

Datenverarbeitung im Recht (DVR)

Archiv für die gesamte Wissenschaft der Rechtsinformatik, der Rechtskybernetik
und der Datenverarbeitung in Recht und Verwaltung

Band 7, 1978



J. Schweitzer Verlag · Berlin

Band 7, 1978

Inhalt

Abhandlungen

F. M. Auburn Retrieval of legal information from non-legal on-line automated information systems: AUSINET	367
Britt-Mari Blegvad Die Rolle des Juristen in der Gesellschaft	53
Willi Egloff Information und Grundrechte	115
Horst Joeres Gesetzesflut und Gesetzesdokumentation Zusammenfassung	17 26
Jaroslav Krecht Deep Structure Language	373
Lothar Philipps Aufgabenform und Auswertungsmodus bei multiple-choice-Aufgaben	341
Leo Reisinger Zur Anwendung der Entscheidungstabellen-Technik in der Legistik	151
Friedrich Schindele Die Sozialdatenbank. Eine informationsrechtliche Untersuchung	189
Karol Sobczak Verwaltung und personenbezogene Daten in Polen	357
Lotte Tuner Die Weiterentwicklung im Datenschutz	29
Georg Werckmeister Informationsrecht – Grundlagen und Anwendung im Überblick Begriff der Information	97 225
Christian Wollschläger Der Umfang der Veröffentlichung von Gerichtsentscheidungen	1
Henner Wolter Rechtsprobleme der Informationssammlung und Informationsweitergabe durch die Ämter für Verfassungsschutz	297

Rechtsprechung

Niedersächsisches Finanzgericht, Urteil vom 13. 4. 1977 IV 184/76	253
---	-----

Literatur

Besprechung der Schrift von Jon Bing und Trygve Harvold: „Legal decisions and information systems“ (Horst Joeres)	257
Besprechung der Schrift von Dworatschek: „Grundlagen der Datenverarbeitung“	258
Besprechung der Schrift von Jon Bing und Trygve Harvold: „Legal decisions and information systems“ (W. Steinmüller)	258
Besprechung der Schrift: „Computer aids to parliamentary work“ (Hans Georg Frohn)	260
Besprechung der Schrift von Gola/Hümmerich/Kerstan: „Datenschutzrecht“ (Thomas Barthel)	260
Besprechung der Schrift von Werner Thieme: „Verwaltungslehre“ (Lothar Messerschmidt)	262
Besprechung der Schrift von Helmut Dreßler: „Problemlösen mit Entscheidungstabellen“	262
Besprechung der Schrift von Dietrich Fürst: „Kommunale Entscheidungsprozesse“	263
Besprechung der Schrift von Peter Eichhorn und Peter Friedrich: „Verwaltungsökonomie I“	263
Besprechung der Schrift: „Computerkriminalität und Strafrecht“	264
Besprechung der Schrift: „Entscheidungstabellentechnik“	264
Besprechung der Schrift von Frits W. Hondius: „Emerging Data Protection in Europe“ (Gerhard Stadler)	265
Besprechung der Schrift von Meyer-Uhlenried, Karl Heinrich: „Methodische Grundlagen für die Planung von Informationssystemen“ (Horst Joeres)	266
Besprechung der Schrift von Reisinger, Leo: „Rechtsinformatik“ (Horst Joeres)	267
Besprechung der Schrift von Lutterbeck, Bernd: „Parlament und Information“ (Detlef W. Weber)	268
Besprechung der Schrift von Frohn, Hans-Georg: „Gesetzesbegriff und Gewaltenteilung“	272

Hinweise

Landesregierung verabschiedet Datenschutzgesetz Nordrhein-Westfalen (DSG NW)	275
Patientenüberwachung mit Computer – aber wer haftet bei Schäden?	277
Computer in der Redaktion	278
Holografische Echtheitsprüfung von Ausweisen	279
Parlaments-Informations-System – Stand der Entwicklung, Probleme, Ziele	280
Datev setzt Laserdrucker ein	281
Fernschreiben über Telefonleitung	281
Verwaltungen in Staat, Wirtschaft und Wissenschaft arbeiten 25% zu teuer	282
Buchführungsprivileg für Steuerberater	282

Computer hilft beim Autokauf	283
IPSIS für's Personal	284
Themen für Forschung und Diskussion	284
Datenbank wahrt Rechte an 700 000 Musikstücken	293
Zur Rolle des Telefons in der Bürokommunikation der Zukunft	294
Jedes Telex spart im Schnitt sieben Mark	295

Die Autoren der Beiträge

F. M. Auburn, Associate Professor, School of Law, University of Western Australia
Nederland, W. A. Australia

Willi Egloff, Dr. jur., Bezirksgericht Zürich, Rennweg 10, CH-8001 Zürich

Horst Joeres, Bundesministerium der Justiz, im Weiler 5, D-5300 Bonn 1

Jaroslav Krecht, Dr. jur., Institut für Staatsverwaltung Prag, 10700 Praha 7 U Uranie,
Tschechoslowakei

Lothar Philipps, Prof. Dr., Institut für Rechtsphilosophie und Rechtsinformatik der
Universität München, Perfallstr. 1, D-8000 München 80

Leo Reisinger, Doz. Dr. Dr., Institut für Statistik der Universität Wien, A-1010 Wien,
Rathausstr. 19/4

Friedrich Schindele, Rechtsreferendar beim Amtsgericht Regensburg, Breslauer Str.
68, D-8300 Landshut

Karol Sobczak, Prof. Dr. habil., Universität Warschau / Dekan der Verwaltungswissen-
schaften, 00-189 Warszawa, Inflancka Fakultät 15/38, Polska

Lotte Tuner, Dr. jur., GMD/Bonn-IDR, Auf dem Scheuel 1c, D-5204 Lohmar 1-Inger

Georg Werckmeister, Universität Regensburg/Abteilung für nichtnumerische Daten-
verarbeitung, Weingartenstr. 18, D-8400 Regensburg

Christian Wollschläger, Prof. Dr. jur., LL. M., Juristische Fakultät der Universität
Göttingen, Ludwig-Beck-Str. 9, D-3400 Göttingen

Henner Wolter, Dr. jur., Haußmannstr. 40a, D-7000 Stuttgart 1

Aufgabenform und Auswertungsmodus bei multiple-choice-Aufgaben

Übersicht

- | | |
|--|---|
| 1 Das medizinische Auswertungssystem: schlechte Diagnosemöglichkeiten. | 3 Das System des Funkkollegs: Pseudo-multiple-choice mit hohen Ratechancen. |
| 2 Das Saarbrücker System: gleiche Ratewahrscheinlichkeit bei verschiedenen Aufgabenformen. | 3.1 Das Hagener System: der programmierte Entscheidungskonflikt |

Es gibt zwei Grundformen von multiple-choice-Aufgaben: bei der einen ist genau eine der Alternativen richtig („1 aus n“, zumeist „1 aus 5“, – „Einfachantwortwahl“), bei der anderen können es mehrere sein („x aus n“, zumeist aus „5“, – „Mehrfachantwortwahl“). Man kann es auch so sehen: bei der einen Form stehen die Alternativen im Verhältnis eines ausschließenden „oder“ („entweder ... oder“), bei der anderen sind sie durch ein nicht-ausschließendes „oder“ verknüpft („und/oder“).

Die didaktischen Institutionen in der Bundesrepublik, die mit mc-Aufgaben arbeiten, beschränken sich vorwiegend auf die Form der Einfachwahl: so das Mainzer „Institut für Medizinische Prüfungsfragen“, das die Aufgaben für die bundeseinheitliche ärztliche Vorprüfung ausarbeitet und auswertet, die Studienstiftung des Deutschen Volkes, bei ihren Tests zur Vorauswahl Hochbegabter; soviel man weiß ist diese Form auch für etwaige Hochschuleignungstests bei „harten“ Numerusclausus-Fächern vorgesehen. Das Funkkolleg andererseits hat sich für Aufgaben mit Mehrfachwahl entschieden, und die Fernuniversität Hagen verwendet beide Formen.

Die Vorliebe für die Form der Einfachantwort ist verständlich. Diese ermöglicht ein einfaches und durchsichtiges System der Auswertung; die Ratechancen (20 % pro Aufgabe) sind gering und bei allen Aufgaben gleichbleibend. Unterschiedlich mehrere richtige Antworten pro Aufgabe tendieren dagegen zu unterschiedlichen Ratechancen. Unterschiedliche Ratenchancen erschweren den Vergleich der Schwierigkeitsgrade der Tests.

Die Beschränkung auf einen Typ läßt freilich die Ausdrucksmittel der mc-Aufgabe, die ohnehin nicht reichhaltig sind, empfindlich verarmen, und sie belastet das Geschäft des Aufgabenkonstruktors, das weiß Gott schon mühselig genug ist. Vor allem der Hochschullehrer, der selbstausgearbeitete, informelle Tests durchführen möchte, um seinen eigenen Lehrerfolg zu überprüfen: der Hochschullehrer außerhalb der Apparate einschlägiger Institutionen würde es begrüßen, wenn er nicht auf *eine* mc-Form angewiesen ist.

Wir haben vor einiger Zeit in Saarbrücken, im Rahmen einer kleinen Arbeitsgruppe für Rechtsinformatik, ein Computer-Programm zur Auswertung von multiple-

choice-Tests entwickelt, das es erlaubt, nebeneinander Aufgaben mit Einfach- und mit Mehrfachwahl zu verwenden, und zwar so, daß die Ratechancen bei allen Aufgaben invariant bleiben.¹

Das Verfahren ist inzwischen in Saarbrücken und München in einem Dutzend strafrechtlicher Tests erprobt worden.

Die Entwicklung des Verfahrens war von der Überzeugung geleitet, daß man beide Aufgabenformen *brauche*: aus inhaltlichen Gründen der Wissenschaften, und nicht einfach, weil zwei Formen benutzen zu können bequemer ist. Hinter dieser Überzeugung stand die Idee, daß die Form einer Aufgabe, auch die der multiple-choice, die Sachlogik eines Problems nachzeichnen müsse und nicht verzerren dürfe. Aber nicht jedes Problem fügt sich zwanglos in jede Aufgabenform.

In der Jurisprudenz beispielsweise bietet sich die mc-Aufgabe geradezu dazu an, Entscheidungssituationen abzubilden – in die ein Richter, ein Staatsanwalt oder ein Verwaltungsbeamter gestellt ist. Nun gibt es aber Situationen, in denen genau eine, die anderen Verhaltensmöglichkeiten erkennbar ausschließende Handlung verlangt wird, während wiederum in anderen Situationen Mehreres zu tun sein mag. Jede der beiden mc-Formen hat, so eingesetzt, ihr Recht und ihre Notwendigkeit.

„Zugegeben“, wird manch einer jedoch skeptisch bleiben, „aber was macht es schon aus, zwei statt einer Aufgabenform zu haben angesichts der unendlichen Vielfalt, in der die Lebenswelt rechtlichen Problemen Gestalt gibt?“

Dies ist zwar richtig; aber bei der mc-Aufgabe handelt es sich um eine Beschränkung besonderer Art und von besonderer Bedeutung: sie korrespondiert mit der Digitalisierung, zu der die Lebenswelt heute, auch im Rechtlichen, tendiert. Die mc-Aufgabe bietet eine einzigartige Möglichkeit, dem Rechtsstudenten, schon vom ersten Semester an, methodologische Probleme der Computerisierung des Rechts nahezubringen, und zwar in der denkbar eindringlichsten Weise: unter dem eigenen Problemlösungsdruck. Sie ermöglicht es ferner, neue Techniken der intersubjektiven Sicherung rechtlichen Formulierens zu erproben, – eine mc-Aufgabe, die nicht intersubjektiv eindeutig ist, funktioniert nicht. Und wir haben in der mc-Aufgabe und ihrer Bewertung ein Modell – überschaubar und mE in mancher Hinsicht generalisierbar – für die Beziehung zwischen Information und Gerechtigkeit, jene sehr eigenständige moderne Variante des alten Themas Freiheit und Gerechtigkeit.²

Alles in allem ist die mc-Aufgabenform mittlerweile viel zu wichtig geworden, als daß man sie allein den Didaktikern und Psychologen zur Diskussion überlassen dürfte.

Die folgende Kritik an einigen Auswertungsverfahren für mc-Aufgaben, auch soweit sie sich gegen Techniken richtet und an technisches Verständnis appelliert, ist nicht zuletzt von einem normativen, man darf sogar sagen: rechtlichen Standpunkt aus konzipiert. Sie setzt sich zunächst mit einem geistreichen und häufig

1 Vgl auch zu den folgenden Ausführungen, *Philipps*, Testaufgaben in der Rechtswissenschaft.

2 Vgl *Philipps*, Recht und Information.

praktizierten Versuch auseinander, die Einseitigkeit des Einfachwahlsystems mit einem logischen Trick zu umgehen (1).

Es folgt die Erläuterung des Saarbrücker Systems (2). Das Auswertungsverfahren des Funkkollegs, statistisch äußerst problematisch, wird in einem dritten Kapitel behandelt – versehen mit einem kurzen Nachtrag über eine jüngst erfolgte Modifizierung dieses Verfahrens in der Fernuniversität Hagen (3).

1 Das medizinische Auswertungssystem³: schlechte Diagnosemöglichkeiten

An und für sich ist es möglich, mit der Aufgabenform, in der nur nach einer Alternative gefragt wird, auszukommen: auch dort, wo von der Sache her mehrere Alternativen richtig sein können. Mit einem einfachen logischen Kunstgriff kann man sich hier helfen: Man verbindet jeweils mehrere der ursprünglichen Alternativen in verschiedenen Kombinationen durch ein „und“ derart, daß nur noch eine dieser Verbindungen richtig ist. Das „und“ ist eine wesentlich stärkere Verbindung als das „oder“; es genügt eine falsche Teilaussage, und die ganze Aussage ist falsch.

Die beiden folgenden Beispiele sind in den „Hinweisen für Sachverständige“ enthalten, die das Mainzer „Institut für Medizinische Prüfungsfragen“ versendet.

Beispiel I:

Für die Harnstoffausscheidung in der Säugetierriere gilt:

- (1) Harnstoff wird im Glomerulum frei filtriert.
- (2) Er wird im proximalen Tubulus aktiv transportiert.
- (3) Die Harnstoff-Clearance ist niedriger als die Inulin-Clearance.
- (4) Die Ausscheidung von Harnstoff ist unabhängig von der Urinmenge.

Wählen Sie bitte unter folgenden Aussagenkombinationen diejenige, die Sie für zutreffend halten.

- | | |
|----------------------------------|-----------------|
| (A) nur 1, 2 und 3 sind richtig | 1, 2, 3, 4' |
| (B) nur 1 und 3 sind richtig (r) | 1, 2', 3, 4'(r) |
| (C) nur 2 und 4 sind richtig | 1', 2, 3', 4 |
| (D) nur 4 ist richtig | 1', 2', 3', 4 |
| (E) alle Aussagen sind richtig | 1, 2, 3, 4 |

3, 3, 3, 3

In der linken Kolumne sind die Alternativen in der *Originalschreibweise* angegeben: das Wesentliche daran ist, daß die als fehlerhaft mitbehaupteten Teilaussa-

³ Mit dem Ausdruck „medizinisches Auswertungssystem“ für die Kombinationsaufgaben-Technik ist natürlich nur die größte pars pro toto benannt, zumal gerade *Hubbard* sich für dies Verfahren stark gemacht hat.

gen nicht explizit angegeben sind. In der rechten Kolumne sind die Alternativen entsprechend ergänzt worden: die negierten Teilaussagen sind jeweils durch einen Index gekennzeichnet. Unter dem Strich ist die Häufigkeit angegeben, mit der die Teilaussagen 1 bis 4 bejaht werden.

Beispiel II:

Zur Fallot'schen Tetralogie gehören:

- (1) Vorhofseptumdefekt
- (2) Dextroposition der Aorta
- (3) Ventrikelseptumdefekt
- (4) Pulmonalstenose
- (5) Tricuspidalinsuffizienz

Welche Antwort halten Sie für zutreffend?

- | | |
|------------------------------------|--------------------|
| (A) nur 1, 2, 3 und 4 sind richtig | 1, 2, 3, 4, 5' |
| (B) nur 3 und 4 sind richtig | 1', 2', 3, 4, 5' |
| (C) nur 2, 3 und 4 sind richtig(r) | 1', 2, 3, 4, 5'(r) |
| (D) nur 1, 2 und 3 sind richtig | 1, 2, 3, 4', 5' |
| (E) nur 2, 3, 4 und 5 sind richtig | 1', 2, 3, 4, 5 |

2, 4, 5, 4, 1

Für diese Art Aufgaben hat sich vor allem *John P. Hubbard* in seinem grundlegenden Buch „Erfolgsmessung der medizinischen Ausbildung“ eingesetzt. Einen naheliegenden Einwand hat er dabei selber vorweggenommen: der Student stehe vor einer Alles-oder-Nichts-Situation, Teilwissen werde nicht honoriert⁴. Hubbard ist aber der Meinung, man solle diesen Mangel um des Vorteils willen in Kauf nehmen, daß nach wie vor „auch jetzt auf ein Item ein Punkt entfällt. Die Punkte für die Fragen einer Prüfung können mit gleichem Gewicht aufsummiert werden, ohne daß der Itemtyp berücksichtigt werden muß.“

Hubbard hat indessen diesen Mangel des Systems vorschnell eingeräumt; tatsächlich wird Teilwissen durchaus berücksichtigt, wenn auch in einer etwas bizarren Weise. Andererseits ist seine Behauptung, der Test bleibe homogen, nur mit Vorbehalt aufzunehmen.

Dies wird sogleich einsichtig, wenn man die Alternativen nicht, wie im Originaltext, in abgekürzter Form niederschreibt, sondern so, daß auch die negierten Teilaussagen explizit angeführt sind. Die beiden Beispiele oben sind in dieser Weise ergänzt worden. Man betrachte zunächst Beispiel II! Wer hierzu nur das eine weiß, daß die Teilaussage (2) falsch ist, hat nun immerhin eine Ratechance von 50 %, die richtige Alternative zu treffen; denn daß (2) falsch ist, wird nur in den beiden Alternativen B und D vorausgesetzt. Ebenso günstig steht, wer nur weiß, daß (4) falsch ist, dies ist nur in A und B vorausgesetzt.

⁴ Erfolgsmessung, S 55.

Trotzdem sind (2') und (4') nicht etwa gleichrangig; (2') ist wesentlich gewichtiger als (4'). Wer weiß, daß (2) falsch ist, braucht nur noch über *eine* der übrigen Teilaussagen – gleichgültig welche – Bescheid zu wissen, dann ist ihm der Punkt sicher; die übrigen Teilaussagen treten nämlich in B und D jeweils mit verschiedenen Vorzeichen auf. Wer dagegen weiß, daß (4) falsch ist, dem nützt weiteres Teilwissen gar nichts – es sei denn eben, er weiß auch, daß (2) falsch ist! Er mag sich aber damit trösten, daß er sich dennoch deutlich besser steht als derjenige, der immerhin weiß, daß (1) *und* (3) richtig ist; denn die Verbindung (1) und (3) tritt in drei Alternativen auf: in A, B und E. Diesem Teilnehmer würde es auch noch nicht zu dem sicheren Punkt verhelfen, daß er weiterhin über (4) Bescheid weiß; er muß schon jenes über alles andere erhobene Teilwissen (2') haben, daß Harnstoff im proximalen Tubulus nicht aktiv transportiert wird. Man sieht: Teilwissen wird durchaus honoriert und sogar in gewichteter Weise und mit der besonderen Finesse einer aleatorischen Gewichtung. Man bekommt zwar keine Teilpunkte, aber eine abgestufte große Ratechance.

Ein solches Verfahren braucht gar nicht einmal sinnlos zu sein; Vergleichbares hat man auch erst unlängst für die Zulassung zum Hochschulstudium vorgeschlagen: die Studienplätze werden ausgelost, aber wer ein gutes Abiturzeugnis hat, erhält ein Los mit höherer Gewinnaussicht als ein schwächerer Abiturient („leistungsgesteuertes Losverfahren“)⁵. Hier jedoch ist zu bezweifeln, daß ein Sinn dahintersteht. Ist das Teilwissen, daß (2) falsch ist, um so viel höher zu bewerten als das Teilwissen, daß (1) *und* (3) richtig ist, daß die Chancen, den Punkt zu erhalten, im Verhältnis von 50 % zu 33,3 % gewichtet sind? Verdient es in Verbindung mit beliebigem anderem Teilwissen gar eine Gewichtung von 100 %? Als Laie kann man das zwar nicht beurteilen, darf aber doch wohl vermuten, daß, wenn dies Ergebnis gewollt wäre, in den „Hinweisen“ des Mainzer Instituts auch ein Hinweis hierauf zu finden wäre. Tatsächlich ist die Aufgabe von *Hubbard* übernommen worden⁶, der das Alternativenschema zwar auch nicht kommentiert; aber sich vermutlich etwas dabei gedacht hat – und zwar, daß es immer noch das kleinere Übel sei.

Die Mängel der Aufgabe sind nämlich keineswegs zufälliger Natur, sondern systembedingt: wenn man sie vermeidet, treten andere auf. In Beispiel II hat man sie vermieden. Wer hier über nur eine Teilaussage Bescheid weiß, hat bestenfalls eine Ratechance von 33,3 %: bei (1'). Hier wird also nicht ein Teilwissen dem andern in auffälliger Weise vorgezogen.

Das Fatale ist freilich jetzt, daß die intelligenteren Ignoranten, zumal die mit der Kodeknacker-Mentalität, gar nicht mehr blind raten werden, sondern strategisch: (3) ist mit Sicherheit richtig; das ist trivial. (2) ist höchstwahrscheinlich richtig, denn es tritt in vier der fünf Alternativen auf; ebenso (4). Das Gleiche gilt für die Annahme, daß (5) falsch sei. Bei (1) kann man nicht so zuversichtlich sein; denn hier ist das Zahlenverhältnis lediglich 2 : 3. Immerhin, wenn man keinerlei Wissen

5 Ein solches System wurde von *Max Martin Scheer* ausgearbeitet; vgl DER SPIEGEL Nr 15, 1976, S 75.

6 aaO, S 202, Aufgabe Nr 97.

darüber hat, ist es rational, sich für das Häufigere zu entscheiden und also anzunehmen, daß (1) falsch sei. Alle diese verschiedenen Annahmen schneiden sich in einer Alternative, in C. Dies ist denn auch die richtige.

Ist das aber nicht ein Spiel des Zufalls? Hätte der Aufgabensteller die Aufgabe nicht auch so formulieren können, daß – bei gleicher Ziffernverteilung in den Alternativen – beispielsweise E richtig gewesen wäre? Zweifellos hätte er das können – doch nur auf Kosten der Gerechtigkeit. Nunmehr würde es schon genügen zu wissen, daß (5) zutrifft, um die Aufgabe zu lösen; wer dagegen über *alle* andere vier Teilaussagen Bescheid weiß, hätte die richtige Alternative immer noch nicht (vgl C und E)! Die Teilaussage (5) wäre also fünfmal so hoch gewichtet wie jede andere Teilaussage – ein solcher Unterschied wird sich unmöglich rechtfertigen lassen.

Die Gesetzmäßigkeit, die hier waltet, ist recht einfach: Will der Aufgabenkonstrukteur verhindern, daß die Aufgabe schon mit Teilwissen gelöst werden kann, so muß er eine zutreffende Teilaussage auf möglichst viel Alternativen verteilen. Eine falsche Teilaussage muß er umgekehrt auf möglichst wenige Alternativen beschränken, sonst kann der, der weiß, daß sie falsch ist, sich durch Streichung dieser Alternativen der Lösung allzusehr nähern.

Wenn der Aufgabenkonstrukteur aber so verfährt, gibt er durch die Form der Aufgabe zu erkennen, welches die wahrscheinlich richtigen Teilaussagen sind, und darüber hinaus auch, welches die wahrscheinlich zutreffende Alternative ist. Denn alle richtigen Teilaussagen müssen ja in einer Alternative zusammentreffen, und eine falsche darf in ihr nicht enthalten sein. Der Konstrukteur steht also vor dem Dilemma, das eine Teilwissen vor dem anderen zu begünstigen oder aber die Lösung anzudeuten. Vermutlich wird er den zweiten Weg bevorzugen.

Wir haben zwei Bände mit mc-Fragen für Medizin-Studenten ausgewertet. Es handelt sich hier freilich nicht um Original-Fragen des Mainzer Instituts – diese bleiben unter Verschuß, die Fragebögen werden nach jeder Prüfung wieder eingesammelt – sondern um Aufgaben nach Art jenes Hauses, die ein privater Verlag herausgibt⁷. Die beiden Bände – Anatomie und Biologie – wurden willkürlich aus dem Verlagsangebot herausgegriffen. Sie enthalten 1297 mc-Fragen, darunter 268 Kombinationsaufgaben.

Wir sind von der Hypothese ausgegangen, daß die Teilaussagen, die in den Alternativen einer Kombinationsaufgabe am häufigsten vorkommen, zumeist richtig sind, und die am wenigsten vorkommenden zumeist falsch. Wir haben dabei unterschieden zwischen einem „schmalen Maximum“ (\max_s), wo eine *einzig*e Teilaussage an der Spitze liegt, und einem „breiten Maximum“ (\max_b), wo mehrere Teilaussagen in gleicher Häufigkeit vor den übrigen rangieren; entsprechend auch zwischen einem „schmalen Minimum“ (\min_s) und einem „breiten Minimum“ (\min_b). In acht der ausgewerteten Aufgaben kommen alle Teilaussagen gleich

7 Der Wert dieser kommerziellen Aufgabensammlungen wird übrigens in der Literatur angezweifelt; vgl Ucke, Zur Verwendung von Antwort-Auswahl-Aufgaben, S 28.

Zur Frage der Geheimhaltung lese man auch die instruktive Schilderung im SPIEGEL Nr 7, 1977, S 38: wie im alten Rom vor Ganeus Flavius.

häufig vor; diese Aufgaben sind in die folgende Übersicht nicht mit aufgenommen worden.

Multiple-choice-Fragen

Biologie und Anatomie

Übersicht zum Aufgabentyp D (Kombinationsaufgaben)

	260	min _s /f 109 41,9 %	min _s /r 32 12,3 %	min _b /f 37 14,2 %	min _b /r 82 31,5 %
max _s /r	122 46,9 %	53 20,4 %	11 4,2 %	25 9,6 %	33 12,7 %
max _s /f	24 9,2 %	6 2,3 %	4 1,5 %	2 0,8 %	12 4,6 %
max _b /r	70 26,9 %	32 12,3 %	13 5,0 %	5 1,9 %	20 7,7 %
max _b /f	44 16,9 %	18 6,9 %	4 1,5 %	5 1,9 %	17 6,5 %

Die Matrix bestätigt die Hypothese in sehr deutlicher Weise.

In 146 von 260 Aufgaben bildet eine Teilaussage ein schmales Maximum; in 122 Fällen ist diese Teilaussage richtig. Also eine Wahrscheinlichkeit von 83,6 %, mit der sich's in 56,2 % der Aufgaben raten läßt!

In 141 von 260 Aufgaben (54,2 %) findet sich ein schmales Minimum. In 109 Aufgaben, das sind 77,3 % der Fälle, ist die entsprechende Teilaussage falsch.

Ein breites Maximum ist richtig in 61,4 % der Fälle (43,8 %). Das wird manchem auf den ersten Blick verhältnismäßig wenig erscheinen, – aber sehr zu unrecht. Daß ein breites Maximum richtig ist, bedeutet, daß jede der darin enthaltenen Teilaussagen richtig ist; der Informationsgehalt dieser Feststellung ist also viel höher als bei einem schmalen Maximum. Wenn man davon ausgeht, daß die einzelnen Teilaussagen mit gleicher Wahrscheinlichkeit richtig wie falsch sein können, und wenn man weiterhin unterstellt, daß die breiten Maxima nur zwei Teilaussagen umfassen, kann man a priori – nach dem Multiplikationssatz der Wahrscheinlichkeitsrechnung – nur eine Wahrscheinlichkeit von 25 % erwarten, und bei näherer Betrachtung der Aufgaben noch viel weniger, da Maxima von drei oder vier Teilaussagen nicht selten sind. Deshalb ist ein Wert von 61,4 % sehr hoch.

Bei den breiten Minima ist die Tendenz viel schwächer ausgeprägt, allerdings auch nicht gegenläufig, wie es nach einem flüchtigen Blick auf die Matrix scheinen könnte; man muß den Multiplikationssatz bedenken. (31 % sind falsch, statt – wie a priori zu erwarten – höchstens 25 %).

In Wirklichkeit sind die Ratechancen aber noch höher, als sich aus unserer Aufstellung ergibt; denn die Einordnung nach Maximum und Minimum ist allzu pauschal und berücksichtigt nicht die Feinstruktur der einzelnen Aufgabe. Ein

Teilnehmer mit einigermaßen Urteilskraft wird unschwer etliche erkennen, bei denen er nicht oder nur mit sehr geringen Aussichten von der Vermutung ausgehen kann, daß das Maximum wahr und das Minimum falsch ist.

In Aufgabe 158 (Anatomie) beispielsweise ist folgendes Alternativenschema vorgeschlagen:

- | | |
|--------------------|--|
| A. nur 1, 2, 4 | Die Teilaussagen (1), (2), (3) und (5) kommen je dreimal vor, die Teilaussage (4) zweimal. |
| B. nur 1, 3, 4 (r) | |
| C. nur 2, 3, 5 | |
| D. nur 2, 5 | |
| E. nur 1, 3, 5 | |

In der Matrix ist diese Aufgabe unter den Fällen aufgeführt, wo weder das (breite) Maximum richtig noch das (schmale) Minimum falsch ist; der Fall scheint also die Ausgangshypothese abzuschwächen. Indessen ist es hier schon aus formalen Gründen nicht möglich, daß das Maximum richtig ist; denn die Kombination 1, 2, 3, 5 kommt nicht vor. Aber selbst wenn sie zur Wahl stünde, würde kein vernünftiger Mensch eine ernsthafte Wette darauf eingehen, daß bei einem derartig breiten Maximum und einem Minimum, das nur um einen Zähler davon differiert, die Ausgangshypothesen zutreffen.

Würde man das analytische Instrumentarium verfeinern, etwa die Breite der Maxima und Minima quantifizieren und auch ihren Abstand vom „Feld“, so käme man sicherlich noch auf wesentlich höhere Ratechancen. Deshalb seien die medizinischen mc-Aufgaben den Wissenschaftstheoretikern ans Herz gelegt und für Übungen in induktiver Logik, semantischer Informationstheorie und Entscheidungstheorie empfohlen: sie sind ein großartiges Spielmaterial.

In welcher Weise diese Tendenzen durch die Psychologie der Aufgabensteller vermittelt sind, wissen wir nicht. Es kann sein, daß man bewußt vermeiden wollte, partielles Wissen zu honorieren; möglicherweise haben sich auch unbewußte Strukturvorstellungen durchgesetzt; vielleicht war aber auch das Bemühen, inhaltlich plausible Alternativen anzubieten, der Grund, sich vorwiegend in der Nähe der zutreffenden Teilaussagen zu halten. Solche Mutmaßungen können dahingestellt bleiben; auch wer das Dilemma durchschaut, entgeht ihm nicht.

Offenbar war sich *Hubbard* des Dilemmas bewußt, wenn er auch nicht ausdrücklich darauf eingeht. Es ist sicher nicht ohne Grund, daß er vorschlägt, für alle Kombinationsaufgaben ein und dasselbe Alternativenschema zu verwenden, und zwar eben das dem Beispiel I zugrundeliegende. Das Schema ist wohlgedacht: Alle Teilaussagen kommen mit derselben Häufigkeit (3) vor; auf den ersten Blick scheint deshalb eine ungleichmäßige Gewichtung vermieden zu sein. Erst wenn man ins Detail geht, erkennt man, daß die Lösung doch nicht ganz gelungen ist. Im übrigen erschwert es das starre Schema erheblich, geeignete Aufgaben zu formulieren. Daß eine einzelne Teilaussage als zutreffend erscheint – wie in dem Schema stets vorgesehen – dürfte bei Kombinationsaufgaben nur ausnahmsweise der Fall sein; andererseits wird man manche attraktive Kombinationsmöglichkeit nicht unterbringen können. Gewiß ist die mc-Aufgabe ohnehin eine sehr strenge Kunst-

form (das Sonett unseres didaktischen Zeitalters); aber dies wäre denn doch zu viel des Verzichts.⁸

Daß Teilwissen scheinbar nicht und genau gesehen in bizarrer Weise honoriert wird, dürfte auf viele Studenten frustrierend wirken. Aber die Kehrseite hiervon ist – diesmal für den Dozenten – nicht weniger unerfreulich: sie besagt nämlich, daß Teil*nicht*wissen nicht genau erkannt wird. Das ist unmittelbar plausibel: je höher die Ratechancen für die Teilnehmer, desto niedriger die Diagnosechancen für den Dozenten; und bei ungleichmäßigen Ratechancen ist eine verzerrte Diagnose zu erwarten.

Es lohnt sich aber, hierauf noch etwas näher einzugehen: Solange es offen ist, welche von ursprünglich vier oder fünf Teilaussagen zutreffen, stehen 16 oder 32 Kombinationen zur Wahl (2^4 und 2^5). Bei dem Mainzer System muß deshalb der Aufgabensteller eine Vorwahl treffen: er muß aus jener Vielzahl von Kombinationen neben der richtigen auch vier falsche heraussuchen, die gleichwohl plausibel sind. Da ist ein ganz erheblicher Aufwand und ein möglicherweise verfehlt: wie kann der Aufgabensteller wissen, ob er wirklich die für den Studenten nächstliegende falschen Kombinationen trifft? Vielleicht gibt es viel attraktivere Alternativen, die aber den Teilnehmern gar nicht vorgelegt werden.

Dadurch wird die Fehlerdiagnose unsicher, dh richtiger gesagt: die Diagnose der *Irrtümer*, denn die Teilnehmer haben nur eine begrenzte Chance, ihre Irrtümer als Fehler offenzulegen. So auch in dem Beispiel II von der „Fallotschen Tetralogie“: Wer der irrigen Meinung ist, der Ventrikelseptumdefekt (3) gehöre *nicht* dazu, kann dies nicht zum Ausdruck bringen; denn (3) kommt in allen fünf Alternativen vor. Und wer fest überzeugt ist, die Tricuspidalinsuffizienz (5) gehöre dazu, ist gezwungen, die Alternative E zu wählen; denn nur in ihr ist (5) enthalten. Welche Vorstellungen dieser Teilnehmer von den anderen vier Phänomenen hat, erfährt man nicht. Möglicherweise hat er (2), (3) und (4) von vornherein für richtig gehalten; dann ist er der Wahrheit immerhin recht nahe gewesen. Vielleicht hat er aber (2), (3) und (4) ursprünglich für falsch gehalten – nur daß sein Glaube an (5) eben noch stärker war –, dann hat er gar nichts gewußt! Sind es viele Teilnehmer,

8 In den 268 Kombinationsaufgaben, die wir untersucht haben, ist es nur achtmal der Fall, daß alle Alternativen mit der gleichen Häufigkeit vorkommen; daran kann man ermes- sen, wie einschneidend *Hubbards* Vorschlag ist. Von 38 Kombinationsaufgaben, die *Hubbard* selber anführt, ist nur einmal die Alternative mit der einzelnen Teilaussage richtig; eigentlich müßte das ungefähr bei jeder fünften Aufgabe der Fall sein. Wenn sich das schon bei diesen Musteraufgaben so verhält, so spricht das sehr dafür, daß das Schema viel zu verpflichtend ist.

In der Schweiz hat man freilich *Hubbards* Schema übernommen. Daneben ist aber noch ein anderes System in Gebrauch, wonach man, um *einen* Punkt zu erhalten, zu *jeder* von vier Alternativen zutreffend angeben muß, ob sie wahr oder falsch ist. Vgl die *iae*-Hinweise S 7 f.

In den *iae*-Hinweisen (S 11) findet sich übrigens ein interessanter Umrechnungsschlüs- sel für *mc*-Aufgaben. Als Einheiten sind einerseits bit und andererseits der Informa- tionsgehalt einer „1 aus 5“-Aufgabe gewählt. Daß nach den *iae*-Hinweisen der Gehalt einer Kombinationsaufgabe gleich dem einer üblichen „1 aus 5“-Aufgabe sein soll, ist freilich nur insoweit richtig, als dabei das partielle Vorwissen der Teilnehmer und der aus dem Prinzip Gerechtigkeit folgende Sachzwang außer acht gelassen sind.

die E gewählt haben, so ist der Wert der Testaufgabe unter dem Gesichtspunkt der Rückmeldung des Lehrerfolgs gesehen ziemlich gering.

Das Mainzer System erinnert an das System der „Blockwahl“ in der Politik, wo man gegebenenfalls auch um eines Kandidaten willen eine ganze Liste akzeptieren oder verwerfen muß⁹; so wenig wie man hier seinen politischen Willen rein zum Ausdruck bringen kann, so wenig dort sein Wissen oder Nichtwissen. In der Didaktik spricht man von „Stellvertreter-Aufgaben“: der Ausdruck deutet treffend an, daß eine Aufgabe nur stellvertretend für die eigentlich gemeinte Aufgabe steht.¹⁰ Durch die Mediatisierung des Problems, durch die Distanz zwischen Problem und formulierter Aufgabe tritt jedoch ein Informationsverlust ein; man sollte es sich deshalb zur Maxime machen, Stellvertreteraufgaben wenn nur irgend möglich zu vermeiden. *John P. Hubbard* betont zwar, Untersuchungen des National Board of Medical Examiners der USA hätten ergeben, daß keine wesentlichen Änderungen in der Rangfolge der Teilnehmer aufträten, wenn man die eine Aufgabenform durch die andere ersetzte.¹¹ Das ist jedoch nur für den tröstlich, der in erster Linie den selektierenden Aspekt der Prüfung im Auge hat. Man sollte aber nicht nur wissen wollen, welche Teilnehmer schwach sind, sondern auch, aufgrund welcher Irrtümer sie es sind.

2 Das Saarbrücker System: gleiche Ratewahrscheinlichkeit bei verschiedenen Aufgabenformen

Im Saarbrücker System¹² können mehrere Alternativen richtig sein (x aus 5); die Teilnehmer haben also die Auswahl aus allen Kombinationen und können deshalb unmittelbar und im Detail zeigen, wo ihre Stärken und Schwächen liegen. Allerdings gerät man nun bei der Bewertung der Antworten in ein anderes Dilemma. Würde man lediglich nach dem Prinzip vorgehen, für jede richtige Wahl einen Punkt zu geben, so brauchte der Teilnehmer nur die robuste Strategie zu verfol-

9 Das Wort von der „Blockwahl“ ist gar nicht einmal als gehässige Metapher gemeint; es könnte durchaus wissenschaftlich fruchtbar sein, didaktische und politische Abstimmungstechniken umfassend zu behandeln. Dies um so mehr, als man sich hier auch auf staatsrechtlich-politischer Seite zuweilen auf tatsächliche oder vermeintliche Sachzwänge beruft: „Das Volk kann nur Ja oder Nein sagen; es kann nicht beraten, deliberieren oder diskutieren; es kann nicht regieren und nicht verwalten; es kann auch nicht normieren, sondern nur einen ihm vorgelegten Normierungsentwurf durch sein Ja sanktionieren. Es kann vor allem auch keine Fragen stellen, sondern nur auf eine ihm vorgelegte Frage mit Ja oder Nein antworten . . .“ *Carl Schmitt*, Legalität und Legitimität, S 93.

10 vgl *Rütter*, Testaufgabe, S 208 f.

11 aaO, S 55.

12 Der Name „Saarbrücker System“ ist ein wenig anmaßend; denn es werden an der Saarbrücker Universität auch andere Systeme für mc-Aufgaben verwandt, wenn auch nicht mit unmittelbarer Auswertung durch den Computer. Es ist uns aber leider keins von diesen geistreichen Akronymen eingefallen. Übrigens ist auch die Bezeichnung „Mainzer System“ für das medizinische Prüfungssystem nicht frei von Ungerechtigkeit; denn es wurde in Marburg entwickelt. Vgl *Schulz*, MMW 1969, 1412.

gen, stets alle fünf Alternativen zu lochen: die richtigen wären immer darunter. Eine solche Umgehungsmöglichkeit muß natürlich ausgeschlossen werden. Das Nächstliegende wäre, wie für jede richtige Antwort einen Pluspunkt, so für jede falsche Antwort einen Minuspunkt zu erteilen. Indessen wird das Prinzip der Minuspunkte in der heutigen Didaktik überwiegend abgelehnt.¹³

Die Ablehnung von Minuspunkten ist berechtigt. Ein Teilnehmer, der glaubt, bei einer Aufgabe einen Punkt erhalten zu haben, soll nicht der Furcht ausgesetzt sein, ihn bei einer der nächsten Aufgaben wieder zu verlieren. Sein Verhalten hinge dann wesentlich von seiner Risikohaltung ab; aber nicht diese soll in einem Leistungstest gemessen werden. Ein Test darf auch nicht die Wirkung haben, den Studenten die Bereitschaft zum wissenschaftlichen Risiko, und das heißt eben auch: zum Raten, auszutreiben; das wäre didaktisch nicht zu verantworten. Neben dem blinden Raten gibt es schließlich auch mit fließenden Übergängen ein intelligentes Raten, das durchaus nicht zu verachten ist; ebenso wie es andererseits auch ein blindes Wissen gibt. Man sollte also den Teilnehmern nicht das Raten verbieten, sondern den Test so gestalten, daß sie statistisch gesehen bei blindem Raten keine nennenswerte Erfolgchance haben.

Deshalb haben wir einen Kompromiß gewählt: die Teilnehmer erhalten zwar für eine falsche Teilantwort einen Minuspunkt, doch werden die Minuspunkte nicht über die einzelne Aufgabe hinweg übertragen: in jedem Falle sind also null Punkte das schlechteste Ergebnis, das bei einer Aufgabe herauskommen kann. Damit ist wenigstens vermieden, daß so etwas wie eine psychische Hypothek anwächst: eine sich mit jeder Aufgabe steigernde Besorgnis, das Erreichte bei den nächsten Aufgaben wieder aufs Spiel zu setzen. Je enger die Alternativen einer Aufgabe um eine Problemachse gelagert sind, desto unproblematischer ist dieser Kompromiß.

Was nun die Ratechancen bei diesem System anlangt, so liegen sie gleichbleibend bei 20 %. Die Ratechancen sind also die gleichen wie bei der Aufgabenform „1 aus 5“, wo die Alternativen sich wechselseitig ausschließen und nur eine richtig sein kann. Daß dieser Fall von dem hier vorgeschlagenen Bewertungsprinzip als Grenzfall miterfaßt wird, ist klar: trifft man die richtige Alternative, so bekommt man einen Punkt und sonst keinen – und die Ratechance beträgt dabei offensichtlich 20 %. Aber daß sich die Ratechance auch nicht ändert, wenn man die einschränkende Voraussetzung aufgibt, ist eine ebenso überraschende wie praktisch bedeutsame Einsicht. Sie besagt nämlich, daß die beiden Grundformen der mc-Aufgabe im Saarbrücker Bewertungssystem ohne weiteres miteinander kompatibel sind. Daß jemand beispielsweise die Hälfte der möglichen Punktzahl in einem Test erreicht, bedeutet das Gleiche, unbeschadet des Mischungsverhältnisses, in dem die beiden Aufgabenformen vorkommen. Die Sorge um die Kompatibilität war es aber, die Hubbard dazu veranlaßt hat, sich dafür einzusetzen, daß die Aufgaben in den medizinischen Prüfungen so homogenisiert werden, wie es jetzt im Mainzer Institut geschieht.

Hinter der Technik des Saarbrücker Systems steht die einfache Idee eines bewegli-

¹³ Vgl. *Ingenkamp*, Tests in der Schulpraxis, S 117 ff; *Herbig*, Lehrzielorientierte Tests, S 165 f.

chen Gleichgewichts: Wenn die Wahrscheinlichkeit, einen Punkt zu erlangen, sich erhöht – weil es in der Aufgabe mehrere richtige Alternativen gibt –, so sinkt im selben Maße die Bedeutung des Punktes – eben weil er nur einer von mehreren ist.

Die folgenden drei Gleichungen drücken die Fälle aus, daß (1) eine, (2) zwei und (3) drei zutreffende Alternativen in der Fünfergruppe enthalten sind. Im Falle (1) beträgt die Wahrscheinlichkeit, einen Punkt zu erhalten, $\frac{1}{5}$. Im Falle (2) steigt sie auf $\frac{2}{5}$; dies ist dann aber auch nur einer von zwei Punkten: die Ratewahrscheinlichkeit pro Punkt beträgt nach wie vor $\frac{1}{5}$. Im Falle (3) verhält es sich entsprechend.

$$(1) \frac{1}{5} \times 1 = \frac{1}{5}$$

$$(2) \frac{2}{5} \times 1 \times \frac{1}{2} = \frac{1}{5}$$

$$(3) \frac{3}{5} \times 1 \times \frac{1}{3} = \frac{1}{5}$$

Vorausgesetzt ist dabei die hier optimale Ratestrategie, stets nur eine Alternative anzukreuzen¹⁴: (3) ist der Grenzfall für diese Strategie. Wenn (3) vorliegt – freilich müßte man das wissen! – ist es genauso rational, zwei Alternativen zu wählen. Man hat jetzt eine Chance von drei zu zehn, zwei Punkte zu erlangen.

$$\frac{3}{10} \times 2 \times \frac{1}{3} = \frac{1}{5}$$

Bei vier zutreffenden Antworten im Fünferbündel wäre es günstiger, auf zwei Alternativen zu setzen. Die Chance, einen Punkt zu erlangen, steigt dann auf $\frac{1}{4}$ an.

Deshalb sollte man sich beim Saarbrücker System mit höchstens drei richtigen Alternativen pro Aufgabe begnügen, eine Einschränkung, die sich auch in den Regeln des Funkkollegs findet, dort aber die ganz andere Auswirkung hat, die Ratechancen ansteigen zu lassen.

3 Das System des Funkkollegs: Pseudo-multiple-choice mit hohen Ratechancen

Minuspunkte tun weh, auch wenn sie nur nach dem Prinzip der (auf eine Aufgabe) beschränkten Haftung erteilt werden. Es ist uns deshalb wiederholt von Studenten vorgeschlagen worden, das Auswertungssystem des Funkkollegs zu übernehmen. Hier kann man für jede Aufgabeneinheit bis zu fünf Punkte erhalten: einen Punkt für jede richtige Teilantwort, die man ankreuzt, und einen Punkt für jede falsche, die man *nicht* ankreuzt. Abzüge sind nicht vorgesehen¹⁵.

Dieses System kommt jedoch nicht ernsthaft in Betracht. Man muß kein Mathematiker sein, um einzusehen, daß es sich hier gar nicht um ein Auswertungssystem von multiple choice-Aufgaben handelt, sondern von einfachen wahr/falsch-Aufgaben, die lediglich in Fünferbündeln zusammengefaßt sind. Ob ich fünfmal frage: „Ist die folgende Behauptung wahr?“ mit dem Modus, daß Schweigen als Verneinung gilt und man für jede richtige Antwort einen Punkt erhält, oder ob ich frage:

14 Eine tiefere mathematische Analyse der Ratestrategien und -chancen bei verschiedenen Auswertungssystemen findet sich bei *Dietmar Weber*: „Verschiedene Bewertungssysteme für Multiple-choice-Aufgaben und ihre Ratewahrscheinlichkeiten“. – *Dietmar Weber* hat auch das Saarbrücker ALGOL-Programm zur Testauswertung geschrieben. Die Mühsal des Fliegenbeinzählens in den Aufgabensammlungen hat *Helmut Scheller* auf sich genommen.

15 Vgl. *Friedrich*, Die Funkkolleg-Prüfungen, S 6, 19.

„Welche der folgenden fünf Behauptungen sind wahr?“ wiederum mit dem Modus, daß Schweigen als Verneinung gilt und man für jede richtige Antwort einen Punkt erhält, – das ist lediglich eine Frage der Formulierung. Und es würde sich auch nichts Wesentliches ändern, wenn man die Alternativen zu viert oder zu sechst bündelte.

Wenn es sich hier der Sache nach um wahr/falsch-Aufgaben handelt, so bedeutet das freilich, daß eine Ratechance von 50 % besteht. Tatsächlich ist sie sogar noch etwas höher. Dies beruht darauf, daß es nach den Regeln des Funkkollegs nicht vorkommt, daß in einem Fünferbündel keine Alternative richtig ist oder daß alle oder daß vier richtig sind (wobei mit dieser dritten Möglichkeit, daß genau eine Alternative falsch ist, offensichtlich fünf verschiedene Kombinationen ausgeschlossen sind). Also scheiden 7 von 32 sonst möglichen Werteverteilungen von vornherein aus.

Durch diese Einschränkungen wird die Stellung des ratenden Teilnehmers deutlich verbessert. Wer an einer Prüfung des Funkkollegs teilnimmt, sollte bei Unwissenheit immer nur *eine* Alternative pro Bündel ankreuzen. Auch im schlimmsten Falle ist ihm dann ein Punkt sicher: wenn nämlich drei Alternativen richtig sind und er trotzdem eine der beiden verbleibenden falschen trifft. Und so schlimm kommt es meistens nicht.

Die optimale Ratestrategie besteht auch hier darin, daß man in jeder Fünfergruppe eine Alternative ankreuzt. Dann hat man folgende Ratechancen:

Bei einer zutreffenden Antwort in der Aufgabe erhält man bei einer von fünf möglichen Wahlen fünf Punkte und in den anderen vieren drei Punkte.

$$\frac{1}{5} \times 5 + \frac{4}{5} \times 3 = \frac{17}{5}$$

Beiläufig: wenn nur dieser Fall vorkäme, erhielte man im Schnitt $\frac{17}{5}$ – also 3,4 – von 5 Punkten. Die Ratewahrscheinlichkeit für den einzelnen Punkt betrüge 0,68.

Bei zwei zutreffenden Antworten gibt es zwei Möglichkeiten, vier Punkte zu erlangen, und drei Möglichkeiten zu zwei Punkten.

$$\frac{2}{5} \times 4 + \frac{3}{5} \times 2 = \frac{14}{5}$$

Bei drei zutreffenden Antworten verhält es sich so:

$$\frac{3}{5} \times 3 + \frac{2}{5} \times 1 = \frac{11}{5}$$

Es ist realistisch anzunehmen, daß jeder dieser drei Fälle ungefähr gleich oft vorkommt. Zwar gibt es 10 Möglichkeiten, zwei – ebenso drei – richtige Alternativen in der Fünfergruppe zu verteilen, und nur fünf Möglichkeiten, *eine* zutreffende Alternative unterzubringen. Aber dieser Unterschied ist unerheblich. Wenn eine Aufgabe mehrere richtige Lösungen enthält, liegt es einfach daran, daß dem Aufgabensteller mehrere eingefallen und als didaktisch wichtig erschienen sind, nicht daran, daß sich die Lösungen in bestimmter Weise verteilen lassen. Das Verteilungsproblem ist ganz nebensächlich. Es ist auch kein Grund ersichtlich, daß dem Aufgabensteller häufiger Aufgaben mit drei richtigen Alternativen einfallen als mit zweien oder mit einer. Also sei das arithmetische Mittel aus den durchschnittlichen Punktwerten gebildet:

$$\frac{17}{5} + \frac{14}{5} + \frac{11}{5} = \frac{42}{5}$$

$$\frac{42}{5} \times \frac{1}{3} = \frac{42}{15} = 2,8$$

Die Wahrscheinlichkeit, einen von fünf Punkten zu erlangen beträgt mithin $2,8 : 5 = 0,56$

Man vergegenwärtige sich einmal, was das heißt: In der ersten Klausur des Quadriga Funkkollegs Biologie (1974) wurden 24 Aufgaben zu 5 Alternativen gestellt; man konnte also bis zu 120 Punkte erreichen. Mit 85 Punkten war der Test bestanden; ab 91 Punkten gab es ein „befriedigend“; ab 103 ein „gut“; ab 115 ein „sehr gut“¹⁶.

Wer in dieser Klausur auch nur 60 von 120 Teillösungen wußte, hatte – bei einem Erwartungswert von weiteren 33,6 Punkten (wenigstens 31,2) – gute Aussichten, ein „befriedigend“ zu erlangen¹⁷.

Das Beispiel ist einem Paper entnommen, das das „Deutsche Institut für Fernstudien“ für die Prüfungskommission des 1976/77 gesendeten Funkkollegs „Literatur“ verfaßt hat; die Notengrenzen dürften also repräsentativ sein. Der Fehler liegt ja auch im System und nicht einfach in einer zu großzügigen Benotungspraxis. Wollte man die Notengrenzen merklich nach oben verschieben, so bliebe kaum noch Platz für eine Differenzierung zwischen den Noten, die von ihren Bezeichnungen her gerechtfertigt und statistisch sinnvoll wäre. Immerhin wäre es zu begrüßen, wenn das Funkkolleg wenigstens die Pseudo-mc-Form aufgäbe und zu ehrlich formulierten wahr/falsch-Aufgaben überginge.

Das System des Funkkollegs hat im übrigen den Nachteil, daß es wiederum auf *eine* Aufgabenform angewiesen ist: diesmal darauf, daß mehrere Alternativen richtig sein können (x aus 5). Denn wenn die Alternativen sich erkennbar wechselseitig ausschließen, sind auch dem Teilnehmer, der die falsche trifft, drei von fünf Punkten sicher: ein absurdes Ergebnis. Solche inhaltliche Abhängigkeiten zu vermeiden, verlangt wiederum erhebliche Einschränkungen bei der Aufgabenkonstruktion, und diesmal ist es schwerlich möglich, durch einen logischen Trick die Grenzen der Aufgabenform zu überspielen.

3.1 Das Hagener System: der programmierte Entscheidungskonflikt

Nachzutragen bleibt noch, daß die Fernuniversität Hagen, die ursprünglich nach dem gleichen großzügigen Modus verfuhr (soweit sie auch Mehrfachwahlaufga-

¹⁶ Friedrich aaO S 26 ff.

¹⁷ Mit diesem Ergebnis ist durchaus vereinbar, daß die Chancen, den Test allein durch Raten zu bestehen, also ohne einen bescheidenen Wissenssockel, nur minimal sind. Friedrich aaO S 27 weist darauf hin, daß die Wahrscheinlichkeit, den Schwellenwert von 84 Punkten rein zufällig zu erreichen oder zu überschreiten, lediglich 0,003 % betrage. Dabei geht er zwar von der nicht ganz zutreffenden Voraussetzung aus, daß die Ratewahrscheinlichkeit bei jeder einzelnen Alternative 50 % sei; der Unterschied von einigen Prozenten fällt hier aber nicht ins Gewicht. Daß der Wahrscheinlichkeitswert sehr niedrig sein muß, kann man sich leicht klarmachen, auch ohne förmliche Statistik. (Die zuständige Formel ist die der Binomialverteilung.) Wenn ein Test nur 12 wahr/falsch-Alternativen enthielte, wäre es zwar bei einigem Glück durchaus möglich, 8 bis 9 Richtige zu treffen. Da aber der erwähnte Test 120 Alternativen enthält, müßte man, um ihn zu bestehen, gleichsam zehnmal nacheinander 8 bis 9 Richtige treffen – so viel Glück hat kaum einer. Indessen ist die Voraussetzung einer völligen Unwissenheit bei einem Teilnehmer unrealistisch; das gibt auch Friedrich zu.

ben verwandt hat) neuerdings zu einem restriktiveren Verfahren übergegangen ist¹⁸. Es handelt sich dabei um ein Verfahren, wie es schon vor einigen Jahren einige Saarbrücker Testteilnehmer angeregt haben: Dietmar Weber hat es daraufhin in einem Paper durchgerechnet und dem Arbeitskreis zur Diskussion gestellt.¹⁹

Die Leitidee war die einer Kreuzung zwischen dem Saarbrücker System und dem des Funkkollegs: zu dem Versuch, die psychologischen Nachteile des einen und die statistischen des anderen zu beseitigen. Man bekommt einerseits (wie beim Funkkolleg) nicht nur für die zu recht gewählten, sondern auch für die zu recht nicht gewählten Alternativen einen Punkt (maximal also 5 pro Aufgabe); aber andererseits auch (wie beim Saarbrücker System) für jede fehlerhafte Wahl – und darüber hinaus jetzt auch für jede fehlerhafte Nicht-Wahl! – einen Minuspunkt; wobei wiederum die Punktzahl null pro Aufgabe nicht unterschritten werden kann.

Aber wie es bei Kreuzungen so geht: nicht immer sind es die erwünschten Eigenschaften, die sich durchsetzen, und die unerwünschten, die unterdrückt werden.

Die Ratechancen sind nunmehr zwar reduziert, ebenfalls fast bis zur idealen Grenze von 20 % (nach *Weber*: 24 %); aber die psychische Situation der Teilnehmer ist gegenüber dem Saarbrücker Modell und erst recht dem des Funkkollegs verschlechtert worden. Der nichtwissende Teilnehmer gerät in einen eigenartig-fatalen Konflikt. Beim Saarbrücker System kann er sich wenigstens vor der Drohung der Minuspunkte zurückziehen und lediglich einen Ratepunkt vergeben: vielleicht hat er sogar Glück damit. Bei dem jetzt in Hagen praktizierten System ist ihm solche Zurückhaltung verwehrt: Er muß damit rechnen, daß er wenn keine Minuspunkte für fehlerhafte Wahlen, dann eben für fehlerhafte Nicht-Wahlen erhält.

Es ist freilich denkbar, daß der durchschnittliche Testteilnehmer dies nicht so deutlich empfindet; daß Juristen, wie es die Mitglieder unseres Arbeitskreises zumeist waren, hierin besonders sensibel sind. Aber auch dann hätten sie Recht: Konstellationen des Wie-man's-macht-man-macht's-verkehrt deuten, wenn sie in einer Rechtsordnung auftreten, auf Anomalien hin: auf Fehler der Gesetzgebung oder auf soziale Grenzsituationen, und man sollte sie jedenfalls nicht noch künstlich herstellen.

Literaturverzeichnis

Fernuniversität Hagen, LOTSE-Informationsheft für das Studienjahr 1976/77.

– dasselbe für das Studienjahr 1977/78.

Friedrich, H. F., Die Funkkolleg-Prüfungen, vervielfältigtes Arbeitspapier, August 1974.

Herbig, Manfred, Praxis lehrzielorientierter Tests, Düsseldorf 1976.

Hubbard, John P., Erfolgsmessung der medizinischen Ausbildung, Bern-Stuttgart-Wien 1974.

18 Vgl die LOTSE-Informationshefte der Fernuniversität Hagen: 1976/77 S 9 einerseits und 1977/78 S 16 f andererseits.

19 Es handelt sich um das „System III“ in *Webers* Papier.

- Ingenkamp, Karlheinz, Die Entwicklung von Schultests, in: Tests in der Schulpraxis, hrsg von K. Ingenkamp, Weinheim und Basel 1973, S 43-73.
- Institut für Ausbildungs- und Examensforschung der Medizinischen Fakultät der Universität Bern (iae). Methodische Hinweise für die Konstruktion von Wahlantwort-Fragen (Multiple Choice Items, MC) sowie für die Zusammenstellung einer schriftlichen Prüfung, vervielfältigtes Arbeitspapier, 2. Fassung, Januar 1974.
- Institut für Medizinische Prüfungsfragen. Konstruktion von multiple-choice-Prüfungsfragen – Hinweise für Sachverständige. Vervielfältigtes Manuskript, Mainz November 1973.
- Multiple-choice-Fragen zum Gegenstandskatalog für die Fächer der Ärztlichen Vorprüfung. Anatomie, Herten 1975. Biologie, Herten 1976.
- Philipps, Lothar, Recht und Information, in: Rechtstheorie, hrsg von Arthur Kaufmann, Karlsruhe 1971, S 125-133.
- ders. Testaufgaben in der Rechtswissenschaft. Konstruktionsprinzipien und Auswertung durch den Computer. Karlsruhe 1978.
- Rütter, Theodor, Formen der Testaufgabe, München 1973.
- Schulz, Ulrich, Der Einsatz elektronischer Rechenanlagen zur Erstellung und Auswertung von Wissenstests, in: Münchener Medizinische Wochenschrift (MMW) 1969, S 1410-1416.
- Ucke, Christian, Zur Verwendung von Antwort-Auswahl-Aufgaben bei Prüfungen im Hochschulbereich – eine Analyse. Magisterarbeit an der Technischen Universität München, vervielfältigtes Manuskript, Januar 1976.
- Weber, Dietmar, Verschiedene Bewertungssysteme für Multiple-Choice-Aufgaben und ihre Ratemwahrscheinlichkeiten. Unveröffentlichtes Manuskript, Saarbrücken 1976.