking, one can speak of an "objective probability" of an event only in terms of the observed frequency of the event in a sequence of experiments generally assumed to be performed under identical conditions. Thus, it is supposed that sequential tosses of a coin are such experiments, and so one can speak of the "objective probability" of Heads or Tails as the supposed limit of the frequency of its occurrence as the number of tosses approaches infinity. Therefore an experiment purporting to compare the objective probability of an event with a subjective probability must give the subject an opportunity to observe the relative frequency of the event in a sequence of experiments in quite the same way as he is given the opportunity to "observe" the weight of an object, the intensity of a light, or the pitch of sound.

Kriz assumes that the "objective probabilities" of the events presented to the subjects (in this case samples taken from a certain universe) are identical with probabilities calculated on the basis of assuming certain elementary events (e.g., all possible samples) to be equally probable. There is no objection to this assumption as such; for it may very well turn out that the relative frequencies of the various samples will correspond to the calculated probabilities in a large

sequence of experiments under proper conditions. However, the subjects are not given an opportunity to observe such frequencies, and so one cannot speak of comparing measurements, a procedure inherent in every experiment purporting to compare "objective" and "subjectively" estimated quantities. It seems, therefore, that the present experiments are not a straightforward extension to the realm of probability of such comparisons.

What then is the analogue of these experiments in the physical realm? It seems to me that the proper analogue would be somewhat like the following. The subject is asked to estimate the weight of an object by viewing it (not being allowed to hold it). Now the "objective" weight (mass) of an object can be gotten by multiplying its volume by its density. In viewing the object, the subject may be able to estimate volume but not density. It is, therefore, natural for him to guess a weight proportional to the estimated volume. If the object has an irregular shape, even the estimation of volume becomes difficult. The subject may be forced to take an even less reliable indicator of weight, say the magnitude of one of the dimensions. The results of such experiments would be analogues to those obtained by Kriz. Namely the estimates of the subjects (their "subjective" weights) would be more strongly correlated with some selected dimension of the object than with the "objective" weight. Thereby the absence of a functional relation between "objective" and "subjective" weights would have been demonstrated. Clearly however, this "absence of functional relation" would be simply a consequence of the fact that the subjects could not obtain all the information for determining the "objective" weight.

It can, of course, be argued that no pertinent information was actually withheld from the subjects in Kriz's experiments. There is a difference, however, between having all pertinent information available and being able to utilize

it. Suppose that subjects not familiar with physics were given all the pertinent information from which they could calculate the weight of an object, but not the opportunity to hold the object. The absence of a functional relation between estimates and objective measurements could still not be attributed (without further discussion) to an absence of a functional relation between "objective" and "subjective" weight. It would have to be attributed to the inability of the subjects to utilize the available information.

To avoid misunderstanding, I would like to emphasize my agreement with the general thesis posed by Kriz, namely that the tacitly assumed functional relationship between an individual's "subjective probability" and the "objective probability" of an event is a highly questionable one. The purpose of my remarks was to elucidate the sources of the difficulties in dealing with these concepts. These difficulties are of theoretical, epistemological, or semantic nature. My only reservation concerning Kriz's paper is that it seems to contain an implication that the absence of a functional relationship between "subjective" and "objective" probability can be demonstrated by a straightforward experimental procedure. That this is not so appears, I believe, from the foregoing analysis, since one can argue that the subjects have simply not been given enough (usable) information and so could not be expected to give evidence of their "subjective probabilities". The crux of the matter, in my opinion, is that the very concept of probability is not in the same category as attributes of objects and events which can be objectively measured and subjectively estimated. The important conclusions to be drawn from Kriz paper is that the extension of the concept of comparing "objective" with "subjective" attributes to the realm of probabilities is not legitimate without far reaching qualifications and elucidations of the meanings of these concepts in various situations.

Literatur

1. ATTNEAVE, F., "Psychological probability as a function of experienced frequency" J.exp.Psychol. 1953, 46, 81-86
2. COOMBS, C.H., BEZEMBINDER, T.G.G. & GOODE, F.M., "Testing expectation theories without measuring utility or subjective probability" J.math. Psychol. 1967, 4, 72-103
3. EDWARDS, W., "The prediction of decisions among bets" J.exp.Psychol., 1955, 51, 201-214
4. " " "The theory of decision making", Psychol. Bull., 1954, 51, 380-417
5. " " "Behavioral decision theory" Ann.Rev. Psychol. 1961, 12, 473-98
6. " " & TVERSKY, A., "Decision making", Penguin modern Psychol. UPS8, 1967
7. GRIFFITH, R.M., "Odds adjustment by American horse-race bettors", Amer.J.Psychol., 1949, 62, 290-94
8. KRIZ, J. "Der Likelihood-Quotient zur Erfassung des subjektiven Signifikanzniveaus", Forschungsber. No.9, Inst.f.Höhere Studien, Wien, Juli 1967
9. " " "Subjektives Signifikanzniveau im zwei-Stichprobenvergleich für Alternativmerkmale" Z.Psychol. (im Druck)
10. MOSTELLER, F. & NOGEE, P., "An experimental measurement of utility", J.polit.Econ., 1951, 59, 371-404
11. PETERSON, M.G. & BEACH, L.R., "Man as intuitive statistician", Psychol.Bull., 1967, 68, 29-46
12. PHILIP, B.R., "Generalization and central tendency in the discrimination of a series of stimuli" Can.J.Psychol., 1947, 1, 196-204
13. PETERSON, M.G., & BARATTA, P., "An experimental study of the auction-value of an uncertain outcome" Amer.J.Psychol., 1948, 61, 183-193
14. SHUFORD, E.H., "A comparison of subjective probabilities for elementary and compound events", Rep. No.20, The Psychometric Laboratory, University of North Carolina, 1959
15. SPROWLS, S.C., "Psychological-mathematical probability in relationship of lottery gamblers" Amer.J.Psychol., 1953, 66, 126-130

16. STEVENS, S.S. & GALANTER, E.H., "Ratio scales and category scales for a dozen perceptual continua", J.exp.Psychol., 1957, 54, 377-409
17. TVERSKY, A., "Additivity, utility and subjective probability", J.math.Psychol., 1967, 4, No.2, 175-202
18. VLEK, C.A.J., "The use of probabilistic information in decision making", Rep. 009-65, Psychol.inst.Univ.of Leiden, december 1965
19. VLEK, C.A.J. & BEINTEMA, K.A., "Subjective likelihoods in posterior probability estimation", Rep. 014-67, Psychol.inst.Univ.of Leiden january 1967
20. VLEK, C.A.J. & VAN DER HEIJDEN, L.H.C., "Subjective likelihood functions and variations in the accuracy of probabilistic information processing", Rep.017-67 Psychol. inst.Univ.of Leiden, november 1967
21. VON NEUMANN, J. & MORGENSTERN, O., "Theory of games and economic behavior", Princeton University Press, 2nd edition 1947
22. WENDT, D., "Versuche zur Erfassung eines subjektiven Verlässlichkeitsniveaus", Z.Psychol. 1966. 172, 40-81
23. " " "Zur Schätzung subjektiver Irrtumswahrscheinlichkeiten", Bericht über den 25. Kongress der Deutschen Gesellschaft für Psychologie, Göttingen, 1967, 302-307