S 22170 F

Sonderdruck aus

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Heft 1 / Band XXXIX 1. Quartal 1992

VERLAG FÜR PSYCHOLOGIE · DR. C. J. HOGREFE
GÖTTINGEN · TORONTO · ZÜRICH

Zeitschrift für experimentelle und angewandte Psychologie
1992, Band XXXIX, Heft 1, S. 18—55

Zur Messung prognostischer Leistung

August Fenk und Josef Vanoucek

Universität Klagenfurt

Antizipation kann als Voraussetzung für effizientes „Lernen" (im weitesten Sinne) und ihre Verbesserung als ein wesentliches Ziel und Erfolgskriterium des Lernprozesses angesehen werden. Dieser grundlegenden Bedeutung von Antizipation entsprechend, beschäftigt das Thema „subjektive Wahrscheinlichkeit" und „Paßgenauigkeit von Vorhersagen" eine ganze Reihe von Disziplinen, wie Induktions-, Entscheidungs-, Kognitions- und Kommunikationstheorie.

Vor diesem Hintergrund wird versucht, Shannons Ratespieltechnik zu einem empfindlicheren und vielseitig einsetzbaren Instrument zur Messung prognostischer Leistung weiterzuentwickeln. Die Shannonschen und alle verwandten Rateverfahren sind blind für das „metakognitive" Wissen des Raters um die (beschränkten) Trefferchancen seiner Prognosen. Es wird vorgeschlagen, diese subjektiven Wahrscheinlichkeiten mitzuerheben und als Gewichtszahlen zu behandeln: Höhere Wahrscheinlichkeitswerte erhöhen das Gewicht der Fehlprognosen (als Fehler) und das Gewicht der Treffer (als Treffer); je geringer die in der Rateprozedur zu beseitigende Fehleinschätzung, umso höher ist die prognostische Leistung einzustufen. Vor allem im Hinblick auf die Ansprüche einer Psychologie der „Informationsverarbeitung" wird ein weiterer Berechnungsmodus eingeführt, welcher es gestattet, den Prognosefehler informational (in bit) zu quantifizieren. Eine erste Erprobung erlaubte die Demonstration der gesteigerten Empfindlichkeit des Diagnose-Instruments sowie die Formulierung einer neuen Hypothese über kognitive Strategien angesichts von Entscheidungen unter Unsicherheit.

1. Zielsetzung

Diese Arbeit versucht, die auf Shannon (1951) zurückgehende Ratespieltechnik weiterzuentwickeln, und zwar zu einem konsequent am Konzept der subjektiven Information orientierten und (daher) valideren, empfindlicheren und möglichst vielseitig verwendbaren Instrument zur Messung prognostischer Leistung.

Bevor wir (in Abschnitt 3) ein wesentliches Defizit der klassischen Ratespieltechnik sowie zwei an diesem Defizit ansetzende Verbesserungsvor-

schläge (in den Abschnitten 4 und 5) diskutieren, ist der Zweck des ganzen Unternehmens zu erläutern.

2. Über „Antizipation" und den (potentiellen) Nutzen der Ratespieltechnik

Das Prinzip der „Ratespiel-Technik" bzw. der „Ratetests" wurde von Shannon (1951) als empirische Methode zur Bestimmung des Informationsgehalts sprachlicher Nachrichten entwickelt: Der Proband versucht den n-ten Buchstaben eines Texts zu erraten. Er macht eine Vorhersage und erfährt, ob sie falsch ist oder stimmt. Im Falle der negativen Rückmeldung versucht er es mit einer weiteren Prognose, und dies so oft, bis er eine positive Rückmeldung erhält; dann wird die Prozedur bei den nächsten Buchstaben (n + 1, n + 2, ...) wiederholt. Der Proband ist instruiert, mit einem Minimum solcher Prognosen (oder solcher „Fragen", die mit „ja" oder „nein" beantwortbar sind) das Auslangen zu finden. Unter der Voraussetzung, daß der Proband sich dieses Ziel einer Minimierung von (Fehl-)Prognosen zu eigen macht und seine Ratestrategie insofern rational ist, als die Reihenfolge der pro Rateplatz prognostizierten Buchstaben der Rangreihe dieser Buchstaben bezüglich ihrer subjektiven Wahrscheinlichkeit entspricht, unter dieser Voraussetzung also erlauben die Anzahl und Verteilung der bis zum Erraten nötigen Vorhersagen bzw. „Binär-Fragen" Rückschlüsse auf den Informationsgehalt des Texts. Genauer gesagt: Sie erlauben abzuschätzen, wieviel Information (Unsicherheit, Neuigkeitswert) der Text *für die ratende(n) Person(en)* enthalten hatte — denn empirisch faßbar ist auf diese Weise nicht der „objektive Informationsgehalt" im Sinne eines nachrichteninhärenten Merkmals bestimmten Ausprägungsgrades, sondern ausschließlich der *subjektive* Informationsgehalt (Frank, z. B. 1971, S. 53 f.), im Grunde genommen also nichts weiter als die informational quantifizierte *Inkongruenz* (Fenk & Lechner, 1986, S. 101) zwischen geäußerten (subjektiven; intersubjektiven) Erwartungen und einem entsprechenden, auf „stimmt/stimmt nicht" reduzierten Feedback.

Die Unmöglichkeit, vom jeweiligen Rater(kollektiv) zu abstrahieren, beeinträchtigt die Tauglichkeit des Ratetests für bestimmte, auf die Beschreibung der Nachricht fokussierende Untersuchungsziele. Für kognitionswissenschaftliche Zielsetzungen hingegen eröffnet die Einsicht, daß Information „nur" als Relation zwischen Nachricht und verarbeitendem System faßbar ist, ein in seiner Tragweite kaum abzuschätzendes Anwendungspotential der Ratespiel-Technik.

Kommunikation und Kognition haben eben viel miteinander zu tun, insbesondere durch ihrer beider Beziehung zur *Antizipation.* Einige Aspekte

dieser Beziehung: Antizipation kann als Voraussetzung und ihre Verbesserung als Ziel „kognitiver", erfahrungsbildender Prozesse aufgefaßt werden, unabhängig davon, ob der Gegenstand der kognitiven Analyse — die „Nachricht" im weitesten Sinne — eine in Verständigungsabsicht produzierte Botschaft ist oder aber die Geräuschproduktion eines defekten Verbrennungsmotors. Und die Einschätzung, wie gering der Zeichenaufwand bei der Produktion einer Nachricht gehalten werden kann, ohne die Verständigung zu gefährden, basiert auf den Erwartungen über die — z. B. „kontextinduzierten" — Erwartungen des Adressaten.

So gesehen kann es nicht verwundern, daß ein auf dem Boden der formalen Kommunikationstheorie (= Informationstheorie) gewachsenes Instrument sich auch oder gerade in der Kognitionsforschung als nützlich erweist: Das auf einen bestimmten Bereich bezogene „Wissen" bestimmter Informationsverarbeiter, also die Paßgenauigkeit („goodness of fit") der ihnen verfügbaren, mehr oder weniger „einschlägigen" Modelle läßt sich mit diesem Instrument, dem Ratetest, ebenso diagnostizieren wie die von bestimmten Bedingungen abhängige Weiterentwicklung ihres Wissens.

Erstaunlich also, daß ein derartiges Instrument in einer ganzen Artikelserie über „Forschungsmethoden der Wissenspsychologie" (Mandl & Spada, 1988, S. 359—469) unerwähnt bleibt; dabei wird in einem dieser Aufsätze — er befaßt sich mit kognitiven Prozessen in der Steuerung von Augenbewegungen (Lüer, 1988) — sogar ausdrücklich darauf hingewiesen, daß die bevorzugt fixierten Stellen sich durch höchsten Informationsgehalt auszeichnen. Reinert (1985) analysiert diesen Aspekt der Steuerung von Blickbewegungen noch viel differenzierter und auch im Hinblick auf die „Tendenz zur Antizipation von visuellen Informationen" (S. 136), aber wiederum ohne Erwähnung der Ratespieltechnik. Dies sowie die generelle Neigung, die Informationstheorie eher als Fundgrube für Redeweisen und Metaphern zu benutzen denn als analytisches Werkzeug, hängt wohl auch damit zusammen, daß die Informationstheorie (samt Ratespiel) meist in zu enger Verbindung mit dem Konzept der limitierten Kanalkapazität gesehen wird.

Die Anwendung dieses Konzeptes in der Modellierung unserer Informationsverarbeitung ist innerhalb der Psychologie in Mißkredit geraten (siehe z. B. Neumann, 1985). Daß der Ausdruck „Kanal" oft zu wörtlich bzw. nicht „in seiner genialen Allgemeinheit" verstanden wird, wie Frank (1990, S. 15) beklagt, mag ein Grund hiefür sein. Andererseits läßt gerade eine konsequente Verallgemeinerung des Konzeptes der „subjektiven Information" das „Kanal-Konzept" entbehrlich erscheinen (Fenk, 1986). Die Informationstheorie als formaler Apparat ist dagegen in ihrer Anwendung nicht auf nachrichtentechnische Fragen beschränkt; sie ist darauf ebensowenig beschränkt wie die Anwendung der Wahrscheinlichkeitstheorie auf das

Glücksspiel, die Anwendung der Spieltheorie auf strategische Spiele oder die Anwendung der „Zinseszins-Formel" auf den Kapitalzuwachs. Sie erweist sich außerhalb ihrer Ursprungsgebiete als „vielseitiges" Analyse-Instrument; mit seiner Hilfe lassen sich alte Fragen der Kognitionspsychologie (z. B. nach dem Ausmaß des Lernzuwachses in Abhängigkeit vom Ausgangsniveau, Fenk, 1986) und aktuelle Fragen aus dem Überschneidungsbereich von Kognitions- und Kommunikationspsychologie (z. B. nach der „Verständlichkeit" von Texten und von Text-Bild-Kompositionen. Siehe unten!) so (re)formulieren, daß sie einer empirischen Prüfung zugänglich werden — und zwar auf dem eigentlich interessierenden, allgemeinen und vom je spezifischen „Wahrnehmungsangebot" und „Lernmaterial" unabhängigen Niveau. Grund genug also, sich mit der Weiterentwicklung dieses Instrumentes zu beschäftigen. Aber das jeweilige Verständnis und die jeweilige Operationalisierung von prognostischer Leistung rührt am Nerv nicht nur der Kognitions- und Kommunikationspsychologie, sondern auch der Testtheorie, der Induktionstheorie und der Entscheidungstheorie.

Die folgenden Andeutungen beziehen sich zuerst auf diese „-theorien", und dann, etwas ausführlicher, auf die „Verständlichkeitsforschung".

2.1. Testtheorie

Das problematischste, aber auch das wichtigste Testkriterium ist bekanntlich die sogenannte „externe Validität", und diese ist im Grunde genommen ein Kriterium für die prognostische Valenz eines (z. B. Eignungs-) Tests. „Soll ein Testwert Kriteriumswerte vorhersagen, welche erst in der Zukunft liegen, so spricht man" laut Dorsch (1987, S. 720) ausdrücklich von „Vorhersage-Validität". In gewissem Sinne hat „Validität" aber ganz grundsätzlich mit „Vorhersage" zu tun: Wissenschaftstheoretiker wie Schlick (1948, S. 18) verwenden den Terminus „Vorhersage" nicht nur im Zusammenhang mit zukünftigen Ereignissen, sondern ganz generell im Zusammenhang mit Merkmalen und Ereignissen, von denen der „Vorhersagende" zum Zeitpunkt der „Vorhersage" noch keine direkte Kenntnis hat. Und im Sinne der Regressionsstatistik kann man aus der Kenntnis des — übrigens seinerseits informationstheoretisch faßbaren (Greeno, 1970) — statistischen Zusammenhanges zweier Variablen (z. B. Alter und Länge bestimmter Fische) und aus Kenntnis des konkreten Ausprägungsgrades einer der beiden Variablen den Ausprägungsgrad der jeweils anderen Variable „vorhersagen". So läßt sich aus einer Theorie „vorhersagen", welche Fossilien in einer Sediment-Schicht bestimmten Alters dominieren, welche Fossilien man also in ihr vorzugsweise finden *wird,* so man danach sucht; und umgekehrt läßt sich so aus der Art der gefundenen Fossilien „vorhersagen", welches Alter diese Schicht hat bzw. welches Alter eine (andere) Altersbestimmungsme-

thode ergeben *wird.* Was in solchen Fällen von Antizipation — oder von Rekonstruktion? — in der Zukunft liegt, ist also nicht das interessierende, z. B. naturhistorische Phänomen, sondern ein weiterer Hinweis auf dieses. Antizipation und Rekonstruktion haben also zumindest eines gemeinsam, nämlich die Projektion oder Extrapolation nomologischen „Wissens" auf noch unbekannte Instanzen. Damit sind wir eigentlich schon beim Thema

2.2. Induktionstheorie

Prognostische Leistung ist ein hervorragendes Kriterium zur Bewertung von (konkurrierenden) Theorien. (Von einer „Konkurrenz" zwischen Theorien sollte nur dann die Rede sein, wenn sich ihre Domänen — also die Bereiche von Phänomenen, welche sie zu erklären und vorherzusagen beanspruchen — zumindest überschneiden.) Denn nicht jede Theorie, welche für ein beobachtetes Phänomen eine plausibel klingende Erklärung parat hat, kann dieses Phänomen auch vorhersagen oder es mit derselben Präzision vorhersagen wie eine konkurrierende Theorie(variante). Und für die insbesondere von Coombs (1984) vorgeschlagenen Dimensionen erfahrungswissenschaftlichen Fortschrittes — „power" und „generality" — läßt sich zeigen (Fenk, 1988), daß ihre Diagnose eine Diagnose prognostischer Leistung voraussetzt:

— Einer Zunahme an „power" entspricht aus dieser Perspektive eine Zunahme der Trefferhäufigkeit, welche nicht auf Kosten des empirischen Gehalts der prognostizierenden Aussage geht, oder aber eine Zunahme des empirischen Gehalts der prognostizierenden Aussage, welche nicht auf Kosten ihrer Trefferhäufigkeit geht.
— Einer Zunahme an „generality" bzw. einer Vergrößerung des Geltungsbereiches (bzw. der „Domäne") entspricht aus dieser Sicht ein höheres Abstraktionsniveau der in den prognostizierenden Aussagen vorkommenden Begriffe, sofern diese Form von Verallgemeinerung nicht auf Kosten der Trefferhäufigkeit geht.

Am Beispiel der folgenden gesetzesartigen Aussagen läßt sich dies illustrieren:

a) „(Alles) Eisen dehnt sich bei Erwärmung aus."
b) „(Alle) Metalle dehnen sich bei Erwärmung aus."
c) „(Alles) Eisen dehnt sich bei Erwärmung von 18 auf 19 Grad Celsius um 10 bis 15 Millionstel Volumenprozent aus."
d) „(Alle) Metalle dehnen sich bei Erwärmung von 18 auf 19 Grad Celsius um 10 bis 15 Millionstel Volumenprozent aus."

Verglichen mit (a) ist (b) eine generellere und (c) eine präzisere, empirisch gehaltvollere Aussage. Und (d) ist sowohl genereller als auch präziser — der sprachlichen Form und dem Anspruch nach; sie wird sich aber, im Unterschied zu den anderen Aussagen, als falsch erweisen, weil sich Metalle mit deutlich größeren Ausdehnungskoeffizienten finden. Wir haben also strikt unterschieden zwischen einerseits semantischen Merkmalen der (Vor-)Aussage — wie Präzision („empirischer Gehalt") und Allgemeinheit —, und andererseits dem Zutreffen der (Vor-)Aussagen; dieses zweite Kriterium muß erfüllt sein, damit auch der erfahrungswissenschaftliche Anspruch an „power" und „generality" erfüllt ist. Die getroffene Unterscheidung erlaubte es uns (siehe oben!), einen speziellen Aspekt der Coombschen Argumentation auf dem elementaren Niveau isolierter Gesetzesaussagen (über Eisen und andere Metalle) zu demonstrieren: „This aspect is the tradeoff between „*power* und *generality.*" (Coombs, 1984, S. 20) „The experimenter /.../ can either seek to strengthen the theory, so it says more about the same domain, or seek to broaden the domain so the theory says the same about more things." (ebd., S. 24) Und vielleicht gilt die Coombsche tradeoff-These *nur* für unsere isolierten Gesetze, nicht aber für eine Theorie i. S. eines komplexeren Aussagengebäudes? Denn möglicherweise oder erwiesenermaßen läßt sich in Experimenten, in denen die Molekülstruktur oder die Dichte verschiedener Metalle die Rolle der unabhängigen Variable spielt, der Ausdehnungskoeffizient eines jeden einzelnen Metalls (hoher Grad an „generality") als Funktion dieser Variable *und* in sehr präziser Weise (hoher Grad an „power") vorhersagen.

Wie auch immer: Bei der Diskussion von Kriterien zur Diagnose der Leistungsfähigkeit von (konkurrierenden) Theorien bzw. zur Diagnose des durch eine Modifikation der Theorie erreichten Fortschrittes liegt es nahe, auf den Begriff der „Information" zu rekurrieren (siehe dazu Oeser, 1976). Auch Ernst von Glasersfeld (1987, S. 219) sieht seine konstruktivistische Interpretation von „Erkenntnis" (und von „Lernen") nicht in Widerspruch zu Norbert Wieners Kybernetik und seine konstruktivistische Interpretation von „Kommunikation" nur als „eine Erweiterung der Theorie Wieners (und Shannons)". Außerdem sieht er (ebd., S. 280f.) seine Position in Übereinstimmung mit einer „instrumentalistischen Erkenntnistheorie", deren Grundprinzip bereits Osiander (1627) formuliert: „Es ist nicht notwendig, daß /.../ Hypothesen wahr sind, oder daß sie überhaupt der Wahrheit ähnlich sind; eines ist für sie ausreichend — sie müssen Berechnungen zulassen, die mit den Beobachtungen übereinstimmen." Was nach v. Glasersfeld den Wert von „begrifflichen Strukturen bestimmt, das ist ihre Erfahrungsadäquatheit, die Güte, mit der sie zur Erfahrung passen..."

„Empirische" Vergleiche zwischen empirischen Theorien bezüglich ihrer prognostischen Leistungsfähigkeit innerhalb einer bestimmten Ereignisklas-

se hängen aber in ihrer Qualität ab von der Qualität der Meßmethode, welche zwischen verschiedenen Ausmaßen des Passens von Prognosen auf Datenstrukturen diskriminieren soll. Das Kriterium des Passens auf Erfahrungsinstanzen geht übrigens auch in verschiedene Fassungen des Begriffes der „Gesetzesartigkeit" ein. Nicht nur bei Schlick, der „das Eintreffen von Voraussagen" explizit als das „wahre Kriterium der Gesetzmäßigkeit" eines Ablaufes bezeichnet (Schlick, 1931; zitiert nach Schlick, 1948, S. 18), sondern auch bei Goodman, der Voraussetzungen der „Gesetzesartigkeit" von Aussagen analysiert: „Projizierbar" und damit gesetzesartig ist eine Aussage nur so lange, als sie als „supported, unviolated, and unexhausted" gelten kann (Goodman, 1973, S. 100). Allerdings scheint das Kriterium „bisher (unterstützt und) unverletzt" nur auf generelle Gesetze anwendbar zu sein; „statistische Gesetze", mit denen wir es zunehmend und keineswegs nur in den „weichen" Wissenschaften zu tun haben (vgl. Salmon, 1971), sind von vornherein nicht wirklich verletzbar, weil ihre Formulierung als „Regularität" oder „Tendenz" von vornherein „Ausnahmen" zuläßt. Angesichts statistischer Gesetze müssen wir uns vermutlich mit einem „mehr oder weniger gut unterstützt" begnügen, also mit dem Kriterium einer größeren oder geringeren Treffergenauigkeit der Vorhersagen.

Mit all diesen Überlegungen über Kriterien zur Entscheidung zwischen Theorien bewegen wir uns gleichzeitig auf einem speziellen Gebiet der

2.3. Entscheidungstheorie

Die subjektive Einschätzung der Wahrscheinlichkeit(en), mit der bestimmte Handlungsmöglichkeiten zu bestimmten, als mehr oder weniger vorteilhaft eingestuften Ergebnissen führen, ist aus der Entscheidungs-Findung und -Begründung nicht wegzudenken. Dies gilt wohl auch dann, wenn man die subjektive Wahrscheinlichkeit nur „als eine Facette von Erwartung" (Westhoff, 1985, S. 64) betrachtet, und unabhängig davon, welcher Variante von „Subjective Theories" (vgl. Smithson, 1989, S. 64) man zuneigt. Und beim Thema „Induktionstheorie" ging es eben um begründbare Entscheidungen für eine bestimmte Theorie, für bestimmte Hypothesen samt Operationalisierung, und damit um *heuristische Entscheidungen* in jenem Vorfeld praxisnäherer Handlungsentscheidungen, in welchem wir uns (innerhalb wie außerhalb der Wissenschaft) möglichst „paßgenaue" Wahrscheinlichkeitsverteilungen erarbeiten. Die gewonnene Bewertung wird u. a. unsere (praktische?) Entscheidung mitbestimmen, welche der denkmöglichen Theorien wir zuerst und mit welchem Aufwand weiterverfolgen und einer genaueren Prüfung unterziehen. Neue Hypothesen können sich Peirce zufolge nur auf dem Wege der „Abduktion" erschließen. Und Anlaß für abduktives Schließen ist — wir zitieren Reichertz (1988, S. 352 f.) — „das Un-

erwartete, das Unheimliche, das Neue, das quer zum gewohnten Leben steht und Unruhe in unseren Plan, was als nächstes zu tun ist, bringt. ‚Abduktion ist jene Art von Argument, die von einer *überraschenden Erfahrung* ausgeht, das heißt von einer Erfahrung, die einer aktiven oder passiven Überzeugung zuwiderläuft' (Peirce, 1983, S. 95). /.../ Die Abduktion ist eine nach rückwärts gerichtete alltägliche Prognostik."

Diese letzten Sätze eignen sich im übrigen auch zur Charakterisierung unserer Strategien der Aufmerksamkeitsverteilung. Diese Strategien sind darauf abgestellt, das (überraschende und daher) interessierende Geschehen rekonstruierbar und „verständlich" zu machen.

2.4. Verständlichkeitsforschung

Ein auch durch die Bedürfnisse der Software-Designer besonders aktuell gewordenes Thema der Verständlichkeitsforschung wird heute unter den Stichworten „Text/Bild-Interaktion, -Integration, -Komplementarität" diskutiert. Im Bild-Verständlichkeitskonzept von Böhme (1987) ist „Komplementarität" das in erster Linie maßgebliche Beurteilungskriterium gerade für jene Sorte von Bildern, die den Charakter von „Arbeitsinstrumenten" besitzen. Ballstaedt (1990) versucht eine Klassifizierung der Arten und Weisen des Zusammenpassens von Text und Bild, rekurriert dabei auf kognitionstheoretische Konstrukte (wie „Schema") und unterscheidet „drei Fälle der audiovisuellen Integration /.../:

1. *Integration durch Redundanz.* Text und Bild werden über Aktivierung desselben Konzepts oder derselben Konzepte integriert. /.../
2. *Integration durch Komplementarität.* Text und Bild aktivieren zwar verschiedene Konzepte, diese sind aber innerhalb eines Schemas miteinander verknüpft. /.../
3. *Integration durch Inferenz.* Text und Bild werden über Schlußfolgerungen zwischen verschiedenen aktivierten Konzepten integriert. Dieser Fall setzt eine Text-Bild-Schere voraus. /.../

Betrachten wir die drei Formen der audiovisuellen Integration, so läßt sich feststellen, daß sie einen zunehmenden *Verarbeitungsaufwand* erfordern." (Ballstaedt, 1990, S. 35).

Ballstaedts Klassifikation erscheint in unserem Zusammenhang aus zwei Gründen interessant. Erstens, weil hier — wie im Konzept der subjektiven Information — Text und Bild nicht von sich aus zusammenpassen, sondern erst durch die mehr oder weniger direkt „passenden" Interpretationen eines beziehungsstiftenden Subjekts „zusammengepaßt" werden. Und zweitens und vor allem, weil Ballstaedts Hinweis auf den vom „einfachsten" Fall 1 bis zum „anspruchsvollsten" Fall 3 zunehmenden *Verarbeitungsaufwand* an

eine zugrundeliegende Dimension denken läßt, an ein irgendwie faßbares *Ausmaß,* in welchem Text und Bild (subjektiv, intersubjektiv) zueinanderpassen (Text-Bild-Korrespondenz) bzw. nicht zueinanderpassen (Text-Bild-Inkongruenz):

Vielleicht ist es nur diese eine Dimension, die uns unter einer Vielzahl von Namen (im folgenden *kursiv*) begegnet:

— Etwa als *interne Redundanz* zwischen verschiedenen Komponenten — Text, Bild — einer Botschaft. (Redundanz ist so gesehen kein den „einfachsten" Fall Ballstaedts kennzeichnendes oder abgrenzendes Merkmal, sondern ein Kontinuum, auf dem sich die Fälle 1 bis 3 aufreihen lassen.) Oder umgekehrt als *Text-Bild-Schere* — je geringer die interne Redundanz (oder die Transinformation; siehe unten!) zwischen der Text- und der Bild-Komponente, umso größer die Schere. (Im „einfachsten" Fall Ballstaedts ist diese Schere relativ gering. Aber zum einen ist sie auch dort vorhanden, denn Text und Bild können nie völlig kongruent sein; keines von beiden kann durch das jeweils andere völlig ersetzt oder exakt in das jeweils andere übersetzt oder im jeweils anderen „wiederholt" werden. Und zum anderen markiert der „einfachste" Fall 1 auch nicht das untere Ende dieses Kontinuums: Zwischen den Zahlenwerten für das Gewicht eines Säuglings nach seiner ersten, zweiten, dritten etc. Lebenswoche und einem entsprechenden Diagramm ist beispielsweise die „Schere" noch geringer bzw. die „Redundanz" noch größer.)

— Oder als *Transinformation* zwischen Text und Bild. Abbildung 1 illustriert die Transinformation als Überschneidungsbereich zwischen einem Informationsfeld X (z.B. Bild) und einem Informationsfeld Y (z. B. Text). Beide Kreisflächen zusammen stehen für die Gesamtinformation der Botschaft. Je größer bei vorgegebenen Kreisflächen ihr Überschneidungsbereich ist, umso geringer ist die „komplementäre" Information, die das Bild oder die den Text zur Totalinformation ergänzt, und umso geringer ist die Gesamtinformation sowie der zu ihrer Verarbeitung notwendige Aufwand.

Ähnliche Überlegungen wurden bereits von Ballstaedt, Molitor und Mandl (1989, S. 122 f.) angestellt: „Der Begriff der *Komplementarität* taucht in der vorhandenen Forschungsliteratur nicht explizit auf, statt dessen der Begriff der Redundanz, mit dem inhaltliche Überlappungen zwischen der Text- und Bildinformation bezeichnet werden. Gemeint sind damit alle inhaltlichen Details, die in beiden Medien präsentiert werden." Der „Redundanzbereich" sowie der komplementäre Bereich und die Gesamtinformation werden nach Art eines Venn-Diagramms durch zwei einander überlappende Ellipsen veranschaulicht. Die beiden auch isoliert gezeichneten Ellipsen umfassen noch einen dritten Bereich, und zwar „diejenigen Text-

bzw. Bildanteile, die weder redundant noch komplementär sind, sondern als textliche oder bildliche Ausschmückungen für den Lerninhalt verzichtbar wären." (In der „Gesamtinformation" scheint dieser Bereich nicht auf.) Wie aus diesen Formulierungen und Darstellungen hervorgeht, werden hier die Begriffe „Information" und „Redundanz" nicht sehr eng mit dem Formalismus der Informationstheorie assoziiert und — wohl aus diesem Grunde — auch nicht mit den Operationalisierungs- und Quantifizierungsmöglichkeiten, welche die Informationstheorie offeriert.

Der informationstheoretische Begriff „Redundanz" weist aber bereits den Weg zu einer Operationalisierung der uns interessierenden Größe: Ratetests erlauben eine empirische Bestimmung des *„Informationsbeitrages"* (Fenk, 1990 a, b), das heißt jenes Ausmaßes, in welchem die Kenntnis von X die Unsicherheit bezüglich Y reduziert bzw. die Antizipierbarkeit und Rekonstruierbarkeit von Y erhöht. Die Ersparnis an Ratefehlern, die ein Raterkollektiv mit Kenntnis von X (z. B. Bild) gegenüber einem „naiven" Raterkollektiv beim Erraten von Y (z. B. Text) erzielt, entspricht dem Informationsbeitrag von X bzw. dem Ausmaß, in welchem X den für Y nötigen „Verarbeitungsaufwand" reduziert. So läßt sich auch untersuchen, welcher von mehreren Bildentwürfen die kognitiven Kosten des Textverstehens am stärksten verringert.

Weltner stellt zwei frühe experimentelle Anwendungen der Transinformationsanalyse in der Bildforschung einander gegenüber: Mit dem Ziel, die „semantische Information" von Bildern zu ermitteln, untersuchte Frank (1967) die „Transinformation" von Bildern über eine Bildbeschreibung und Heinrich (1970) die „Transinformation" einer „Bildbeschreibung" — bzw. einer sprachlichen Bezeichnung für das im Rasterbild Dargestellte („Elefant") — auf das Bild. Unser Vorschlag entspricht demgegenüber

- einer Erweiterung des Anwendungsbereiches — weg vom Spezialfall Bild/Bildbeschreibung, hin zu verschiedensten Text-Bild-Kompositionen.
- einer veränderten Zielsetzung — weg von dem Bemühen, die sprachlich codierte oder codierbare „Bedeutung" von Bildern zu erfassen, hin zur Erfassung der Dimension Text-Bild-Integration bzw. Text-Bild-Kohärenz.
- einer gewissen Skepsis
  - gegenüber allen Versuchen, auf empirischem Wege den Inhalt von der syntaktischen Form zu isolieren (vgl. Luther & Fenk, 1984) bzw. die „semantische Information" von anderen (z. B. „ästhetischen") Komponenten zu isolieren, sie mit „Bedeutung" gleichzusetzen und so letztere zur — vermeintlich! (vgl. Fenk, 1980, S. 11) — skalaren Größe zu machen.
  - bezüglich der Möglichkeit, dem Begriff der „Transinformation" in der Operationalisierung gerecht zu werden: Will jemand die „Transinformation" von X (z. B. Bild, oder Textversion X) auf Y (z. B. Text, oder Textversion Y) und vice versa messen, so setzt er voraus — „Transinformation" ist ja symmetrisch! —, daß die beiden Arten der Messung zu Werten führen, die sich zumindest nicht systematisch unterscheiden. Will man diese Implikation bzw. diesen Anspruch an die Operationalisierung vermeiden, so empfiehlt es sich, den Transinformationsbegriff — evtl. zugunsten des Begriffes „Informationsbeitrag" — letztlich wieder zu verlassen.

Wenn man die drei Formen Ballstaedts „nur" als verschiedene Abschnitte *einer* meßbaren Dimension auffaßt, so macht dies natürlich Ballstaedts Klassifikation nicht hinfällig. Die Begriffe „Kälte" und „Wärme" sind ja auch nicht überflüssig geworden, seit der Begriff und die Messung der „Temperatur" eingeführt worden sind. (Dies gilt auch für Wissenschaftssprachen: Man denke nur an die in der Sinnesphysiologie beschriebenen „Wärme-Rezeptoren" und „Kälte-Rezeptoren".) Umgekehrt war die Einführung des Temperatur-Begriffes bekanntlich nicht überflüssig, sondern erwies sich — mit zunehmender Meßgenauigkeit in zunehmendem Maße — als sehr nützlich. Und für die Schätzung des Informations-*Beitrages* gilt: Sie kann nicht präziser sein als der Ratetest zur Erfassung des subjektiven Informations-*Gehaltes,* da sie ja nichts weiter ist als die Differenz aus zwei unter verschiedenen Bedingungen erhobenen Werten des subjektiven Informationsgehaltes einer Nachricht(enkomponente).

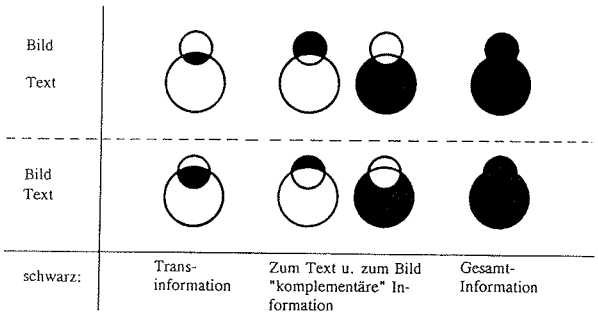

Abb. 1:
Größere Transinformation (untere Bilderzeile!) bedeutet unter sonst gleichen Bedingungen eine entsprechende Verringerung der „Total-Information" sowie der zum Text und der zum Bild „komplementären" Information (aus Fenk, 1991, S. 22)

Die beschriebene Methode zur Diagnose der Text-Bild-Kohärenz ist natürlich nur ein Beispiel aus einer Vielzahl von Anwendungsmöglichkeiten der Ratespieltechnik auf dem weiten Feld der Verständlichkeitsforschung. Verglichen mit den dort üblichen Diagnose-Methoden — z. B. jeder Art von Hinterher-Prüfung — ist sie „dichter" am Geschehen bzw. enger verknüpft mit dem Prozeß des forschenden, hypothesenprüfenden Erkennens und Verstehens. Zum anderen liefert sie — anders als etwa die Methode des „lauten Denkens" — quantitative Resultate auf sehr direkte und eindeutige, im testtheoretischen Sinne „objektive" Art und Weise.

3. An welchem Defizit des Ratespiels kann seine Weiterentwicklung ansetzen?

Die eingangs beschriebenen Ratetests lassen eine relevante „Informationsquelle" zur Schätzung der subjektiven Information — oder auch eine Komponente subjektiver Information — außer acht:

Instruktionsgemäßes Verhalten des Raters immer vorausgesetzt, verraten uns herkömmliche Ratetests zwar, welches Ereignis der Rater an einer bestimmten Stelle der Rateprozedur für das wahrscheinlichste hält, nicht aber, für WIE wahrscheinlich er es hält.

Dies stellt sich in zweifacher Hinsicht als Manko dar:

a) *Herkömmliche Ratetests bilden die subjektive Wahrscheinlichkeitsverteilung des Raters nur auf dem reduzierten Niveau von Rangskalen ab.* Wenn die ratende Person an einer bestimmten Stelle der Prozedur auf „D" tippt, dann bedeutet das, daß sie „D" im Vergleich zu anderen Möglichkeiten für wahrscheinlicher (oder zumindest für gleich wahrscheinlich) hält; ob sie „D" für gleich wahrscheinlich oder für doppelt so wahrscheinlich hält wie irgendeine andere Möglichkeit oder für wahrscheinlicher als alle anderen Möglichkeiten zusammengenommen, das alles erfahren wir nicht aus der bloßen Prognose des Buchstabens. Obwohl theoretisch gesehen die subjektive Information eine Funktion der subjektiven Wahrscheinlichkeit(sverteilung) ist — ganz so wie die „objektive" Information als Funktion der „objektiven" Wahrscheinlichkeit(sverteilung) zu sehen ist — verzichtet der herkömmliche Ratetest darauf, Genaueres über diese subjektive Wahrscheinlichkeit(sverteilung) in Erfahrungen zu bringen.

b) *Herkömmliche Ratetests (und Lücken-Tests) sind blind für das „metakognitive" Wissen des Raters um die (beschränkten) Trefferchancen seiner Prognosen.* Wie sicher sich der Rater ist, daß sich seine Prognose als Treffer erweisen wird, bzw. wie unsicher er ist, ob sich seine Prognose als richtig erweisen wird, geht in die Ratetest-Daten nicht ein.

Wir müssen hier nicht näher untersuchen, ob die Punkte (a) und (b) zwei verschiedene Defizite beschreiben oder ob sie *ein* Defizit in verschiedenen Formulierungsvarianten beschreiben. Denn für die im folgenden diskutierten Auswege macht dies keinen Unterschied.

4. Ein Lösungsweg und ein probabilistischer Score für prognostische Leistung

4.1. Das Prinzip des Lösungsweges

Eine bei den oben formulierten Defiziten ansetzende Weiterentwicklung ist denkbar „by means of an estimate made by the subject on the chance of success he ascribes to each guess" (Fenk, 1988, S. 41; vgl. auch Fenk, 1990 a, S. 369). In einem entsprechenden Verfahren gibt also die ratende Person nicht nur preis, welches Ereignis sie am ehesten erwartet, sondern auch die Wahrscheinlichkeit, mit der sie „es" — dieses Ereignis und damit das Eintreffen der Prognose — erwartet.

Wie sind aber nun diese Angaben über die subjektive Wahrscheinlichkeit (in Prozenten oder in Form von Wettquotienten) zu behandeln? Der „subjektive Informationsgehalt" ist leistungstheoretisch gesehen ein Schlechtpunktescore, und in diesem Sinne kann man die Fehlprognosen einfach mit dem jeweiligen Wahrscheinlichkeitswert gewichten: Eine Fehlprognose wiegt als Fehler umso schwerer, je sicherer sich der Vorhersagende seiner Sache gewesen war. Bei den Treffern ist das Gegenteil der Fall — je sicherer sich der Vorhersagende seiner Sache gewesen war, umso besser, je unsicherer er sich gewesen war, umso schlechter (... war er darüber im Bilde, was da bevorstand)! Nicht nur für Fehlprognosen, sondern auch für Treffer gilt also: Je besser die geäußerten „apriori"-Wahrscheinlichkeiten auf die Verteilung der später einlangenden Daten bzw. Feedbacks passen, umso höher die prognostische Leistung bezüglich der jeweiligen Ereignisfolge. Anders gesagt: Wenn zwei Rater bei einem Text exakt gleich viele Treffer erzielen und exakt gleich viele Fehler in Kauf nehmen müssen, kann dennoch ein Leistungsunterschied zwischen ihnen bestehen. Und zwar ist derjenige der besser „Informierte", bei dem die „Wettquotienten" dem folgenden Ideal am nächsten kommen: Bei Wahrscheinlichkeitsangaben von „100%" werden in 100% der Fälle Treffer erzielt, bei den Wettquotienten von „50 (zu 50)!" in 50% der Fälle, und so fort. (Vgl. dazu auch die Überlegungen von Garner, 1962, z. B. S. 24 und 94, sowie das Prinzip von Untersuchungen zum Thema „Calibration". Vgl. Smithson, 1989, S. 166 f.).

Damit ist auch einem Einwand begegnet, der da lauten könnte: Zwei Personen (X, Y) würfeln und sind bei ihren ersten beiden Würfen und damit in 100% der Versuche erfolgreich in der Aufgabe, die nächste von ihnen erreichte Augenzahl vorherzusagen. X hatte seine Trefferchance mit 99%, Y die seine mit 16,6% (= 1/6) beziffert. Im Sinne der weiter oben angestellten Überlegungen hätte also X „besser" abgeschnitten, obwohl — „objektiv" und „logisch" gesehen — nur Y seine Trefferchance richtig beurteilt hat. Der Gegeneinwand: Der Einwand ist nur berechtigt auf Basis der Vorannahme der *Zufälligkeit.* (Diese Vorannahme schließt unvorsichtigerweise von vornherein jede Möglichkeit eines „unfairen" Würfels und Manipulationen in der Würfelprozedur aus. Und sie schließt — unvorsichtigerweise? — „hellseherische" Begabungen aus, die dem Probanden sogar angesichts von Zufallsfolgen zu überzufällig hohen

Trefferquoten samt entsprechendem Wettquotienten verhelfen würden.) Die subjektive „Sicherheit" bezüglich dieser Vorannahme würde auf seiten eines durch Erfahrung belehrbaren Beobachters wohl sinken, wenn die Serie genau so weitergeht wie oben beschrieben; er würde sich der Möglichkeit bewußt werden, daß er selbst am allerschlechtesten, Y schon besser und X am allerbesten im Bilde ist über das, was da alles „im Spiel ist".

Sich auf ein bestimmtes Ereignis mit einem bestimmten Wahrscheinlichkeitswert festzulegen — die Wahrscheinlichkeit, daß die Augenzahl 3 oder eine ungerade Augenzahl oder der Buchstabe D folgt, mit einem Prozentwert zu beziffern — bedeutet, allen restlichen Möglichkeiten zusammengenommen den komplementären Wert, also die Ergänzung auf 100%, zuzuschreiben. (Dies entspricht der in der Wahrscheinlichkeitstheorie axiomatisch begründeten „Komplementarität der Negation". Siehe Smithson, 1989, S. 136.) Ein Extremfall besteht also darin, dem Ereignis D die Wahrscheinlichkeit von 100% bzw. von 1 zuzuschreiben; der Komplementär-Wert, die Wahrscheinlichkeit der Alternative(n) — also aller Nicht-D — hat dann die Größe Null. (Der andere Extremfall, daß eine Prognose abgegeben und die Trefferchance dann mit Null beziffert wird, kann uns nur bei nicht instruktionsgemäß handelnden Ratepersonen begegnen. Vgl. 4.2.!) Erweist sich D bei der subjektiven Wahrscheinlichkeit von 100% als Treffer, so ist dies für die Rateperson nicht informativ, sondern informationslos; durch das Feedback „D stimmt!" wird keine Unsicherheit beseitigt, keine Fehleinschätzung korrigiert. Wäre sich die Rateperson nur zu 90% sicher gewesen, so wäre auch im Falle des Treffers, also durch das positive Feedback, Unsicherheit beseitigt worden. Erweist sich aber D bei einer subjektiven Wahrscheinlichkeit von 100% als Fehlprognose („D stimmt nicht!"), dann ist der Überraschungs- bzw. Informations-Wert maximal, und rückblickend gilt dies auch für die Fehleinschätzung des Raters. Hätte die subjektive Wahrscheinlichkeit nur 90% betragen, so hätte dies einer weniger krassen Fehleinschätzung entsprochen, die negative Rückmeldung hätte weniger überrascht. (Vgl. Tabelle 1)

Tabelle 1

Der Prognose-Fehler „PF" bei richtigen und bei falschen Vorhersagen unterschiedlicher „subjektiver Wahrscheinlichkeit"

	PF bei	
subj. W. in %	richtiger Vorhersage	falscher Vorhersage
99	.01	.99
90	.10	.90
50	.50	.50
10	.90	.10
1	.99	.01

Tabelle 1 könnte, für sich allein betrachtet, den Eindruck erwecken, als wollten wir eine richtige Prognose (mit z. B. 10% subjektiver Wahrscheinlichkeit) auf dieselbe Leistungsstufe stellen wie eine falsche Prognose (mit z. B. 90% subjektiver Wahrscheinlichkeit); denn in diesen Beispielsfällen beträgt der „Fehler" jeweils 90%. Im Falle des *negativen Feedbacks* („D stimmt nicht!") wird aber — wenn es mehr als nur ein Nicht-D gibt — nur die Unsicherheit bezüglich der beiden Feedback-Möglichkeiten völlig beseitigt, nicht aber die Unsicherheit bezüglich der richtigen „Lösung", also bezüglich des wirklich folgenden Buchstaben. Wollen wir das *Gesamtausmaß* der Fehleinschätzung des Raters ausloten und irgendwie abbilden, so

— dürfen wir als Feedback nicht gleich die richtige Lösung („Es kommt ein N!") offerieren. Denn wer sagt uns, daß N für den Rater an dieser Stelle die einzige oder nächstbeste Alternative zu D gewesen wäre? Nach einem *negativen* Feedback („Stimmt nicht!") müssen wir die Rateprozedur so lange fortsetzen, bis der Rater eine richtige Prognose abgibt.
— müssen wir auch die durch das *positive* Feedback beseitigte Unsicherheit einkalkulieren. (Und diese kann beträchtlich sein. Muß ich eine sechsstellige Zahl frei erraten, so taxiere ich meine Trefferchance sehr gering und wäre selber entsprechend überrascht, hätte ich auf Anhieb Erfolg.)
— müssen wir die subjektiven Wahrscheinlichkeitswerte bei Treffern und Nicht-Treffern unterschiedlich behandeln. Höhere Wahrscheinlichkeitswerte erhöhen das Gewicht der Fehler (als Fehler) und das Gewicht der Treffer (als Treffer); in der Leistungsbilanz müssen sie also „gegensinnig" zu Buche schlagen und einander — ganz im Sinne des besagten Komplementaritätsprinzips — ergänzen.

Zwar könnten auch in einem Verfahren, in welchem pro Rateplatz nur einmal geraten und dann die richtige Lösung präsentiert wird, die geäußerten Wahrscheinlichkeitswerte „gegensinnig" bewertet werden. (Die Werte könnten z. B. mit positivem Vorzeichen im Falle von Treffern und mit negativem Vorzeichen im Falle von Fehlprognosen aufsummiert werden, so daß die über die Rateplätze gemittelte „Leistung" Werte zwischen maximal +1 und −1 annimmt.) Ein derartiges Verfahren würde aber den oben genannten Anforderungen nicht genügen.

4.2. Ein probabilistischer Score für prognostische Leistung

Der oben in seinen Grundzügen skizzierte Lösungsweg führt zur folgenden Methode der Quantifizierung prognostischer Leistung:

Je geringer der Prognose-Fehler — je geringer also die in der beschriebenen Rateprozedur zutage tretende und beseitigte Fehleinschätzung — umso

höher die prognostische Leistung. Der Prognose-Fehler entspricht der Summe aus zwei Teilsummen (a, b):

a) Summe aus den subjektiven Wahrscheinlichkeiten bei den sich als falsch herausstellenden Prognosen.

b) Summe aus den „Komplementär-Werten" — den Ergänzungen der subjektiven Wahrscheinlichkeitswerte auf 1 — bei den sich als richtig herausstellenden Prognosen.

Dieser Score orientiert sich also in der Behandlung richtiger und falscher Prognosen sehr direkt an dem im Konzept der „Wahrscheinlichkeit" angelegten, „dualistischen" Prinzip der Komplementarität. Wir wollen ihn daher — und in Unterscheidung zu einem „informationalen" Score, der die prognostische (Fehl-)Leistung in bit ausdrückt (s. Abschnitt 5!) — als „probabilistisch" bezeichnen.

In Tabelle 2 sind oben die fiktiven Ratetest-Daten zweier Rater (linke und rechte Spalte) beim Erraten der Buchstabenfolge DER angeführt. Der „linke" und der „rechte" Rater machen gleich viele Fehler an den gleichen Stellen, unterschiedlich sind nur die subjektiven Wahrscheinlichkeiten beim mittleren Buchstaben E. Darunter finden sich die unseren Überlegungen entsprechenden Umformungen, die aber nur die *richtigen* Prognosen betreffen: Bei diesen wird der jeweilige Wahrscheinlichkeitswert auf 1 (bzw. auf 100%) ergänzt. Diese Werte sind nun einfach zu addieren und die Summe entspricht dem Leistungsscore, wobei höhere Werte für schlechtere Leistung stehen. In unserem Beispiel war der „linke" Rater der schlechtere — mehr

Tabelle 2

Oben die fiktiven Ratedaten zweier verschiedener Personen bei einem „Text", darunter die aus diesen Daten berechneten Prognose-Fehler und — ganz unten — die nach Weltner (1970, S. 70) berechnete subjektive Information

„Text"	D	E	R		D	E	R
geratene Buchstaben und subj. Wahrscheinlichkeiten (in %)	D 10	A 20 I 30 E 40	R 30		D 10	A 5 I 5 E 95	R 30
die einzelnen Prognose-Fehler	.90	.20 .30 .60	.70		.90	.05 .05 .05	.70
Prognose-Fehler gemittelt		0,90000		≠		0,58333	
$\overline{H}$ nach Weltner		1,91333		=		1,91333	

noch als der rechte war er sich bei den Fehlprognosen zu sicher und bei den richtigen Prognosen zu unsicher gewesen.

Herkömmliche Informationsmaße können diesen Leistungsunterschied nicht erfassen. Stellvertretend für diese Maße ist in Tabelle 2 unten der Wert angegeben, der sich aus der Berechnung nach der folgenden, von Weltner (1970, S. 70) vorgeschlagenen Formel ergibt:

$$H = 0{,}27 + 4{,}93 \cdot C \quad \text{mit } C = \frac{\text{n der Fehler bei der 1. Prognose}}{\text{n der Rateplätze}}$$

Diese Formel hat Weltner für ein mit geringem Aufwand durchführbares Rateverfahren — pro Rateplatz bzw. pro Zeichen wird nur einmal geraten, dann wird das richtige Zeichen präsentiert — entwickelt. Sie ist nicht unmittelbar informationstheoretisch begründet, sondern gründet sich auf die von Weltner empirisch ermittelte Beziehung zwischen den Rate-Daten dieses Sparverfahrens und den Ergebnissen des Shannon-Verfahrens; es handelt sich also um eine Formel, welche es erlaubt, aus den im Sparverfahren gewonnenen Daten regressionsstatistisch „vorherzusagen" (siehe Abschnitt 2!), welches Resultat das aufwendigere, aber informationstheoretisch begründete Shannon-Verfahren erbringen würde.

Im folgenden soll das Ergebnis dieser Überlegungen zur Quantifizierung des Prognose-Fehlers etwas präziser und differenzierter gefaßt werden.

Bezogen auf den einzelnen Rateplatz j kann der *P*rognose-*F*ehler „*PF*" wie folgt berechnet werden:

$$PF = (1 - p_j) + \sum_{i=1}^{f'} p_i \tag{1}$$

p_j = Wettquotient bei richtiger Vorhersage
p_i = Wettquotient bei falscher Vorhersage
f' = Anzahl falscher Vorhersagen auf Rateplatz j

Für Ereignis*folgen* (Anzahl der Rateplätze > 1) gilt dann

$$PF = \sum_{j=1}^{r} (1 - p_j) + \sum_{i=1}^{f} p_i \tag{2}$$

p_j und p_i wie unter (1)
r = Anzahl aller richtigen Vorhersagen
f = Anzahl aller falschen Vorhersagen

Die in einer Ereignisfolge mit n Ereignissen bzw. n Rateplätzen beseitigte Fehleinschätzung beträgt daher im Durchschnitt (pro Rateplatz)

$$\overline{PF} = \frac{\Sigma_{j=1}^{r} (1 - p_j) + \Sigma_{j=1}^{f} p_i}{n} \tag{3}$$

r, f, p_j und p_i wie unter (2)
n = Anzahl der Rateplätze
$\overline{PF}$ = die im Durchschnitt (pro Rateplatz) beseitigte Fehleinschätzung

(3) ist die mathematisch einfache Formel, nach der man bei der Berechnung prognostischer Leistung normalerweise vorgehen wird.

Will man neben den (aufsummierten) Wahrscheinlichkeitswerten auch die zweite relevante Dimension — die Anzahl nötiger Versuche bzw. Vorhersagen „V" — eigens ausgewiesen sehen bzw. ihren Anteil am Score $\overline{PF}$ transparent machen, so bieten sich zwei Möglichkeiten an. Entweder man bildet das arithmetische Mittel der Werte $(1 - p_r)$ sowie das arithmetische Mittel der Werte p_f und schreibt

$$\overline{PF} = \frac{\Sigma_{j=1}^{r} V_r \cdot \overline{(1 - p_j)} + \Sigma_{i=1}^{f} V_f \cdot \overline{p}_i}{n} \tag{4}$$

r, f, p_j, p_i, n und $\overline{PF}$ wie unter (3)
V_r = Anzahl richtiger Vorhersagen
V_f = Anzahl falscher Vorhersagen

Oder man begreift den mittleren Prognose-Fehler als Produkt aus der pro Rateplatz nötigen Anzahl von Vorhersagen mit der pro Vorhersage beseitigten Fehleinschätzung — eine Formulierung, die sich auch für die Instruktion der Rateperson anbietet. Die entsprechende Formel:

$$\overline{PF} = \frac{V}{n} \cdot \frac{\Sigma_{i=1}^{r} (1 - p_j) + \Sigma_{i=1}^{f} p_i}{V} \tag{5}$$

r, f, p_j, p_i, n und $\overline{PF}$ wie unter (3)
V = Anzahl der (richtigen und falschen) Vorhersagen

Die Formeln (3) bis (5) führen zu denselben Ergebnissen, und man muß aus (5) nur V wegkürzen, um wieder zu (3) zu gelangen.

Welche Extremwerte kann der auf diese Weise berechnete Prognosefehler annehmen?

Ein *unterer Extremwert* von Null (keinerlei Fehleinschätzung, theoretisches Maximum der Prognoseleistung!) ergibt sich klarerweise dann, wenn der Rater jedes Ereignis auf Anhieb richtig und mit völliger subjektiver Gewißheit („p = 100%") vorhersagt.

Ein *oberer Extremwert:* Der Prognosefehler ist pro Rateplatz identisch mit der Anzahl möglicher Ereignisse. Dieses Ergebnis erhalten wir angesichts eines Raters, der erstens alle Möglichkeiten einer Fehlprognose ausschöpft (er tippt auf „A", „B", „C" ... bis „Y", bevor er „Z" errät) und der sich zweitens bei den Fehlprognosen jeweils völlig sicher ist („p = 100%"), während er ausgerechnet beim letzten und einzig richtigen Tip die subjekti-

ve Wahrscheinlichkeit mit „0%" angibt. (Er schließt also beim Treffer völlig aus, daß das, was er soeben prognostiziert hat, an dieser Stelle auch nur als Möglichkeit in Betracht kommt. Laut Instruktion ist aber dasjenige Ereignis anzugeben, welches nach seiner Einschätzung von den verbleibenden Möglichkeiten noch am ehesten in Betracht kommt. Dieser obere Extremwert setzt demnach bereits ein logisch inkonsistentes, aber konsequent durchgehaltenes Rateverhalten voraus.)

Wenn die Rateperson, aus welchen Gründen auch immer, überhaupt nie zur richtigen Prognose vorstößt, wird ihr Prognosefehler (beim betreffenden Rateplatz) unberechenbar, weil es nicht zum abschließenden positiven Feedback kommt.

4.3. Eine erste Erprobung und eine spekulative Annahme

Die Erprobung des Instrumentes an einer „echten" Versuchsperson sollte erste Anhaltspunkte über den nötigen Zeitaufwand für die Rateprozedur und die „Verständlichkeit" der Instruktion liefern. Vor allem aber sollte sie einen Datensatz liefern, der eine Darstellung unterschiedlicher Resultate der hier — und insbesondere unter 5! — vorgeschlagenen Auswertungsmethoden mit den herkömmlichen Methoden (von Shannon, von Weltner) erlaubt. Der Vergleich mit diesen anderen Verfahren wird dadurch vereinfacht, daß unsere Methode all jene Daten mitproduziert, welche für die anderen Auswertungen gebraucht werden.

Die *Instruktion* bestand aus der vereinfacht angeschriebenen Formel 5 (siehe 4.2.!), einer mündlichen Erläuterung der Formel und einem Probedurchgang an den ersten beiden Wörtern irgendeines Satzes. „Wortende" war genauso zu erraten wie ein Buchstabe, die Interpunktion wurde vom Versuchsleiter beigesteuert. Der Versuchsperson wurde geraten, auf einem ersten Zettel die pro Rateplatz falsch geratenen Buchstaben zu notieren, um die Wiederholung eines falschen Buchstabens zu vermeiden. Auf einem zweiten Zettel sollte sie die richtigen Buchstaben aneinanderreihen. Vor den Augen der Versuchsperson entwickelte sich also Buchstabe für Buchstabe der Text und war, soweit bisher erraten, ständig präsent. Und zwar in Blockschrift, mit Umlauten und mit zwei S anstelle eines ß.

Die *Versuchsperson,* eine 41jährige Sprachwissenschaftlerin, wurde darüber informiert, daß die *Textprobe* von dem Naturphilosophen Moritz Schlick und aus dem Jahre 1931 stammt. Dann wurde ein Absatz aus diesem Werk (S. 17f.) vorgelesen; nur der letzte Satz dieses Absatzes — hier in BLOCKBUCHSTABEN — war dann zu erraten:

> Nachdem es uns gelungen ist, eine Funktion zu finden, welche eine Menge von Beobachtungsresultaten befriedigend miteinander verbindet, sind

wir im allgemeinen noch keineswegs zufrieden, auch dann nicht, wenn die gefundene Funktion einen sehr einfachen Bau hat; sondern nun kommt erst die Hauptsache, die unsere bisherigen Betrachtungen noch nicht berührt hat: wir sehen nämlich zu, ob die erhaltene Formel nun auch solche Beobachtungen richtig darstellt, die wir zur Gewinnung der Formel noch nicht benutzt hatten. Für den Physiker als Erforscher der Wirklichkeit ist es das einzig Wichtige, das schlechthin Entscheidende und Wesentliche, daß die aus irgendwelchen Daten abgeleiteten Gleichungen sich nun auch für neue Daten bewähren. Erst wenn dies der Fall ist, hält er seine Formel für ein Naturgesetz. Mit anderen Worten: DAS WAHRE KRITERIUM DER GESETZMÄSSIGKEIT, DAS WESENTLICHE MERKMAL DER KAUSALITÄT IST DAS EINTREFFEN VON VORAUSSAGEN.

Das *Ergebnis:* Für die insgesamt 115 Rateplätze wurden 213 Vorhersagen benötigt, wobei der Prognosefehler in Summe 55,49 ausmachte. Nach Formel (5) entspricht der mittlere Prognose-Fehler $\overline{PF}$ dem Produkt aus der pro Rateplatz nötigen Anzahl von Vorhersagen (= 1,85) mit der pro Vorhersage beseitigten Fehleinschätzung (= 0,26): $\overline{PF}$ = 0,48.

Die Rateperson hat also im probabilistischen Score besser abgeschnitten als der bessere unserer beiden fiktiven Rater in Tabelle 2. Eine Komponente dieses Leistungsunterschiedes — die relative Häufigkeit der auf Anhieb richtigen Prognosen — erfaßt übrigens auch der nach Weltner berechnete Score: Die fiktiven Rater erzielten einen Score von 1,9 bit, die wirkliche Rateperson blieb unter 1,4 bit.

Interessanter erscheint ein Blick auf die im Anhang wiedergegebenen Ratetest-Daten. Diese verraten einiges über die Ratestrategie und Hypothesenbildung des Raters, das herkömmlichen Ratetests verborgen bleibt: Offenbar werden Hypothesen über *möglichst umfassende* Superzeichen (Satz, Satzteil, ...) gebildet, also über größere Kontexte, in welche das jeweils zu erratende Zeichen „passen" könnte. Erst wenn diese weittragenden Pfeile verschossen sind, greift man auf weniger weitreichende Überlegungen bezüglich (bedingter) Buchstaben-Häufigkeit zurück; die Wahrscheinlichkeitsangaben werden plötzlich — eingedenk der vielen möglichen Alternativen — eher „vorsichtig"; höhere Werte nehmen sie gegebenenfalls wieder an, wenn das Repertoire der noch in Frage kommenden Buchstaben sich dem Ende zuneigt und/oder zum „Ausprobieren" eines neuen Superzeichens inspiriert. (Zum Begriff des „Superzeichens" siehe Lánsky & Polák, 1977.)

Künftige Untersuchungen könnten die Hypothese prüfen, daß diese Strategie — samt dem in den Augen des (Versuchsleiters und) Beobachters vielleicht übergroßen Vertrauen auf die zuerst ins Auge gefaßten Superzeichen

— charakteristisch ist für die Bildung und Modifikation von Plausibilitäts- und Wahrscheinlichkeitsurteilen unter Entscheidungszwang. Und sie könnten prüfen, ob diese Strategie einem guten Abschneiden hinsichtlich „prognostischer Leistung" sogar besonders förderlich ist. Wenn ja, dann ist es eher der besserwisserische und aus der Distanz Beobachtende, der ein „übergroßes Vertrauen" in seine eigenen Maßstäbe setzt und nach diesen das Prognose-Verhalten anderer als zu „(un)vorsichtig", zu konservativ oder zu vorschnell qualifiziert.

Ein Beispiel: Riedl (1980) analysierte die Änderungen in der subjektiven Sicherheit von Probanden in zwei Münzwurfexperimenten. In einem der beiden Experimente schwindelten die Versuchsleiter — die verwendete Münze hatte zwei gleiche Seiten; für jeden von sechs aufeinanderfolgenden Würfen mit naturgemäß gleichem Ergebnis wurde erhoben, wieviele der Probanden noch an Zufall bzw. wieviele schon an Absicht glaubten. Die Ergebnisse interpretiert Riedl so: „Der Gewißheitsgrad wächst /.../ bei der Entdeckung einer Gesetzmäßigkeit rascher, bei einer Kette von Zufallsereignissen ungleich langsamer als das vernünftigerweise geschehen kann" (S. 63). Andererseits: Die Würfelergebnisse für sich allein erlauben den Probanden eigentlich überhaupt keine Schlüsse in der Frage „Zufall" versus „Absicht"; und jedes „externe" Wissen, das sie aktivieren können, ist asymmetrisch zugunsten positiver Hypothesen (die eine „Regularität" oder einen „Effekt" unterstellen) im Vergleich zur „Null-Hypothese", die jede systematische Abweichung vom „Zufall" in Abrede stellt. (Zur Illustration dieser Asymmetrie: Jemanden beim Mogeln zu ertappen bestätigt positiv den Verdacht der Mogelei; nichts von Mogeleien zu bemerken bestätigt dagegen noch lange nicht die Unschuldsvermutung bzw. die Null-Hypothese.)

Daß gerade die von Riedl analysierten Experimente als Beispiel angeführt wurden, hat einen weiteren Grund: Sie sind auch ein Beispiel für Untersuchungen des sogenannten „induktiven Denkens" — und Beispiel für verwandte Aufgaben vom Typus des „Zahlenreihen-Fortsetzens" —, in denen unser neues Analyse-Instrument sehr viel differenziertere Diagnosen leisten könnte.

4.4. Eine vorläufige Bewertung

Im skizzierten Verfahren wird neben der Häufigkeit von Prognosefehlern auch erhoben, mit welcher Sicherheit oder Wahrscheinlichkeit der Rater erwartet, daß das von ihm vorhergesagte Ereignis eintreffen wird bzw. daß er mit seiner Prognose recht behalten wird. Diese subjektiven Wahrscheinlichkeiten werden dann — in der oben entwickelten Art und Weise — im Sinne von *Leistungsdaten* behandelt und mit den anderen Leistungsdaten („Trefferquoten") zu einem Leistungs-Index synthetisiert. Was sind die Vorzüge und was die Nachteile dieses Prinzips und des probabilistischen Scores?

Indem das Verfahren eine Facette „subjektiver (Un)Sicherheit" und prognostischer Leistung erfaßt, für welche die herkömmlichen Ratetests blind sind, ist es — in diesem trivialen Sinne — „valider" als diese. Daß es auch

„empfindlicher" mißt als diese, nämlich bei gleicher Anzahl und Verteilung der Ratefehler noch Leistungsunterschiede registriert, wurde ausführlich dargestellt. Und das Studium systematischer Änderungen der subjektiven Wahrscheinlichkeit kann zu neuen Gesetzeshypothesen über unsere Entscheidungs-Strategien inspirieren (vgl. 4.3.!).

Und wie steht es mit dem Entwicklungsziel der „universellen" oder zumindest „vielseitigen" Anwendbarkeit? Der probabilistische Score ist zumindest im Prinzip — sofern es sich eben arrangieren läßt, daß das Feedback „dosiert" wird (siehe nächsten Absatz) — bei Wetterprognosen, Wirtschaftsprognosen und Buchstabenprognosen ebenso einsetzbar wie für die Theoriebewertung oder die Wissensdiagnose. Erfolgversprechend erscheint auch eine Anwendung in der EEG-Forschung. Dort interessiert man sich zunehmend (vgl. Korunka et al., 1990, S. 52 f.) für gesetzmäßige Beziehungen zwischen einerseits Gleichspannungsänderungen im EEG und andererseits der „erlebten Auftrittswahrscheinlichkeit" von Reizen bzw. dem „Ausmaß der Reduktion der Unsicherheit" durch Reize. (Wobei der Messung der Gleichspannungsänderungen sehr viel mehr Aufmerksamkeit gewidmet wird als der Messung und Operationalisierung der „subjektiven Wahrscheinlichkeit", „Überraschung" und „Unsicherheit".) Ein Grund für diese vielseitigen Einsatzmöglichkeiten des Scores ist wohl darin zu sehen, daß er keine theoretischen Annahmen über bestimmte „subjektive" oder „objektive" Wahrscheinlichkeitsverteilungen voraussetzt. Im Prinzip setzt er nicht einmal voraus, daß dem Rater, dem „Rückmelder" oder sonst irgend jemandem der Umfang des Repertoires (die Größe des Alphabets, die Anzahl möglicher Zustände oder Ereignisse, ...) bekannt ist, geschweige denn, daß irgend jemand schon zum Zeitpunkt der Prognose Bescheid weiß über das nächstfolgende Ereignis aus diesem Repertoire.

Die Fortschritte hinsichtlich dieser Entwicklungsziele (bzw. die am Ende von Kapitel 3 formulierten Notwendigkeiten) standen allerdings der Realisierung anderer Wünsche im Wege:

Angepeilt worden war ein Verfahren, welches auch insofern „universell" anwendbar ist, als es auf keine spezielle Form des Feedbacks angewiesen ist. Dieses (vom Zweitautor von Anfang an als unrealistisch beurteilte) Entwicklungsziel wurde verfehlt: In unserem Verfahren wird dem Rater die Lösung solange vorenthalten, bis er selber draufkommt. Dies ist zwar im Experiment immer machbar, aber irgendwie künstlich im Vergleich zum Normalfall: Daß seine Prognose falsch war, und was die richtige Prognose gewesen wäre, das erfährt der Meteorologe und der Meinungsforscher gewöhnlich „in einem".

Damit wurde auch die Hoffnung auf eine „zügig" bzw. mit geringem Aufwand durchführbare Rateprozedur zunichte gemacht. Hätte man auch im Falle einer Fehlprognose mit *einem* Feedback das Auslangen finden kön-

nen, wäre der durch die zusätzlichen Wahrscheinlichkeitsangaben entstandene Mehraufwand mehr als kompensiert worden.

Die von „außen" zu erwartende Kritik könnte u. a. vom jeweiligen — je nach Theoretiker wechselnden und bei den (anderen, „praktischen") Entscheidern vermuteten — *Verständnis von Wahrscheinlichkeit* bestimmt sein.

Nun gehen wir aber davon aus, daß der Proband in unserem Verfahren skaliert, wie *plausibel* ihm das Eintreffen eines Ereignisses und damit seiner Prognose erscheint, und daß es für die Berechtigung einer darauf basierenden Leistungsmessung irrelevant ist, ob, wie und in welchem Ausmaß sich der Rater dabei auf eine (mehr oder weniger intuitive) Statistik stützt. Wie immer der Theoretiker zur Unterscheidung zwischen „frequentistischer Wahrscheinlichkeit" und „Aussagenwahrscheinlichkeit" steht, und wo immer er sich in dem von Smithson (1989, S. 64 f.) skizzierten Spektrum mehr oder weniger „subjektivistischer" Konzepte von Wahrscheinlichkeit ansiedelt: Daß es sich bei den Vorhersagen und Wahrscheinlichkeitsangaben unserer Rater um *subjektive* Einschätzungen handelt, dürfte kaum zu bestreiten sein.

Und falls man doch streiten will: Eigentlich sind wir alle Rater. Die Kenntnis „objektiver Wahrscheinlichkeiten" kann keiner für sich beanspruchen,

— wenn es um synthetische Urteile geht (und nicht um die logischen apriori-Wahrscheinlichkeiten angesichts ideal gedachter Würfel, Münzen etc.),
— wenn er das Induktionsproblem in seiner ganzen Radikalität begriffen hat, und
— wenn er „Wahrscheinlichkeit" auf noch nicht bekannte Instanzen bezieht, in ihr also mehr sieht als die bisher — „objektiv" oder „intuitiv" — registrierte relative Häufigkeit.

Subjektive Wahrscheinlichkeiten sind als „internal representations of objective probabilities" zu betrachten — was z. B. Rachlin (1989, S. 25) tut oder zumindest dem „Kognitivisten" unterstellt — bedeutet dagegen, den „objektiven" Wahrscheinlichkeiten eine „externe", vom jeweiligen Subjekt unabhängige und vom Induktionsproblem unangefochtene Realität zuzubilligen.

Ernster zu nehmen ist der Einwand, daß der Rater in unserem Verfahren gleichsam gezwungen wird, etwas auf das Hundertstel genau anzugeben, was er vielleicht nur sehr vage oder „nur" als Vertrauensbereich im Kopf hat. Einige der bei Smithson (1989) erwähnten Untersuchungen weisen darauf hin, daß wir uns zumindest dort, wo wir nicht als „Experte" urteilen und entscheiden, meist mit derlei „groben" Einschätzungen begnügen; und psychologische Tests sehen meistens von einer „zu feinen" (mehr als 9- oder 10stufigen) Skalierung subjektiver Dimensionen (z. B. subjektiver Sicherheit) ab, weil diese „scheingenau" wäre oder die Testperson überfordere.

Eine kurze Rechtfertigung: Zum einen läßt sich unser probabilistischer Score ohnehin wie bisher berechnen, wenn die Rateperson eine noch feinere (z. B. Promille-)Skala oder eben eine gröbere Skala (z. B. 0, .1, .2, .3, ... 1) benutzt oder benutzen muß. Zum anderen aber könnte sich bei einer von vornherein groben Skala die sehr unterschiedlich große und — wenn sie „gefordert" wird — oft unerwartet große Expertise gar nicht artikulieren,

die potentielle Diskriminationsfähigkeit des Tests wäre ohne Not von vornherein beschnitten. Wenn die Rateperson wirklich „nur" oder zuerst einmal Vertrauensbereiche im Kopf hat — was im übrigen keineswegs mit Vagheit zu tun haben muß — wird es ihr ein Leichtes sein, die Mitte dieses Bereiches als die intuitiv wie auch testtheoretisch beste Punkt-Schätzung des „wahren Wertes" preiszugeben. Jedenfalls hat sich bislang keine der Ratepersonen (eines soeben abgeschlossenen Experimentes von Kurat, 1991) über eine zu feine Skalierung beklagt. Was wohl auch damit zu tun hat, daß wir alle mit der Prozentskala im allgemeinen und als „Wahrscheinlichkeits-Skala" im besonderen bestens vertraut sind.

Der probabilistische Score ist ein *Durchschnittsmaß,* und gerade deswegen ist den Probanden leichter als bei jedem Verteilungsmaß klarzumachen, auf welche Weise der Score aus ihren Äußerungen (Prognosen, Wahrscheinlichkeitsangaben) berechnet wird. Das ist von größter Bedeutung, damit der Proband in einem solchen „Leistungs-Experiment" seine Verarbeitungs- und Beantwortungsstrategien an genau jenen Kriterien ausrichten kann, nach denen seine Leistung bemessen wird.

Es hat aber auch einen Nachteil: Gerade die Prüfung der wirklich „informational" formulierten Theorien und Hypothesen über menschliche Informationsverarbeitung verlangt einen informationalen Maßstab (z. B. bit oder bit/sec). Wer die *Ersparnis an (Zeit und) „Unsicherheit" („Unsicherheit" im informationstheoretischen Sinne)* bestimmen will — sei es, weil er die „Informations(konstanz)hypothese" prüft (Fenk, 1986), sei es, weil er sich für eine informationale Messung des „Transfer"-Effektes interessiert, den ein Unterricht A auf den Unterricht B hat (Frank, 1990 b) — ist auf einen solchen Maßstab angewiesen. Und, nicht ganz unwichtig: Ein wesentlicher Teil empirischer Ergebnisse z. B. über die Perzeptions- und Gedächtniskapazität ist informational beschrieben (vgl. Mittenecker & Raab, 1973), so daß künftige Untersuchungen schon aus Gründen der Vergleichbarkeit ebensolche Beschreibungen ermöglichen sollten.

Ein informationaler Score wollte sich aber aus dem skizzierten Lösungsprinzip keineswegs so direkt und zwanglos ergeben wie der probabilistische Score. An dieser Stelle muß der für die bisherigen Ausführungen verantwortliche Erstautor kapitulieren und alles weitere der vom Zweitautor vertretenen Kybernetik überlassen[1]).

---

1) Angesichts dieser unterschiedlichen Zuständigkeiten sowie der Überlänge dieses Aufsatzes war auch die Publikation in zwei aufeinanderfolgenden, eigenständigen Artikeln überlegenswert. Den Ausschlag für die vorliegende Variante gab der Umstand, daß sie — zum Beispiel wegen der Bezugnahme auf *ein* Set von Daten und Beispielen — die insgesamt „ökonomischere" Lösung ist.

5. Ein informationaler Score für prognostische Leistung

Der informationale Score IS unterscheidet sich von dem probabilistischen nur in der Auswertung, die Rateprozedur ist hingegen für beide Scores dieselbe: Der Proband versucht, ein Superzeichen S zu erraten, indem er diverse Zeichen z aus einem bestimmten Zeichenrepertoire Z nennt. Nach jedem Schritt wird er vom Testleiter darüber informiert, ob das eben genannte Zeichen richtig oder falsch war. Um allerdings den Ratetest regulär durchführen zu können, muß die Vp zumindest über die Art der zu erratenden Ereignisse informiert sein.

Nach dem Abschluß des Ratetests könnte das Protokoll des Versuchsleiters in allgemeiner Form etwa folgendermaßen ausschauen (vgl. z. B. mit Tab. 2):

Tabelle 3
Das Rateprotokoll in Form einer Zeichenmatrix

S	z_1	z_2	...	z_k	
	z_{11}	z_{21}	...	z_{k1}	k = Länge von S
	z_{12}	z_{22}	...	z_{k2}	r_i = Ratezahl pro $z_i \in Z$
	⋮	⋮		⋮	
	$\underline{z_{1r_1}}$	$\underline{z_{2r_2}}$	...	$\underline{z_{kr_k}}$	Unterstrichen sind die letzten und richtig erratenen Zeichen
r_i	r_1	r_2	...	r_k	

Hier stellen die Buchstaben z_{ij} alle geratenen Zeichen und r_i die Zahl der Rateversuche pro Zeichen z_i dar.

Aus der Tabelle 3 läßt sich nun erkennen, daß der Proband versucht hat, das Superzeichen bzw. die horizontale Zeichenfolge z_1 bis z_k aufgrund bedingter Wahrscheinlichkeiten p_{ij} zu erraten, oder daß er bei jedem Rateschritt (vertikal) an irgendein reduziertes Zeichenrepertoire

$$z \subset Z$$

mit der Mächtigkeit

$$R_{ij} = | z_{ij} | < | Z |$$

gedacht hatte (z. B. die Vokale aus dem gesamten Alphabet).

Dementsprechend könnte man die Tabelle 3 in Tabelle 4 umwandeln.

Tabelle 4
Das Rateprotokoll in Form einer Mächtigkeits- bzw. Wahrscheinlichkeitsmatrix

S	z_1	z_2	...	z_k
	R_{11}	R_{21}	...	R_{k1}
	R_{12}	R_{22}	...	R_{k2}
	⋮	⋮		⋮
	R_{1r_1}	R_{2r_2}	...	R_{kr_k}
r_i	r_1	r_2	...	r_k

S	z_1	z_2	...	z_k
	p_{11}	p_{21}	...	p_{k1}
	p_{12}	p_{22}	...	p_{k2}
	⋮	⋮		⋮
	p_{1r_1}	p_{2r_2}	...	p_{kr_k}
r_i	r_1	r_2	...	r_k

Für die Berechnung der subjektiven Information (Entropie H) aus solchen Ratespiel-Daten gibt es nun verschiedene Methoden (Formeln). Die bekanntesten stammen von Shannon (1951):

$$H_1 = \frac{1}{k} \sum_{i=1}^{k} \mathrm{ld}\, r_i \qquad \text{/bit/} \quad 1.$$

$$H_2 = \sum_{i=1}^{k} \frac{r_i}{N} \mathrm{ld} \frac{N}{r_i} =$$

$$= \mathrm{ld}\, N - \frac{1}{N} \sum_{i=1}^{k} r_i \,\mathrm{ld}\, r_i \quad \text{mit} \quad N = \sum_{i=1}^{k} r_i \qquad \text{/bit/} \quad 2.$$

$$H_3 = \frac{1}{k} \sum_{i=1}^{k} r_i \,\mathrm{ld}\, r_i - (r_i - 1)\,\mathrm{ld}\,(r_i - 1) =$$

$$= \sum_{i=1}^{n} (p_i - p_{i+1})\, i\, \mathrm{ld}\, i \quad \text{mit} \quad \sum_{i=1}^{n} p_i = 1 \qquad \text{/bit/} \quad 3.$$

Die erste Formel bestimmt hier die untere, die zweite die obere und die dritte die mittlere bzw. die „höhere untere" Schranke.

Darüber hinaus gibt es zur Berechnung der subjektiven Entropie noch zwei Formeln von Weltner, und zwar für das sogenannte Sparverfahren — pro Rateplatz wird nur einmal geraten — bei längeren Superzeichen (Näheres in Kap. 4.2.):

$$H_4 = \frac{1}{k}(3{,}9.F - 0{,}08.k) = 3{,}9\,\frac{F}{k} - 0{,}08 \qquad \text{/bit/} \quad 4.$$

$$H_5 = 4{,}93\,\frac{F}{k} + 0{,}27 \qquad \text{/bit/} \quad 5.$$

F sind hier alle falsch geratenen Zeichen, *k* ist wiederum die Gesamtlänge des Superzeichens. Die Differenz zwischen 5. (in Weltner 1970) und 4. (in Frank, 1971, S. 83), also

$$H_5 - H_4 = 1{,}03(F/k) + 0{,}35$$

ist möglicherweise als Anpassung an neuere empirische Ergebnisse zu verstehen.

Aus den Gleichungen 1. bis 3. läßt sich nun erkennen, daß bei der Berechnung der Zeichenentropien lediglich die Verteilung der Ratezahlen r_1 bis r_k berücksichtigt wurde, also eigentlich nur die letzte Zeile in den Tabellen 3 und 4, während man vertikal jeweils eine Gleichverteilung

$$R_i = r_i \text{ bzw. } p_i = 1/r_i$$

angenommen hatte.

Dies entspricht allerdings nicht der anfangs gestellten Bedingung, daß nämlich der Proband jedes Zeichen pro Spalte nur einmal nennen darf und daß dieses Zeichen somit aus dem Spiel ausscheidet. Setzt man also diese Bedingung wiederum in die drei Shannonschen Gleichungen ein, so erhält man:

$$H_1' = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{r_i} \operatorname{ld}(r_i!) \qquad \text{/bit/} \quad 6.$$

$$H_2' = \frac{1}{k} \sum_{i=1}^{k} \operatorname{ld} N_i - \frac{1}{N_i} \sum_{j=1}^{r_i} j \operatorname{ld} j \qquad \text{mit} \quad N_i = \frac{(r_i + 1).r_i}{2} \qquad \text{/bit/} \quad 7.$$

$$H_3' = H_1 \qquad \text{/bit/} \quad 8.$$

So wie die Shannonschen haben auch diese Formeln eine allgemeine Geltung. Um die Entropien laut Gleichung 1. bis 3. und 6. bis 7. schneller und mit einfacheren Mitteln bestimmen zu können, wurden diese für die Ratezahlen 1 bis 20 berechnet und tabelliert (Tab. III im Anhang). Es läßt sich nun auch hier feststellen, daß die Ergebnisse aus den Gleichungen 6. und 7. aufgrund der Bedingung

$$(r_i!)^{1/r_i} < r_i$$

etwas niedriger ausfallen als jene aus Gleichung 1. bis 3. Die Gleichung 6. stellt hier wiederum die untere, die Gleichung 7. allerdings die mittlere und die Gleichung 8. die obere Informationsschranke dar.

In Wirklichkeit (s. Kapitel 4.1.) kann der Proband bei jedem Versuch, das richtige Zeichen zu erraten, an verschieden große Zeichenrepertoires bzw. Wahrscheinlichkeiten denken. Ist er nun imstande, diese mit jedem Zeichen

dem Versuchsleiter mitzuteilen, dann könnte das Protokoll etwa die Form von Tabelle 4 annehmen. Die ersten Zeichenrepertoires bzw. Wahrscheinlichkeiten pro Spalte gelten hier für alle falsch geratenen Zeichen, die letzten und unterstrichenen für die richtige Antwort.

Im weiteren wollen wir, so wie beim probabilistischen Score, lediglich mit den Wahrscheinlichkeiten rechnen. Nehmen wir dabei an, daß diese im gedachten Zeichenrepertoire R = 1/p gleich verteilt sind, dann können wir wiederum die drei Shannonschen Formeln verwenden, diesmal jedoch auf die gesamte Matrix erweitert:

$$H_1' = -\frac{1}{r} \sum_{i=1}^{k} \sum_{j=1}^{r_i} \mathrm{ld}\, p_{ij} \qquad \text{/bit/} \qquad 9.$$

$$H_2' = -\sum_{i=1}^{k} \sum_{j=1}^{r_i} p_{ij}\, \mathrm{ld}\, p_{ij} \qquad \text{mit} \quad \sum_{i=1}^{k} \sum_{j=1}^{r_i} p_{ij} = 1 \qquad \text{/bit/} \qquad 10.$$

$$H_3' = -\frac{1}{r} \sum_{i=1}^{k} \sum_{j=1}^{r_i} \frac{1}{p_{ij}} [p_{ij}\, \mathrm{ld}\, p_{ij} + (1-p_{ij})\, \mathrm{ld}\,(1-p_{ij})] \qquad \text{/bit/} \qquad 11.$$

mit

$$r = \sum_{i=1}^{k} r_i$$

Um zum informationalen Score zu gelangen, muß man diese Gleichungen noch um die Art der Antwort (richtig R oder falsch F) ergänzen. Dabei nimmt man so wie beim probabilistischen Score an, daß sich für die falsche Antwort mit einer niedrigeren Wahrscheinlichkeitsabschätzung bzw. für die richtige mit einer höheren Wahrscheinlichkeitsabschätzung ein kleinerer Wert ergibt als im umgekehrten Fall. Aufgrund dieser Überlegungen erhalten wir

$$p_R = p_i \text{ und } p_F = 1 - p_j$$

Auch diese Bedingungen könnten wir nun problemlos in die Gleichungen 9. bis 11. einsetzen, nur hätten wir dann wiederum drei verschiedene Ergebnisse. Zur Berechnung des informationalen Scores wählen wir daher nur die Gleichung 11.: Diese ist zwar am kompliziertesten, ermöglicht aber laut Shannon eine genauere Schätzung (als 9.) und bedarf außerdem keiner Umrechnung von Wahrscheinlichkeiten (wie 10.):

$$H_R = -\sum_{i=1}^{R} \frac{1}{p_i} [p_i\, \mathrm{ld}\, p_i + (1 - p_i)\, \mathrm{ld}\,(1 - p_i)] \qquad \text{/bit/} \qquad 12.$$

$$H_F = -\sum_{j=1}^{F} \frac{1}{1-p_j} [p_j\, \mathrm{ld}\, p_j + (1 - p_j)\, \mathrm{ld}\,(1 - p_j)] \qquad \text{/bit/} \qquad 13.$$

Das R ist hier die Summe aller richtigen und F die aller falschen Antworten in der Matrix. Der durchschnittliche informationale Score beträgt dann

$$\overline{IS} = \frac{1}{r}\,(H_R + H_F), \qquad \text{/bit/} \qquad 14.$$

wobei

$$r = \sum_{i=1}^{k} r_i = R + F \qquad 15.$$

ist.

Da man nicht völlig ausschließen kann, daß der Proband bei der Wahrscheinlichkeitsabschätzung die Extremwerte von Eins oder Null wählt, und da diese Extreme Null oder Unendlich bei der Entropie zur Folge hätten, limitieren wir sicherheitshalber den Bereich auf

$$0 < p < 1\,, \text{ konkret auf } 0{,}01 \leq p \leq 0{,}99$$

Es wird also statt Null mit 0,01 und statt Eins mit 0,99 gerechnet.

Um auch hier die Teilentropien aus den Gleichungen 9. bis 13. schneller und mit einfacheren Mitteln feststellen zu können, wurden diese für die Wahrscheinlichkeiten von Null bis Eins bzw. von 0,01 bis 0,99 berechnet und tabelliert (Tab. IV im Anhang).

Die Anwendung der Gleichungen 1. bis 5. und der Gleichung 14. demonstrieren wir nun an zwei fiktiven Beispielen mit dem Superzeichen „DER" (vgl. Tab. 2):

z	D	E	R	z	D	E	R	k = 3
p	0,1	0,2 0,3 0,4	0,3	p	0,1	0,05 0,05 0,95	0,3	
r_i	1	3	1	r_i	1	3	1	r = 5
		Links				Rechts		

Da die r_i-Verteilung in beiden Beispielen gleich ist, liefern die Gleichungen H_1 bis H_5 für „Links" und „Rechts" dasselbe Ergebnis.

Für die Berechnung nach Gleichung 1. verwenden wir die Gleichung 1'. aus der Tabelle III (Anhang):

$$H_1 = \frac{1}{3}(2.\text{ld } 1 + \text{ld } 3) = \frac{1}{3}\ 1{,}58496 = \mathbf{0{,}52832\ bit}$$

Dies ist die untere Informationsschranke. Die obere Schranke berechnet man nach Gleichung 2. und mit Hilfe der Gleichung 1'. und 2'. (Ergebnisse in Tab. III, Anhang):

$$H_2 = \text{ld } 5 - \frac{1}{5}(2.\ 1 \text{ ld } 1 + 3 \text{ ld } 3) = 2{,}32193 - \frac{1}{5}\ 4{,}75489 = \mathbf{1{,}37095\ bit}$$

Die dritte Gleichung (3.), welche die mittlere „Schranke" darstellt, berechnen wir mit Hilfe der Gleichung 2'. (Ergebnisse in Tab. III, Anhang):

$$H_3 = \frac{1}{3}[2(1 \text{ ld } 1 - 0 \text{ ld } 0) + 3 \text{ ld } 3 - 2 \text{ ld } 2] =$$

$$= \frac{1}{3}\ (4{,}75489 - 2{,}00000) = \mathbf{0{,}91830\ bit}$$

Die vierte (4.) und fünfte Gleichung (5.) läßt sich ganz einfach berechnen, wenn man die mittlere Spalte in beiden Beispielen mit F = 1 belegt (einzige falsche Antwort mit nachfolgender Bekanntgabe der richtigen):

$$H_4 = 3{,}9\frac{1}{3} - 0{,}08 = \mathbf{1{,}22000\ bit}$$

$$H_5 = 4{,}93\frac{1}{3} + 0{,}27 = \mathbf{1{,}91333\ bit}$$

In den Gleichungen 12. bis 14. für den durchschnittlichen informationalen Score $\overline{IS}$ werden nun die Wahrscheinlichkeiten p_{ij} beider DER-Matrizen eingesetzt, die unterschiedlich sind und daher zu unterschiedlichen Ergebnissen führen müssen. Auch diese Berechnung ist mit Hilfe von Tabelle IV (s. Anhang) möglich:

Links:

$$H_R = -\left[\frac{1}{0{,}1}(0{,}1 \text{ ld } 0{,}1 + 0{,}9 \text{ ld } 0{,}9) + \frac{1}{0{,}4}(0{,}4 \text{ ld } 0{,}4 + 0{,}6 \text{ ld } 0{,}6) + \right.$$
$$\left. + \frac{1}{0{,}3}(0{,}3 \text{ ld } 0{,}3 + 0{,}7 \text{ ld } 0{,}7)\right] =$$
$$= 4{,}68995 + 2{,}42738 + 2{,}93764 = \mathbf{10{,}05497}$$

$$H_F = -\left[\frac{1}{0{,}8}(0{,}2 \text{ ld } 0{,}2 + 0{,}8 \text{ ld } 0{,}8) + \frac{1}{0{,}7}(0{,}3 \text{ ld } 0{,}3 + 0{,}7 \text{ ld } 0{,}7)\right] =$$
$$= 0{,}90241 + 1{,}25899 = \mathbf{2{,}16140}$$

$$\overline{IS} = \frac{1}{5}(10{,}05497 + 2{,}16140) = \mathbf{2{,}44327\ bit}$$

Rechts:

$$H_R = -\left[\frac{1}{0{,}1}(0{,}1 \text{ ld } 0{,}1 + 0{,}9 \text{ ld } 0{,}9) + \frac{1}{0{,}95}(0{,}95 \text{ ld } 0{,}95 + 0{,}05 \text{ ld } 0{,}05) + \right.$$
$$\left. + \frac{1}{0{,}3}(0{,}3 \text{ ld } 0{,}3 + 0{,}7 \text{ ld } 0{,}7)\right] =$$
$$= 4{,}68995 + 0{,}30147 + 2{,}93764 = \mathbf{7{,}92906}$$

$$H_F = -\left[\frac{2}{0{,}95}(0{,}05 \text{ ld } 0{,}05 + 0{,}95 \text{ ld } 0{,}95)\right] = 2 \cdot 0{,}30147 = \mathbf{0{,}60294}$$

$$\overline{IS} = \frac{1}{5}(7{,}92906 + 0{,}60294) = \mathbf{1{,}70640 \text{ bit}}$$

Tabelle 5

Der nach verschiedenen Methoden berechnete Informationsgehalt in bit für zwei fiktive Ratetestdaten beim Wort „DER" und für echte Ratetestdaten bei einem längeren Satz

		DER	DER	Text
Untere Schranke	H_1 (Gl.1.)	0,52832		0,40144
Obere Schranke	H_2 (Gl.2.)	1,37095		6,11345
Mittlere Schranke	H_3 (Gl.3.)	0,91830		0,65797
Weltner I	H_4 (Gl.4.)	1,22000		0,80174
Weltner II	H_5 (Gl.5.)	1,91333		1,38461
Inf. Score	$\overline{IS}$ (Gl.14.)	2,44327	1,70640	1,35660
		Links	Rechts	

Die Berechnung von H und $\overline{IS}$ wurde also anhand des sehr kurzen Textes „DER" demonstriert, aber auch für die von der Versuchsperson produzierten Ratedaten durchgeführt. Die Ergebnisse aller informationstheoretischen Berechnungen sind in Tabelle 5 zusammengefaßt.

Die in Abschnitt 4.4. vorgenommene Bewertung des neuen Rateverfahrens gilt sinngemäß auch für den informationalen Score. Die Unterschiede: Der Prognosefehler läßt sich nun auch in bit ausdrücken. Zu diesem Zwecke mußten allerdings Implikationen und (Verteilungs-)Annahmen, welche in die Shannonsche Berechnung der Information aus Ratedaten eingehen, übernommen und Extremwerte subjektiver Wahrscheinlichkeiten (0%, 100%) als unzulässig erklärt werden.

Summary

Anticipation can be considered a precondition for efficient „learning" (in the broadest sense) and its enhancement an essential aim and criterion of success in the learning process. Due to this principal importance of anticipation the topics „subjective probability" and „goodness of fit of predictions" are dealt with in a great variety of disciplines such as the theories of induction, decision, cognition and communication.

Given this background an attempt is made to develop Shannon's guessing game technique into a more sensitive and widely applicable instrument for measuring prognostic performance. Shannon's guessing game technique as well as all related techniques are blind to the guesser's „metacognitive" knowledge of the (limited) chances of correctness of his prognoses. The suggestion is made that these subjective probabilities should also be collected and treated as weight indices: Higher probability values increase the weight of incorrect prognoses (as errors) and the weight of correct prognoses (as „hits"). The smaller the prognostic error to be eliminated in the guessing procedure, the higher the level of prognostic performance. In particular, taking into account the requirements of a psychology of „information processing" a further algorithm is introduced, which makes it possible to quantify the prognostic error in bits. In a first probe it was possible to demonstrate the enhanced sensitivity of the diagnostic instrument as well as to formulate a new hypothesis about cognitive strategies to be applied when faced with decisions under uncertainty.

Literatur

Ballstaedt, S. P. (1990). Wenn Hören und Sehen vergeht. Grenzen der audiovisuellen Integration. In D. Meutsch & B. Freund (Hrsg.), *Fernsehjournalismus und die Wissenschaften* (S. 29—46). Wiesbaden: Westdeutscher Verlag.

Ballstaedt, S. P., Molitor, S. & Mandl, H. (1989). Wissen aus Text und Bild. In J. Groebel & P. Winterhoff-Spurk (Hrsg.), *Empirische Medienpsychologie* (S. 105—133). Psychologie Verlagsunion.

Böhme, G. (1987). *Ein Ansatz zur Konzeptualisierung und Messung der Verständlichkeit von Bildern unter besonderer Berücksichtigung ingenieurwissenschaftlicher und mathematisch-informatischer Bildersprachen.* Dissertation, Universität Klagenfurt.

Coombs, C. H. (1984). Theory and experiment in psychology. In K. Pawlik (Hrsg.), *Fortschritte der Experimentalpsychologie* (S. 20—30). Berlin, Heidelberg: Springer.

Dorsch (1987). *Psychologisches Wörterbuch* (11. ergänzte Aufl.). Bern: Huber.

Fenk, A. (1980). „Ein Bild sagt mehr als tausend Worte...!?" Lernleistungsunterschiede bei optischer, akustischer und optisch-akustischer Präsentation von Lehrmaterial. *AV-Forschung, 23,* S. 3—50.

Fenk, A. (1986). Informationale Beschränkungen der Wissenserweiterung? *Zeitschrift für Experimentelle und Angewandte Psychologie, 33,* 208—253.

Fenk, A. (1988). Cognitive process and scientific progress. In H. Janig & A. Vrbinc (Hrsg.), *Report on Alps-Adria Symposium 1987* (S. 41—42). Klagenfurt: Universitätsverlag Carinthia.

Fenk, A. (1990a). Graphische Darstellung und kognitive Repräsentation. In N. Derner, C.-D. Heinze, F. Klaus & A. Melezinek (Hrsg.), *Technik und humane Daseinsgestaltung* (S. 365—371). (Referate des Int. Sympos. Ingenieurpädagogik '89). Alsbach: Leuchtturm-Verlag.

Fenk, A. (1990b). Der Informationsbeitrag von Bildern als meßbare Größe. In A. Haug, A. Melezinek & V. K. Schutz (Hrsg.), *Ingenieurausbildung 2000 — Engineering Education 2000* (S. 182—184). (Proceedings des Int. Sympos. Ingenieurpädagogik '90). Alsbach: Leuchtturm-Verlag.

Fenk, A. (1991). Sprachbilder — Bildsprachen. In H. Kautschitsch & W. Metzler (Hrsg.), *Anschauliche und experimentelle Mathematik I* (S. 9—26). Schriftenreihe Didaktik der Mathematik, Bd. 20. Wien: Hölder—Pichler—Tempsky; Stuttgart: B. G. Teubner.

Fenk, A. & Lechner, B. (1986). Das nachrichtentechnische Konzept der limitierten Kanalkapazität in Pädagogik und Psychologie. *grkg/Humankybernetik, 27* (3), 99—104.

Frank, H. (1967). Über den Informationsgehalt von Bildern. *grkg/Humankybernetik, 8* (1), 23—31.

Frank, H. (1971). *Kybernetische Grundlagen der Pädagogik.* Stuttgart: Kohlhammer.

Frank, H. (1990a). Bedenken gegen einen radikalen Informationskonstruktivismus. *grkg/Humankybernetik, 31* (1), 15—25.

Frank, H. (1990b). Rechtfertigung von Rechner- und Sprachmodellen durch die kybernetische Transfertheorie. *grkg/Humankybernetik, 31* (3), 100—108.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts.* New York, London: John Wiley & Sons.

Goodman, N. (1973). *Fact, fiction and forecast.* University of London, 1955 (3. Aufl.). The Bobbs-Merrill-Company.

Greeno, J. G. (1970). Evaluation of statistical hypotheses using information transmitted. *Philosophy of Science, 37,* 279—293.

Heinrich, P. B. (1970). Durchführung von Rateversuchen mit Hilfe eines Rechners. *grkg/Humankybernetik 11,* 45—56.

Korunka, C., Bauer, H., Wolek, A. & Leodolter, M. (1990). Die Erwartungswelle (CNV) als Indikator neuronaler Informationsverarbeitung in Abhängigkeit von spontanen Gleichspannungsänderungen im EEG. *Zeitschrift für Experimentelle und Angewandte Psychologie, 37* (1), 52—68.

Kurat, I. (1991). *Das Substantiv als Satzzentrum in der französischen Werbesprache.* Diplomarbeit, Universität Klagenfurt.

Lansky, M. & Polak, V. (1977). *Studien zur Superierung durch Komplexbildung.* Hannover, Paderborn: Schroedel/Schöningh.

Lüer, G. (1988). Kognitive Prozesse und Augenbewegungen. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie* (S. 386—399). München, Weinheim: Psychologie Verlagsunion.

Luther, P. & Fenk, A. (1984). Wird der Wortlaut von Sätzen zwangsläufig schneller vergessen als ihr Inhalt? *Zeitschrift für Experimentelle und Angewandte Psychologie, 31* (1), 101—123.

Mandl, H. & Spada, H. (Hrsg.). (1988). *Wissenspsychologie.* München, Weinheim: Psychologie Verlagsunion.

Mittenecker, E. & Raab, E. (1973). *Informationstheorie für Psychologen.* Göttingen: Hogrefe.

Neumann, O. (1985). Die Hypothese begrenzter Kapazität und die Funktion der Aufmerksamkeit. In O. Neumann (Hrsg.), *Perspektiven der Kognitionspsychologie* (S. 185—229). Berlin, Heidelberg: Springer.

Oeser, E. (1976). *Wissenschaft und Information* (3 Bände). Wien: Oldenbourg.

Rachlin, H. (1989). *Judgment, decision, and choice.* New York: Freeman.
Reinert, G. (1985). Schemata als Grundlage der Steuerung von Blickbewegungen bei der Bildverarbeitung. In O. Neumann (Hrsg.), *Perspektiven der Kognitionspsychologie* (S. 113—145). Berlin, Heidelberg: Springer.
Riedl, R. (1980). *Biologie der Erkenntnis.* Berlin, Hamburg: Verlag Paul Parey.
Salmon, W. C. (1971). *Statistical explanation and statistical relevance.* University of Pittsburgh Press.
Schlick, M. (1948). Die Kausalität in der gegenwärtigen Physik. *Die Naturwissenschaften 1931, Bd. 19* (zitiert aus einem Wiederabdruck im Band „Gesetz, Kausalität und Wahrscheinlichkeit", S. 3—55). Wien: Gerold.
Shannon, C. E. (1951). Prediction and entropy of printed English. *The Bell System Technical Journal, 30,* 50—54.
Smithson, M. (1989). *Ignorance and uncertainty.* New York: Springer.
Weltner, K. (1970). *Informationstheorie und Erziehungswissenschaft.* Quickborn: Schnelle.
Westhoff, K. (1985). *Erwartungen und Entscheidungen.* Berlin, Heidelberg: Springer.

Anschrift der Verfasser: August Fenk, Universität Klagenfurt, Abteilung für Kognitionsforschung, Universitätsstraße 65—67, A-9022 Klagenfurt.

Anhang

Tabelle I
Die Ratetest-Daten

D	80
I	99, A 99
S	100
+	100
N	70, W 60
E	100, I 80, A 60
H	99
R	100
E	100
+	100
G	50, I 50, D 50, E 50, N 50, W 50, U 50, M 50, K 50
A	99, E 80, R 80
I	100
T	100
E	100
R	100
I	100
U	100
M	100
+	100
I	95, D 70
E	80
R	100
+	100

E 70, W 70, P 30, F 30, N 20, G 30
L 99, E 80
N 95, S 90
E 99
T 100
Z 100
E 100, S 99, M 95
Ä 98
S 100
S 100
I 100
G 100
K 100
E 100
I 100
T 100
+ 100
W 80, D 80
A 100
S 100
+ 100
A 40, F 40, D 40, S 40, E 30, Ä 20, V 50, I 10, G 70, J 90, N 20, M 50, K 20, L 10, R 10, T 5, U 5, W 70
A 80, E 80
N 80, S 95
E 100
N 100
T 100
L 100
I 100
C 100
H 100
E 100
+ 100
E 70, K 70, D 30, I 30, G 30, P 60, F 70, S 30, W 10, B 10, C 5, A 5, H 5, I 5, J 2, L 10, M 15
E 100
R 100
K 100
M 100
A 100
L 100
+ 100
D 90
E 100
R 100
+ 100
G 80, N 80, E 80, W 50, D 10, B 10, J 10, K 10
R 80, E 10, O 10, L 10, A 10
N 10, U 60
M 100, F 10, S 95
A 100

L 100
I 100
T 100
Ä 100
T 100
+ 100
V 70, D 80, F 50, S 10, A 10, I 80
S 100
T 100
+ 100
D 80
I 70, A 90
S 100
+ 100
E 70
R 90, I 90
N 90
B 20, Z 80, M 50, F 60, H 60, S 70, E 60, D 5, G 5, I 5, K 10, L 10, P 5, R 5, T 5
R 80
A 95, E 80
F 95
F 100
E 100
N 100
+ 100
V 80
O 100
N 100
+ 100
E 70, V 70
O 95
R 99
H 100, A 80
U 100
S 100
S 100
A 100
G 100
E 100
N 100
+ 100

Tabelle II
Die auf den einzelnen Rateplatz entfallende Zahl von Vorhersagen

1 2 11 2 3 1 1 1 1 9 3 11 1 1 11 1 1 2 1 1 1
D A S + W A H R E + K R I T E R I U M + D E R +

6 2 2 1 1 1 3 1 1 1 1 1 1 1 1 1 1 2 1 1 1
G E S E T Z M Ä S S I G K E I T +, D A S +

18 2 2 1 1 1 1 1 1 1 1 1 1 17 1 1 1 1 1 1 1 1 1 1 1 1
W E S E N T L I C H E + M E R K M A L + D E R +

8 5 2 3 1 1 1 1 1 1 1 6 1 1 1 1 2 1 1
K A U S A L I T Ä T + I S T + D A S +

1 21 15 1 2 1 1 1 1 1 1 1 1 1 1
E I N T R E F F E N + V O N +

2 1 1 2 1 1 1 1 1 1 1 1
V O R A U S S A G E N +.

Anzahl der Rateplätze: 115
Anzahl der Vorhersagen: 213

Tabelle III

r	Gl.1'.	Gl.2'.	Gl.3'.	Gl.6'.	Gl.7'.
1	0,00000	0,00000	0,00000	0,00000	0,00000
2	1,00000	2,00000	2,00000	0,50000	0,91829
3	1,58496	4,75489	2,75489	0,86165	1,45916
4	2,00000	8,00000	3,24511	1,14624	1,84644
5	2,32193	11,60964	3,60964	1,38138	2,14926
6	2,58496	15,50977	3,90013	1,58197	2,39831
7	2,80735	19,65148	4,14171	1,75703	2,61000
8	3,00000	24,00000	4,34852	1,91240	2,79420
9	3,16992	28,52932	4,52932	2,05212	2,95729
10	3,32193	33,21928	4,68996	2,17911	3,10364
11	3,45943	38,05375	4,83447	2,29550	3,23639
12	3,58496	43,01955	4,96580	2,40295	3,35787
13	3,70044	48,10572	5,08617	2,50276	3,46984
14	3,80735	53,30297	5,19725	2,59595	3,57371
15	3,90689	58,60336	5,30039	2,68334	3,67056
16	4,00000	64,00000	5,39664	2,76563	3,76129
17	4,08746	69,48687	5,48687	2,84339	3,84662
18	4,16992	75,05865	5,57178	2,91708	3,92718
19	4,24783	80,71062	5,65197	2,98713	4,00346
20	4,32193	86,43856	5,72794	3,05387	4,07589

Gl.1'.: $\mathrm{ld}\, r$

Gl.2'.: $r\, \mathrm{ld}\, r$

Gl.3'.: $r\, \mathrm{ld}\, r - (r-1)\, \mathrm{ld}\,(r-1)$

Gl.6'.: $\frac{1}{r}\, \mathrm{ld}\,(r!)$

Gl.7'.: $\mathrm{ld}\, N - \frac{1}{N} \sum_{j=1}^{r} j\, \mathrm{ld}\, j \qquad \text{mit} \qquad N = \frac{(r+1).r}{r}$

Tabelle IV

p	H_1	H_2	H_3	p_R	p_F	$\overline{H}$
1,00	0,00000	0,00000	0,00000	0,99	0,01	0,08161
0,95	0,07400	0,07030	0,21610	0,95	0,05	0,30147
0,90	0,15200	0,13680	0,33219	0,90	0,10	0,52111
0,85	0,23446	0,19929	0,41054	0,85	0,15	0,71746
0,80	0,32193	0,25754	0,46438	0,80	0,20	0,90241
0,75	0,41504	0,31128	0,50000	0,75	0,25	1,08170
0,70	0,51457	0,36020	0,52109	0,70	0,30	1,25899
0,65	0,62149	0,40397	0,53010	0,65	0,35	1,43703
0,60	0,73696	0,44218	0,52877	0,60	0,40	1,61825
0,55	0,86250	0,47437	0,51840	0,55	0,45	1,80504
0,50	1,00000	0,50000	0,50000	0,50	0,50	2,00000
0,45	1,15200	0,51840	0,47437	0,45	0,55	2,20616
0,40	1,32193	0,52877	0,44218	0,40	0,60	2,42738
0,35	1,51457	0,53010	0,40397	0,35	0,65	2,66876
0,30	1,73696	0,52109	0,36020	0,30	0,70	2,93764
0,25	2,00000	0,50000	0,31128	0,25	0,75	3,24511
0,20	2,32193	0,46438	0,25754	0,20	0,80	3,60964
0,15	2,73696	0,41054	0,19929	0,15	0,85	4,06560
0,10	3,32193	0,33219	0,13680	0,10	0,90	4,68995
0,05	4,32193	0,21610	0,07030	0,05	0,95	5,72794
0,00	∞	0,00000	0,00000	0,01	0,99	8,07931

$H_1 = -\ \mathrm{ld}\, p$

$H_2 = -\ p\, \mathrm{ld}\, p$

$H_3 = -\ (1-p)\, \mathrm{ld}\,(1-p)$

$\overline{H} = \frac{1}{p_R}\,(H_2 + H_3) \qquad$ (gilt nur für $0 < p < 1$)