

Förderung
Bayesianischen Denkens
-
Kognitionspsychologische
Grundlagen und
didaktische Analysen

Christoph Wassner

Kassel, Januar 2007

Kurzbeschreibung der Schriftenreihe

In den Kasseler Online-Schriften zur Didaktik der Stochastik (KaDiSto) werden vom Herausgeber und ggf. weiteren Gutachtern geprüfte Materialien publiziert, z.B. Staatsexamensarbeiten, Dissertationen, Berichte von Forschungs- und Entwicklungsprojekten, Unterrichtsmaterialien und „Occasional Papers“, die sich mit der Didaktik der Stochastik und dem Stochastikunterricht beschäftigen. Die Arbeiten werden oft zusammen mit weiteren elektronischen Materialien, z.B. Dateien von Computerprogrammen zur Stochastik verfügbar gemacht.

Die Reihe wurde ins Leben gerufen, um Materialien zu veröffentlichen, die in der Arbeitsgruppe des Herausgebers oder bei Kooperationspartnern in Wissenschaft und Schulpraxis entstanden sind. Die Reihe steht grundsätzlich auch anderen Autorinnen und Autoren offen.

Kurzbeschreibung des Dokuments

Der in dieser Arbeit wesentliche Fokus ist die Realisierung eines anwendungsbezogenen Konzeptes zur Förderung stochastischer Kompetenzen im Mathematikunterricht, die sich auf *Entscheiden und Urteilen unter Unsicherheit* beziehen. Von zentraler Bedeutung ist hierbei die alltagsrelevante Kompetenz, mit Problemen um bedingte Wahrscheinlichkeiten und Anwendungen des Satzes von Bayes umgehen zu können, die i.w.S. mit „*Bayesianischem Denken*“ bezeichnet wird. Die historische und theoretische Grundlage der Arbeit sind kognitionspsychologische Erkenntnisse zum menschlichen Urteilen unter Unsicherheit: Intuitive Formen probabilistischen Denkens basieren auf *Häufigkeitsanschauungen* (z.B. Piaget & Inhelder, 1975; Gigerenzer, 1991). Meine didaktischen Analysen ergaben aber, dass der Umgang mit Unsicherheit im üblichen Stochastikunterricht nach einer häufigkeitsbasierten Einführung des Wahrscheinlichkeitsbegriffes (der ja bekanntlich vielfältige Interpretationsmöglichkeiten aufweist) nur noch auf Basis der numerischen Formate für Wahrscheinlichkeiten (z.B. Prozentwerte, Dezimalbrüche) und entsprechenden Regeln gelehrt wird. Damit werden m.E. grundlegende Intuitionen von Schülern leider nur unzureichend beachtet. Das in dieser Arbeit detailliert entwickelte „*Didaktische Konzept der natürlichen Häufigkeiten*“ schlägt somit die konsequente Modellierung probabilistischer Probleme mit Häufigkeitsrepräsentationen vor.

Auf Grundlage empirischer Laborbefunde und didaktischer Analysen wurde im Rahmen der Arbeit eine Unterrichtsreihe „*Authentisches Bewerten und Urteilen unter Unsicherheit*“ für die Sekundarstufe I entwickelt (Wassner, Biehler, Schweynoch & Martignon, 2004 auch als Band 5 der KaDiSto-Reihe veröffentlicht). Zum einen erfolgte eine Umsetzung des „*Didaktischen Konzeptes der natürlichen Häufigkeiten*“, zum anderen wurde ein Zugang mit hohem Realitätsbezug verwirklicht, in dem so genannte „*allgemeinere Bildungsaspekte*“ wie Lebensvorbereitung, eigenständige Problemlösefähigkeit, kritischer Vernunftgebrauch, Sinnstiftung, motivationale Faktoren etc. wesentliche Beachtung fanden. Die Reihe wurde auch im Rahmen dieser Arbeit in der Sekundarstufe I (fünf 9. Klassen, Gymnasium) implementiert und daraufhin der Unterrichtsgang detailliert bewertet und analysiert.

Diese Arbeit stellt die Dissertation des Verfassers dar, die an der Universität Kassel von Rolf Biehler betreut wurde. Sie ist identisch mit der Erstveröffentlichung 2004 im Franzbecker Verlag, Hildesheim, der der elektronischen Veröffentlichung im Rahmen von KaDiSto zugestimmt hat.

Kasseler Online-Schriften zur Didaktik der Stochastik:

<https://kobra.bibliothek.uni-kassel.de/handle/urn:nbn:de:hebis:34-2006062213595>

Herausgegeben von Rolf Biehler, Fachbereich Mathematik/Informatik, Universität Kassel,

biehler@mathematik.uni-kassel.de

Bd.4:

<http://nbn-resolving.org/urn:nbn:de:hebis:34-2006092214705>

© Christoph Waßner (Nürnberg), ehemals Fachbereich Mathematik/Informatik, Universität Kassel,

c.wassner@gmx.de

Christoph Wassner

**Förderung Bayesianischen Denkens –
Kognitionspsychologische Grundlagen und
didaktische Analysen**

Dissertation

Universität Kassel

Fachbereich Mathematik und Informatik

**Mit Arbeitsmaterialien und didaktischen Kommentaren zum Thema
„Authentisches Bewerten und Urteilen unter Unsicherheit“ für den
Stochastikunterricht der Sekundarstufe I**

von Christoph Wassner, Rolf Biehler, Stefan Schweynoch und Laura Martignon

INHALTSVERZEICHNIS

VORWORT	5
EINFÜHRUNG	7
1 Die Relevanz des Bayesianischen Denkens	7
2 Aufbau und Ziele der Arbeit	11
KAPITEL I	
KOGNITIONSFORSCHUNG ZU BAYESIANISCHEM DENKEN	13
1 Frühe Befunde: „Konservativismus“	14
2 Forschungsprogramm „Heuristics and Biases“	15
2.1 Befunde und Folgerungen zum Bayesianischen Denken	15
2.2 Theoretische Kritik am Forschungsprogramm	18
2.3 Bewertung aus didaktischer Sicht	19
3 Forschungsprogramm „Ökologische Rationalität“	20
3.1 Natürliche Häufigkeiten und „natural sampling“	21
3.2 Grafisch-visuelle Repräsentation	24
4 Empirische Ergebnisse des Forschungsprogramms „Ökologische Rationalität“ zum Bayesianischen Denken	26
4.1 Einfluss des Informationsformates auf Bayesianisches Denken	27
4.1.1 <i>Variation des Informationsformats in verschiedenen Bayesianischen Kontexten</i>	28
4.1.2 <i>Variation der Informationselemente innerhalb der Bayesianischen Situation</i>	31
4.1.3 <i>Verallgemeinerung der Bayesianischen Basissituation</i>	34
4.1.4 <i>Bayesianisches Denken bei Kindern</i>	36
4.2 Einfluss grafischer Darstellung auf Bayesianisches Denken	37
4.2.1 <i>Aktive und passive Verwendung des Häufigkeitsrasters</i>	38
4.2.2 <i>Vergleich verschiedener grafischer Häufigkeitsdarstellungen</i>	39
5 Fazit: Anwendungen der Ergebnisse des Forschungsprogramms „Ökologische Rationalität“ zum Bayesianischen Denken	43
KAPITEL II	
DIDAKTISCHE UND CURRICULARE ANALYSEN ZUM THEMENGEBIET „BEDINGTE WAHRSCHEINLICHKEIT UND BAYESSCHE REGEL“	47
1 Curriculare Diskussion	47
1.1 Aktuelle Einbettung in den gymnasialen Lehrplänen	47
1.2 Aktuelle Empfehlungen aus der Mathematikdidaktik	50
1.2.1 <i>Empfehlungen des Arbeitskreises „Stochastik in der Schule“</i>	50
1.2.2 <i>Empfehlungen aus den „NCTM Standards“</i>	51
1.3 Bewertung	52
2 Mathematikdidaktische Analyse	52
2.1 Traditionelle Zugänge	53
2.2 Innovative Elemente	56
2.2.1 <i>Darstellungsmittel</i>	56
a) <i>Vierfeldertafeln und umgekehrte Baumdiagramme</i>	

	<i>b) Flächendiagramme („Unit-Square“)</i>	
	<i>c) Verbindung von Darstellungsmitteln: Baum-Mengen-Diagramme</i>	
2.2.2	<i>Bayesianischer Ansatz</i>	60
2.2.3	<i>Häufigkeitsbasierte Regelherleitung</i>	62
2.2.4	<i>Chancenkonzept und Konzept des „Begünstigens“</i>	63
2.2.5	<i>Betonung des Realitätsbezugs</i>	65
3	Das didaktische Konzept der „natürlichen Häufigkeiten“	67
3.1	Modellierungsprozess	67
3.2	Realisierung in Lernmitteln	69
4	Vergleichende Bewertung	72

KAPITEL III

TRAINING ZUR FÖRDERUNG BAYESIANISCHEN DENKENS

		75
1	Wahrscheinlichkeitsbasiertes Training	75
1.1	Trainingsbeschreibung	75
1.2	Methodisches Vorgehen	76
1.3	Ergebnisse	76
1.4	Diskussion	77
2	Training mit dem Konzept der „natürlichen Häufigkeiten“	78
2.1	Trainingsbeschreibung	78
2.2	Methodisches Vorgehen	80
2.3	Ergebnisse	80
2.4	Diskussion	82
3	Trainingsvergleiche mit Schülern	82
3.1	Trainingsbeschreibung	83
3.2	Methodisches Vorgehen	83
3.2.1	<i>Stichproben</i>	83
3.2.2	<i>Erhebungsinstrumente</i>	83
3.2.3	<i>Design</i>	86
3.3	Ergebnisse	87
3.3.1	<i>Analyse der Modellierungsprozesse und Operationalisierung</i>	87
3.3.2	<i>Leistungen im Kompetenztest</i>	89
3.3.3	<i>Leistungen differenziert nach Itemtypen bzw. Art der Modellierung</i>	91
3.3.4	<i>Verwendung von Modellierungsformen</i>	94
3.3.5	<i>Leistungsunterschiede differenziert nach Modellierungsschritten</i>	97
3.4	Zusammenfassung und Diskussion	98

KAPITEL IV

REALISIERUNG DES KONZEPTE DER „NATÜRLICHEN HÄUFIGKEITEN“ IM UNTERRICHT DER GYMNASIALEN SEKUNDARSTUFE I

		101
1	Entwurf einer Unterrichtsreihe „Authentisches Bewerten und Urteilen unter Unsicherheit“	103
1.1	Ziele und Einordnung	103
1.2	Inhalte	104
1.2.1	<i>Übersicht</i>	105

1.2.2	<i>Einführung mit einer realen Problemstellung</i>	105
1.2.3	<i>Vertiefung und Erweiterung der Modellierung</i>	108
1.2.4	<i>Vertiefung außermathematischer Kontexte</i>	110
1.2.5	<i>Verallgemeinerung und Begriffsvertiefung</i>	110
1.2.6	<i>Erweiterung des Anwendungsbezuges</i>	112
1.2.7	<i>Übung</i>	112
2	Ziele und Methoden einer Unterrichtsstudie in der Sekundarstufe I	112
2.1	Stichprobe	113
2.2	Ablauf der Unterrichtsstudie	113
2.3	Erhebungsinstrumente	114
2.3.1	<i>Klassenarbeit</i>	114
2.3.2	<i>Kompetenztest zur Überprüfung der Nachhaltigkeit des Lernerfolges</i>	115
2.3.3	<i>Schülerbefragung nach Unterricht</i>	115
2.3.4	<i>Lehrerbefragung nach Unterricht</i>	116
2.3.5	<i>Beobachtung des Unterrichts</i>	116
3	Ergebnisse der Unterrichtsstudie	117
3.1	Klassenarbeit	117
3.1.1	<i>Leistungsverteilung</i>	117
3.1.2	<i>Benotung</i>	119
3.1.3	<i>Leistungen differenziert nach Modellierungsschritten</i>	121
	<i>a) Darstellen in Häufigkeitsbäumen und Häufigkeitstabellen</i>	
	<i>b) Lösen mit Häufigkeitsbäumen und Häufigkeitstabellen</i>	
	<i>c) Bezeichnen von Wahrscheinlichkeiten</i>	
	<i>d) Berechnen (technische Fähigkeiten)</i>	
	<i>e) Vergleich der Modellierungsschritte</i>	
3.1.4	<i>Qualitative Ergänzungen der Leistungsanalyse</i>	125
	<i>a) Lösungsstrategien</i>	
	<i>b) Darstellungsverwendung</i>	
	<i>c) Interpretation als Neubewertung von Wahrscheinlichkeiten</i>	
3.2	Kompetenztest zur Überprüfung der Nachhaltigkeit des Lernerfolges	127
3.2.1	<i>Leistungsstabilität und Vergleich mit Labortrainingsgruppen</i>	127
3.2.2	<i>Qualitative Ergänzungen der Leistungsanalyse</i>	129
	<i>a) Darstellung und Modellierung</i>	
	<i>b) Lösungsstrategien</i>	
3.3	Schülerbefragung	135
3.4	Lehrerbefragung	140
3.4.1	<i>Standardisierte Befragung</i>	140
3.4.2	<i>Nichtstandardisierte Befragung</i>	144
3.5	Unterrichtsbeobachtung	145
3.5.1	<i>Einstieg und erste Modellierung mit Häufigkeitsbäumen</i>	145
3.5.2	<i>Vertiefung und Interpretation</i>	148
3.5.3	<i>Spontaner Anstoß von realitätsbezogenen Fragestellungen</i>	151
3.5.4	<i>Vertiefungen zum Einfluss der Basisrate und Darstellungserweiterung</i>	154
3.5.5	<i>Vergleich zwischen Darstellungsformen und begriffliche Vertiefungen</i>	158
3.5.6	<i>Denkweise des „Lernen aus Erfahrung“</i>	159
4	Zusammenfassung und Ausblick	163

KAPITEL V	
ZUSAMMENFASSENDES FAZIT UND IMPLIKATIONEN	166
LITERATURVERZEICHNIS	169
ANHÄNGE	180
Anhang A. Klassifizierung und Analyse der dichotomen Bayesaufgabe	180
Anhang B. Unterrichtsreihe „Authentisches Bewerten und Urteilen unter Unsicherheit“	182
Anhang C. Klassenarbeit zur Unterrichtsreihe	221

VORWORT

Der Titel mag für eine Arbeit in der Mathematikdidaktik ungewöhnlich erscheinen, weist er doch an erster Stelle auf das Denken und an zweiter Stelle auf das Lehren und Lernen hin. Ich habe ihn bewusst so gewählt, da die Entstehung dieser Arbeit zunächst kognitive Aspekte des Denkens unter Unsicherheit betrifft und schließlich sukzessive immer mehr die Didaktik in den Vordergrund tritt. Demgemäß werde ich mich ausgehend von kognitionspsychologischen Fragestellungen, wie der Mensch unter Unsicherheit urteilt und welche individuellen Denkprozesse beim Umgang mit unsicheren Situationen ablaufen („*Judgment and decision making under uncertainty*“), der Frage nähern, wie der vernünftige Umgang mit Unsicherheit verbessert und gelehrt werden kann. Die Motivation für diese Arbeit ist demnach die Überzeugung, dass die Didaktik der Stochastik von Befunden der psychologischen Urteilsforschung profitieren kann, die von vielen Didaktikern geteilt wird. Andererseits gehe ich davon aus, dass die didaktische Analyse auch der psychologischen Forschung zu neuen Impulsen verhelfen kann. Die Zusammenführung und wechselseitige Befruchtung der Disziplinen ist ein Hauptanliegen meiner Arbeit.

Die Integration von psychologischer und didaktischer Forschung ist auch das Resultat der Genese eines interdisziplinären Forschungsprojektes, in dem ich als Mitarbeiter tätig war. Die Resultate in dieser Arbeit sind Ergebnisse dieses Forschungsprojekts. Das Projekt „*Entscheidungsfindung unter Unsicherheit als fächerübergreifende Kompetenz - Alltagsorientierter Stochastikunterricht am Gymnasium*“ wurde von Prof. Dr. Laura Martignon (Mathematik und Kognitionsforschung) und Prof. Dr. Peter Sedlmeier (Kognitionspsychologie und Methodenlehre) im Rahmen des BIQUA-Schwerpunktes begonnen und zunächst am Max-Planck-Institut für Bildungsforschung in Berlin in der Forschungsgruppe „*Adaptives Verhalten und Kognition*“ von Prof. Dr. Gerd Gigerenzer durchgeführt¹. Seit September 2002 wurde es gemeinsam mit Prof. Dr. Rolf Biehler

¹ Förderung durch die Deutsche Forschungsgemeinschaft seit 1.1.2001 (Geschäftszeichen: Ma 1544 / 10).

(Mathematikdidaktik) an der Universität Kassel weitergeführt. Dadurch entstand der Nährboden für eine integrative, empirische Forschungsarbeit, die von Theorien und empirischen Ergebnissen der Kognitionspsychologie über vergleichende Trainingsstudien bis hin zur Realisierung und Untersuchung eines konkreten Lehr-Lernszenarios im Klassenzimmer reicht.

Um einen abgegrenzten inhaltlichen Rahmen für die von mir durchgeführten empirischen Untersuchungen zu schaffen, habe ich das Bayesianische Denken („*Bayesian Reasoning*“), also den Umgang mit stochastischen Situationen, die wahrheitstheoretisch die Anwendung des Satzes von Bayes (Bayes, 1763) erfordern, herausgegriffen. Das Bayesianische Denken, wie ich anschließend genauer ausführen werde, halte ich dabei als Fokus für schuldidaktische Untersuchungen im Fach Stochastik für besonders relevant. Diese Überzeugung wird in der derzeitigen curricularen Verwirklichung meist nicht geteilt. In diesem Sinne möchte ich auch Argumente dafür liefern, dass eine Betonung und vor allem Vorverlegung dieser Inhalte in allen Stochastikcurricula sinnvoll ist.

Ich danke allen, die mir geholfen haben, dass diese Arbeit entstehen konnte.

Besonderer Dank gilt:

Prof. Dr. Laura Martignon und Dr. Stefan Krauss, die mir den Weg in die wissenschaftliche Welt und den Zugang zur Forschungsarbeit am Max-Planck-Institut für Bildungsforschung in Berlin eröffneten, für eine spannende, kreative und produktive gemeinsame Zeit,

Prof. Dr. Rolf Biehler, der mir den weiteren Weg der Forschung in der Didaktik der Mathematik an der Universität Kassel ermöglichte, für die Unterstützung, Bemühungen und entscheidende Impulse zur didaktischen Umsetzung der Ideen und zum Gelingen dieser Promotionsarbeit,

Prof. Dr. Peter Sedlmeier und Prof. Dr. Gerd Gigerenzer für die vielen Impulse und Ideen zu meiner Arbeit,

Herrn Stefan Schweynoch für die Mitarbeit bei der Erstellung der Unterrichtsreihe und für die Organisation und Durchführung der Unterrichtsstudie an der Schule, sowie den beteiligten Lehrenden Frau Reinhild Niederkleine, Herrn Bernd Gößling, Herrn Michael Hötger und Herrn Friedhelm Weller und allen mitwirkenden Schülerinnen und Schülern des Freiherr-vom-Stein Gymnasiums Bünde (Westfalen),

Herrn Erhard Waßner und Herrn Dr. Helmuth Waßner für literarische Hinweise und Korrekturen.

Der Deutschen Forschungsgemeinschaft danke ich für die finanzielle Unterstützung.

Kassel im April 2004

Christoph Waßner

EINFÜHRUNG

1 Die Relevanz des Bayesianischen Denkens

Seit den 80er Jahren des letzten Jahrhunderts wurde das Thema AIDS sehr stark in der Öffentlichkeit diskutiert, da immer mehr Menschen erkrankten und man die Gefahr einer Weltseuche befürchtete. In der Diskussion stand immer wieder die Einführung obligatorischer Bluttests der gesamten sexuell aktiven Bevölkerung oder speziell benannter Risikogruppen. Eine unkontrollierte Verbreitung in der Bevölkerung sollte vor allem durch Prävention verhindert werden. In Gesundheitsämtern wurden Beratungszentren eingerichtet, über Broschüren und Medien wurden eindringliche Botschaften verbreitet: „Gib AIDS keine Chance – Kondome schützen!“. Zusätzlich wurde zu regelmäßigen „AIDS-Tests“ aufgerufen. Viele ließen sich daraufhin freiwillig testen, um sich über ihren Gesundheitszustand Sicherheit zu verschaffen. Die meisten waren sich aber nicht darüber im Klaren, dass die Ungewissheit über eine eventuelle HIV-Infektion nicht durch den „AIDS-Test“ zweifelsfrei in Gewissheit überführt werden konnte. Sie unterlagen einem fatalen Glauben an den Determinismus wissenschaftlicher Verfahren und Möglichkeiten. Sie vertrauten Ärzten und AIDS-Beratern, von denen sie annahmen, sie würden ihnen endgültige Sicherheit über ihren Gesundheitszustand bringen können.

Diese so vermittelte Sicherheit war leider oft eine Illusion. Das Testverfahren – das gilt sogar bei wiederholtem Testen – kann nicht völlig fehlerfrei HIV-Infektionen aufdecken, vor allem wenn es sich um Personen mit besonders geringem Risikoverhalten handelt. Der Sicherheitsgrad nach einem Test, etwa die Wahrscheinlichkeit HIV-infiziert zu sein, wenn ein Test positiv ausfällt, kann sich als so gering erweisen, dass eine endgültige Diagnose nicht zu treffen ist. Was aber viel schlimmer war: Einem hohen Anteil der medizinischen Fachleute war diese Tatsache auch nicht bewusst! Einer Studie von Gigerenzer, Hoffrage & Ebert (1998) zufolge wurden bei durchgeführten AIDS-Beratungen mit freiwilligen Versuchspersonen ohne besonderes Risikoverhalten alarmierende Fehleinschätzungen über die Verlässlichkeit von Testergebnissen abgegeben. Als Resultat solcher Urteilsfehler sind eine Fülle tragischer Einzelschicksale dokumentiert, die an die Diagnose „AIDS“ glaubten, obwohl sie nicht zutraf (vgl. Gigerenzer, 2002, Kap.7). Heutzutage weiß man, dass viele dieser Schicksale vermeidbar gewesen wären, wäre die Öffentlichkeit und insbesondere medizinische Fachleute besser darüber informiert gewesen, welchen Sicherheitsgrad diagnostische Testergebnisse liefern können.

Die Illusion von Sicherheit kann auch bewusst eingesetzt werden, um politische oder wirtschaftliche Ziele zu verfolgen. In der BSE-Krise der letzten Jahre war „Sicherheit“ ein gern benutztes Wort von Politikern oder dem Handel (vgl. Kurzenhäuser, 2001). Zunächst wurde nach dem Auftreten von BSE-Fällen in Großbritannien deutsches Rindfleisch als „sicher“ deklariert. Später, als auch deutsche BSE-Fälle bekannt wurden (erstmalig am 2.11.2000), argumentierte man mit einer „neuen“ vermeintlichen Sicherheit. Beim Verbraucher wurde z.B. durch selektive Information der Eindruck erweckt, dass mit Hilfe von „BSE-Tests“ kein verseuchtes Fleisch in den Handel gelangen könne. Tatsächlich konnten aber negative Testergebnisse BSE-Freiheit nicht garantieren; die Verfahren – ohne auf Details eingehen zu wollen – konnten durchaus zu fehlerhaften Ergebnissen führen. Als erste Fälle von „falsch-negativen“ Testungen auftraten, war auch diese Sicherheit einzuschränken. Die Folgen waren Unglaubwürdigkeit der Verantwortlichen, Gefährdung, Konfusion und Verunsicherung der Verbraucher, dramatische Umsatzeinbußen der Erzeuger und Händler etc.

Auch diese negativen Konsequenzen für alle Beteiligten zeigen, dass ein verantwortungsvoller Umgang mit (Un-)Sicherheit unverzichtbar ist.

Nun ist womöglich die Frage aufgekommen, was diese „Weltprobleme“ mit einer Arbeit über Didaktik der Stochastik zu tun haben? Ich habe sie deshalb an den Anfang gestellt, da sie in einprägsamer Weise verdeutlichen, warum das Wissen um den Umgang mit Unsicherheit ein unverzichtbarer Bestandteil der allgemeinen Bildung und des Weltbildes eines mündigen Menschen sein muss. Sie führen die *praktische Relevanz* des probabilistischen Denkens vor Augen, das der Mensch ständig – bewusst oder unbewusst – zur vernünftigen Bewertung und Entscheidung von unsicheren Situationen einsetzen sollte: Wie ist eine Alternative (z.B. Vorliegen einer HIV-Infektion) im Lichte von Informationen (z.B. ein Testergebnis) zu bewerten? Welche Entscheidungen (z.B. „Rindfleisch sollte zur Zeit lieber nicht mehr gegessen werden“, „Die Testdiagnose HIV ist nicht sicher, auf jeden Fall muss der Test wiederholt werden“) folgen aus dieser Bewertung? Nach dem entsprechenden wahrscheinlichkeitstheoretischen Kalkül, dem *Satz von Bayes*, wird es oft als *Bayesianisches Denken* („Bayesian Reasoning“), oder auch *Bayesianische Inferenz* („Bayesian inference“) bezeichnet, da von vorliegenden Informationen auf unsichere Hypothesen geschlossen wird. Neben den beispielhaft erwähnten AIDS- oder BSE-Tests finden sich augenfällig viele Anwendungen Bayesianischen Denkens im täglichen Leben. Aus dieser Tatsache leitet sich die Notwendigkeit ab, über die Art und Weise der Förderung vernünftigen Bayesianischen Denkens als Bestandteil der schulischen Ausbildung genauer nachzudenken.

Bevor ich zu einer detaillierteren Ziel- und Inhaltsangabe dieser Arbeit komme, sei ein kurzer Abriss der Historie des für diese Arbeit relevanten stochastischen Denkens gestattet. Er soll zur Verdeutlichung der (*erkenntnis-)**theoretischen Relevanz* Bayesianischen Denkens beitragen. Die Idee einer Art statistischen Inferenz kam, soweit bekannt ist, erstmals 1710 von John Arbuthnot. Er ging mit einer bis dahin geradezu revolutionären Methode an die alte Frage, ob es einen aktiven Gott gibt. Aus der Beobachtung, dass die Sterblichkeit bei jungen Männern höher als bei Frauen ist, leitete Arbuthnot geschickt sein statistisches Argument für die Existenz eines aktiven Gottes und für die Monogamie ab: Wenn Gott Monogamie vorsähe, müsste die Wahrscheinlichkeit einer Jungengeburt höher sein als einer Mädchengeburt. In den Geburtenstatistiken der letzten 82 Jahre fand er dies bestätigt und gab die Wahrscheinlichkeit dieser Beobachtung (Daten) unter der Voraussetzung (Hypothese), es gäbe keinen Gott, sondern nur blinden Zufall, mit $(1/2)^{82}$ an. Heute würde man von einem Likelihood-Argument sprechen. Der Gebrauch statistischer Inferenz wurde erst im letzten Jahrhundert in Form von Signifikanztests vor allem durch Arbeiten von Sir Ronald Fisher, Jerzy Neyman und Egon Pearson angestoßen und die Methoden wurden zum sine qua non wissenschaftlicher Hypothesenprüfung. Die Vorgehensweise wird heute gerne als *klassischer Ansatz* aufgrund einer weitreichenden Institutionalisierung in der Wissenschaft bezeichnet (z.B. Gigerenzer, 1993a).

Eine andere Methode, Hypothesen zu testen bzw. statistische Inferenz zu betreiben, geht auf das bekannte Theorem von Thomas Bayes (1763; Veröff. posthum) zurück und wird heute als *Bayesianischer Ansatz* bezeichnet.² Die grundlegende Denkweise wurde in obigen Beispielen bereits beschrieben. Der Unterschied der Methoden ist prinzipiell folgender: Während klassisch beurteilt wird, ob ein beobachtetes Ergebnis (objektive „Daten“) mit einer als gegeben vorausgesetzten Annahme (der „Nullhypothese“) verträglich ist, wird im

² Zu beachten ist, dass die Ausführungen von Bayes von der modernen Bayesianischen Theorie zu unterscheiden sind, und sogar zu bezweifeln ist, dass Bayes selbst ein Bayesianer im modernen Sinne gewesen ist (Wickmann, 2001, S.123).

Bayesianischen Ansatz eine ursprüngliche, auch möglicherweise subjektiv bewertete Annahme (Hypothese) aufgrund von Beobachtungen (erhobenen Daten) neu gewichtet. Formaler gesprochen ist klassisch die *bedingte Wahrscheinlichkeit* des Auftretens der vorliegenden Daten unter der Annahme der Nullhypothese $P(D | H_0)$ von zentraler Bedeutung, während Bayesianisch die „inverse bedingte Wahrscheinlichkeit“ der Hypothese unter den erhobenen Daten $P(H | D)$ von wesentlichem Interesse ist. Die Leistung von Bayes (und später auch von Laplace) ist der allgemeine Beweis eines Zusammenhangs, der die Berechnung von inversen aus gegebenen Wahrscheinlichkeiten erlaubt, also die Beziehung zwischen $P(H | D)$ und $P(D | H)$ zum Inhalt hat. Im Falle dichotomer Ereignisse lautet der Zusammenhang:

$$P(H | D) = \frac{P(D | H) \cdot P(H)}{P(D | H) \cdot P(H) + P(D | \bar{H}) \cdot P(\bar{H})} \quad (0)$$

Bayes veröffentlichte sein Theorem nicht selbst, da zur Anwendung eine problematische Annahme nötig war: Wenn nichts über die konkurrierenden Hypothesen³ bekannt ist, dann ist ihr Zutreffen a priori gleich wahrscheinlich. Bayes belegte diese Annahme allerdings in seinen Experimenten durch Argumente. Laplace war hierbei sorgloser und hat die Annahme der Gleichverteilung bei Unwissenheit zum Prinzip (des unzureichenden Grundes) erhoben (vgl. Gigerenzer et al., 1999, S.52).

Gerade dieses Prinzip wurde aber bald heftig kritisiert und als *subjektiv*⁴ und damit unwissenschaftlich bezeichnet. In die Bewertung von Hypothesen a priori würden in vielen Situationen zwangsläufig auch qualitative, subjektive Vorstellungen einfließen, die nicht auf objektiven Daten beruhen. Sogenannte „Frequentisten“ (oder: Objektivisten) lehnten Methoden ab, die nicht auf einer objektiven Häufigkeitsinterpretation der Wahrscheinlichkeit basieren. Verschiedene subjektive Wahrscheinlichkeiten zum Ausdruck genau desselben Zustands der Unsicherheit zuzulassen, wird als unadäquat gesehen. Gerade diese Berücksichtigung subjektiver Überzeugungen und qualitativen, personenbezogenen Vorwissens (das zu einer intersubjektiven Größe gemacht wird) ist aber das Paradigma der „Bayesianer“ (oder: Subjektivisten) (vgl. Gigerenzer et al., 1999, S.57 und S.115; Borovcnik, 1992, S.357f.).

„Bayesianer“ warfen wiederum dem klassischen Ansatz vor, dass er Antworten auf wesentliche Fragestellungen nicht liefern und deshalb zu irreführenden Interpretationen verleiten kann. Die Argumentation ist, dass mit einem Signifikanztest die eigentlich interessierende Frage nicht beantwortet wird, *ob* bzw. mit welcher Wahrscheinlichkeit eine Hypothese aufgrund der zur Verfügung stehenden Daten zutrifft. Hingegen wird nur die eher belanglos wirkende Wahrscheinlichkeit bestimmt, mit der eine Hypothese zurückgewiesen wird, *wenn* sie tatsächlich zutrifft (z.B. Buth, 1991). Leider ist der sogenannte klassische Zugang zur Statistik auch alles andere als ein einheitliches Konzept, denn die „Väter“ des klassischen Ansatzes, Fisher (z.B. 1935; 1956) und Neyman & Pearson (z.B. 1933), waren sich über grundlegende Fragen uneinig (vgl. für historische Übersichten z.B. Savage, 1976; Gigerenzer et al., 1999; Barnett, 1999). Als Folge wurden beide Ansätze leider in manchen Wissenschaftsbereichen (z.B. in den Sozialwissenschaften) inkonsistent vermischt und in eine Art „Theorieneintopf“ verwandelt, der heute oft mechanische Anwendung findet (vgl. Gigerenzer et al., 1999, S.230). Vielfach wird beklagt, dass diese hybride Praxis statistischer

³ In der allgemeinen Formulierung von Bayes wurde eine beliebige Menge von Hypothesen betrachtet.

⁴ Die Begriffe „subjektiv“ bzw. „intersubjektiv“ und „objektiv“ werden in diesem Zusammenhang bei Wickmann (1998, S.51f.) genauer geklärt.

Methoden zu einem zwanghaften Ritual geworden ist, das ohne *methodisches Nachdenken* erfolgt und weder von Fischer noch von Neyman und Pearson so akzeptiert worden wäre (vgl. Gigerenzer, 1993b). In anderen Disziplinen (z.B. Mikroökonomie, künstliche Intelligenz, theoretische Physik) spielen wiederum Bayesianische Methoden bereits eine große Rolle, deren Weiterentwicklung im letzten Jahrhundert z.B. durch Arbeiten von Jeffreys (1961), Savage (1954) oder De Finetti (1931/1989) angestoßen wurde.

Das grundsätzliche Problem besteht für die Theoretiker und Anwender darin, dass die verschiedenen Ansätze, klassisch à la Fisher vs. à la Neyman-Pearson vs. Bayesianisch, auch zu verschiedenen praktischen Folgerungen führen können. Fundamentale Kritikpunkte gegen jeweils die anderen Seiten nährten einen Grundlagenstreit in der Statistik, der teilweise apologetisch geführt wurde, aber nicht endgültig entscheidbar scheint (Borovcnik et al., 2001, S.7). So vertreten viele Statistiker der heutigen Zeit ein Zusammenwirken, das zu einem produktiven und progressiven Ergebnis dieser Kontroverse führen soll (z.B. Box, 1983; für eine aktuelle Diskussion: Berger, 2003 mit Kommentaren von Christensen, Johnson, Lavine et al.). Der Aufruf gegen Dogmatismus und zur reflektierten wissenschaftlichen Anwendung der Vielfalt der statistischen Inferenzmethoden oder auch neuer Verfahren der explorativen Datenanalyse (Tukey, 1977) erschallte in den letzten Jahrzehnten immer häufiger, wird aber leider immer noch zu oft ignoriert (z.B. Biehler, 1982; Gigerenzer, 1993a; Sedlmeier, 1996).

Ich habe bewusst einen recht ungewöhnlichen Anfang für eine didaktische Arbeit gewählt, um zu verdeutlichen, welche praktische und theoretische Bedeutung das Bayesianische Denken hat. Welche Folgerungen sind aber für das Lehren und Lernen von Stochastik zu ziehen? Welche *bildungspolitische und didaktische Relevanz* sollte dem Bayesianischen Denken im Rahmen der stochastischen Schulausbildung zukommen? Dazu muss man sich vergegenwärtigen, dass Bayesianisches Denken bisher eine marginale Rolle in Lehrplänen, -büchern oder in der Ausbildung von Mathematiklehrern spielte und Bayesianische Inferenzstatistik so gut wie gar nicht vorkam (vgl. genauer Kap.II.1). Eine eigene Befragung von 40 Mathematiklehrenden zeigte, dass nur 3 offensichtlich mit dem Begriff „Bayesianische Inferenz“ oder „Bayes-Statistik“ konkretere Vorstellungen verbinden konnten. Symptomatisch für die Relevanzeinschätzung Bayesianischen Denkens in der Schule ist meines Erachtens folgende Aussage einer befragten Lehrerin: „Der Satz von Bayes spielt in der Schule eine so untergeordnete Rolle, dass man auf ihn verzichten kann. Man beschäftigt sich höchstens 1-2 Stunden im Leistungskurs damit“.

Im Hinblick auf das Ziel einer allgemeinen Bildung mag diese schuldidaktische Unterbewertung zwar bedauerlich sein, aber sie ist nicht verwunderlich. Sie entspringt derselben Institutionalisierung des klassischen Zugangs wie in manchen Wissenschaftsgebieten. Es gibt sehr wenige ausgearbeitete didaktische Vorschläge, die Bayesianisches Denken unterstützen oder gar in den Mittelpunkt stellen. Beispiele sind ohne Anspruch auf Vollständigkeit Arbeiten von Borovcnik (1984), Riemer (1986) oder Götz (1997). In einigen Lehrbüchern (z.B. Schmid & Weidig, 1996; Cukrowicz & Zimmermann, 2001) wurden solche Ideen auch für den Unterricht bereits aufgegriffen und umgesetzt. Im Großen und Ganzen steht aber die Realität des Stochastikunterrichts an deutschen Schulen ganz im Zeichen der klassisch-frequentistischen Tradition und Bayesianisches „Lernen aus Erfahrung“ fristet weiterhin ein „Schattendasein“ (vgl. genauer Kapitel II).

Wieso passt aber Bayesianisches Denken so gut in einen „modernen“ Mathematikunterricht? In den letzten Jahren wurden Ideen mathematischer Grundbildung neu und umfassend im Rahmen von TIMMS und PISA formuliert. Der Kompetenzbegriff bei PISA fußt auf der „realitätsbezogenen Mathematik“ (didaktische Phänomenologie) von Hans

Freudenthal (1973; 1983). Mathematische Begriffe und Strukturen sind Gedankenobjekte, die reale Phänomene organisieren sollen. Demnach ist die Realität der Ausgangspunkt für mathematische Aktivität. Das Lehren und Lernen muss von der „Phänomenologie mathematischer Begriffe“ ausgehen, nicht von der „fertigen Mathematik“. Bayesianisches Denken zeichnet sich, wie durch obige Beispiele angedeutet wurde, gerade durch Nähe zu lebensrelevanten Fragestellungen und praktische Bedeutung aus. Bayesianisches Vorgehen wurde geradezu durch den Druck der Praxis (z.B. in der Physik oder Medizin) erzwungen, da bei kleinen Stichproben klassische Methoden für vernünftige und genaue Aussagen nicht mehr ausreichten bzw. die frequentistische Deutung zu eng abgesteckte Grenzen aufwies.

Nach diesem „modernen“ Verständnis von mathematischer Bildung müssen im Unterricht vielfältige Gelegenheiten zum Durchlaufen eines *Prozesses des mathematischen Modellierens* geschaffen werden (vgl. z.B. Schupp, 1984; Blum, 1996; Klieme, Neubrand & Lüdtke, 2001), der neben der Mathematisierung und innermathematischen Verarbeitung realer Situationen auch Interpretation und Validierung des Modells beinhaltet. Das wechselseitige Verhältnis von „Mathematik und Welt“ (z.B. Heymann, 1995) stellt in dieser Sicht die wesentliche Herausforderung an den Mathematikunterricht dar. Bayesianisches Denken bietet vielerlei Möglichkeiten, den Prozess des Lernens aus Erfahrung und des Interpretierens und Validierens unter Einbeziehung einer subjektiven Komponente im Unterricht zu integrieren (z.B. Borovcnik, 1992; Wickmann, 1998; Götz, 2001). Ohne auf Ergebnisse meiner Untersuchungen vorgreifen zu wollen, möchte ich bereits an dieser Stelle betonen: Bayesianisches Denken in adäquater Form eignet sich hervorragend, Lernende dazu zu bringen, Vorgehensweisen und Ergebnisse kritisch zu betrachten und sie eher als Entscheidungshilfen denn als Entscheidungen anzusehen (vgl. Schupp, im Druck).

2 Aufbau und Ziele der Arbeit

Motivation zu dieser Arbeit ist der Entwurf eines Zugangs zu Bayesianischem Denken, der Lernenden die Einsicht ermöglichen soll, wie Mathematik eingreift und wofür sie gebraucht wird. In den bisherigen Ausführungen wurde bereits deutlich, welche Grundhaltung mit diesem Zugang verbunden wird. Das Bedeutungsdefizit in der schulischen Stochastikausbildung und der damit einhergehende Mangel an systematischen Untersuchungen zu diesem Thema sollen durch diese Arbeit ausgeglichen werden.

Meine persönliche Motivation zur intensiven Beschäftigung mit dem Thema kam aus der kognitionspsychologischen Diskussion. Mathematiker und Philosophen der Aufklärung (z.B. Pierre Simon de Laplace) sahen den gesunden Menschenverstand des „homme éclairé“ und die Wahrscheinlichkeitsrechnung im Einklang. In den siebziger Jahren des letzten Jahrhunderts kamen jedoch Psychologen um Daniel Kahneman und Amos Tversky bei der Erforschung menschlichen Denkens unter Unsicherheit zur entgegengesetzten Überzeugung, dass der Mensch oft systematischen, kaum zu verändernden Fehlvorstellungen (z.B. beim Bayesianischen Denken) unterliegt, da er sich auf eine kleine Anzahl von Heuristiken verlässt. Diese Auffassung – meist unter dem Schlagwort „Heuristics & biases“ – manifestierte sich auch teilweise in der didaktischen Literatur. Erkenntnisse der Kognitionspsychologie der neunziger Jahre widerlegten diese Einschätzung weitgehend (z.B. Gigerenzer & Hoffrage, 1995). Unter Verwendung geeigneter Informationsformate gehen Menschen, z.B. beim Bayesianischen Denken, sehr kompetent mit Unsicherheit um. In *Kapitel I* der Arbeit werden diese Kontroverse in der *psychologischen Urteilsforschung* und

ihre Auswirkungen genauer dargestellt sowie Ergebnisse kognitionspsychologischer Studien zum Bayesianischen Denken berichtet.

Auch in der *Didaktik der Wahrscheinlichkeitsrechnung* wurde Bayesianisches Denken, als Inhaltsbereich „Anwendung der bedingten Wahrscheinlichkeit und des Satzes von Bayes“, unter Berücksichtigung der psychologischen Urteilsforschung kontrovers diskutiert. Didaktische Ansätze wurden aber meist theoretisch und nicht empirisch begründet. In der Schulrealität nimmt Bayesianisches Denken noch immer eine Außenseiterrolle ein, wie auch in Lehrplänen und -büchern festzustellen ist. In **Kapitel II** werden die curriculare Einbettung und mathematikdidaktische Zugänge zum Thema ausführlich dargestellt und analysiert. Ziel eines eigenen Vorschlags ist, neuere kognitionspsychologische Befunde für die Didaktik der Stochastik zu erschließen.

Kapitel III beschäftigt sich mit Möglichkeiten zum *Training Bayesianischen Denkens*. Auf Basis der empirischen Ergebnisse von Gigerenzer & Hoffrage (1995) zur Bedeutung des Informationsformats bei Bayesianischem Denken, entwickelte Sedlmeier (1997) ein computergestütztes Trainingstutorial auf Basis von Häufigkeitsdarstellungen, das in Trainingsstudien mit Universitätsstudenten im Vergleich mit anderen Varianten sehr erfolgreich eingesetzt wurde. Dieses Ergebnis ist umso bemerkenswerter, da die Wirksamkeit von Training zu Bayesianischem Denken vorher nicht deutlich gezeigt werden konnte oder gänzlich bestritten wurde. Ich habe daraufhin ähnliche Trainingsvergleiche auch bei Schülern der Sekundarstufe I und II durchgeführt. Die Trainingsstudien und erzielte Ergebnisse werden ausführlich diskutiert.

Die Resultate der Trainingsstudien mit Schülern und zeitgemäße Grundpositionen zur mathematischen Bildung (Realitätsbezogenheit, Lebensrelevanz etc.), die bereits einführend angedeutet wurden, wurden mit Unterstützung von Rolf Biehler, Stefan Schweynoch und Laura Martignon in einer *Unterrichtsreihe* zur Einführung Bayesianischen Denkens in der Sekundarstufe I didaktisch umgesetzt. Der Unterrichtsvorschlag wurde in fünf Klassen der Jahrgangsstufe 9 als Unterrichtsgang in Wahrscheinlichkeitsrechnung realisiert. Eine Studie untersuchte die Realisierung und Bildungswirkungen mit qualitativ-interpretativen und quantitativen Methoden. Die genaue Darstellung der Unterrichtsreihe und der Erhebungsmethoden, Ergebnisse und Folgerungen der Unterrichtsstudie erfolgen in **Kapitel IV**. Anregungen für künftige Forschungsarbeiten werden gegeben.

In **Kapitel V** werden schließlich die zentralen Ergebnisse der empirischen Untersuchungen noch einmal zusammengefasst und vor dem Hintergrund der aktuellen Diskussion zur mathematischen Bildung bewertet. Implikationen für die schulische Stochastikausbildung beschließen diese Arbeit.

Im Anhang ist insbesondere die noch unveröffentlichte Unterrichtsreihe „Authentisches Bewerten und Urteilen unter Unsicherheit“ von Wassner, Biehler, Schweynoch und Martignon als Arbeitsblattsammlung mit didaktischen Kommentaren vollständig abgedruckt.

KAPITEL I

KOGNITIONSFORSCHUNG ZU BAYESIANISCHEM DENKEN

„The theory of probability is at bottom nothing more than good sense reduced to a calculus which evaluates that which good minds know by sort of instinct, without being able to explain how with precision”

PIERRE SIMON DE LAPLACE (1814/1951, S.196)

Die Frage, der in diesem Kapitel nachgegangen wird, hat eine lange und bewegte Geschichte: Inwieweit entspricht die menschliche Intuition den Gesetzen der Wahrscheinlichkeit, insbesondere bei Bayesianischem Denken? Genauer ausgedrückt: Haben Menschen kognitive Mechanismen, um gut mit Bayesianischen Situationen umgehen zu können und wenn ja, welche? Ich gebe im Folgenden einen historischen Überblick über verschiedene wissenschaftliche Standpunkte und Antworten, die bis heute gegeben wurden. Die Frage wurde im Kontext der psychologischen Urteilsforschung (unter Unsicherheit) meist unter dem Schlagwort „Bayesian Reasoning“ untersucht, das in dieser Arbeit mit dem Begriff „Bayesianisches Denken“ übersetzt ist.⁵

Von Mathematikern und Philosophen der Aufklärung (z.B. M.J. Condorcet, S. Denis Poisson, Pierre S. de Laplace, siehe obiges Zitat) wurde die gerade einmal 100 Jahre alte Wahrscheinlichkeitstheorie zum Inbegriff des gesunden Menschenverstandes erhoben. Mathematische Theorie und das Denken eines vernünftigen Menschen („*homme éclairé*“) wurden als zwei Seiten derselben Medaille angesehen. Die verfügbaren Werkzeuge der klassischen Wahrscheinlichkeitstheorie, wie z.B. das Theorem von Bayes, waren Beschreibung und Norm vernünftigen Urteilens zugleich (Daston, 1988). Mit dieser Gleichsetzung von Theorie und Vernünftigkeit ging auch eine Verschmelzung von subjektiver und objektiver Wahrscheinlichkeitsinterpretation einher. Das Kalkül wurde durch die Anwendungen und umgekehrt bestimmt (Gigerenzer et al., 1999, S.39).

Diese positive Einschätzung der Ausgangsfrage wurde im Laufe der Zeit relativiert. Beschreibung und Norm der Vernünftigkeit drifteten mehr und mehr auseinander, objektiv und subjektiv verstandene Wahrscheinlichkeit wurden unterschieden. Im 19. Jahrhundert wich die Vision der Vernünftigkeit des subjektiven Urteils dem Streben nach objektiven Gesetzmäßigkeiten (z.B. Adolphe Quetelet). Durch die wachsende Verfügbarkeit von sozialstatistischen Daten war das Zeitalter der Statistik als numerischer Wissenschaft angebrochen und die Anwendungen der Wahrscheinlichkeitstheorie basierten nun auf beobachteten Häufigkeiten und Mittelwerten, also rein auf Tatsachen und nicht mehr auf Vermutungen oder Intuitionen. Allerdings widersprach diese Regelmäßigkeit nicht unbedingt der menschlichen Intuition, wie bereits Jacob Bernoulli in einem Brief an Leibniz zum Ausdruck brachte, dass sein Gesetz der großen Zahlen „auch der dümmste Mensch mittels eines naturgegebenen Instinkts von allein und ohne vorherige Instruktion weiß“ (Leibniz, 1703, Bd.3 Teil I, S. 77f.).

Die philosophischen und mathematischen Standpunkte und Diskussionen um gesunden Menschenverstand, Intuition oder Naturinstinkt zeigen die Nähe zu psychologischen Fragestellungen zur Kognition des Menschen. Mitte des 20. Jahrhunderts haben Psychologen begonnen, menschliches Denken unter Unsicherheit auch empirisch zu untersuchen und begründete Antworten auf obige Fragen zu finden. Die folgenden

⁵ Dieser Überbegriff wird im Folgenden i.w.S. für jegliches Umgehen mit Bayesianischen Situationen bzw. entsprechenden Textaufgaben benutzt.

Ausführungen geben einen Überblick über die empirische Erforschung des Urteilens unter Unsicherheit – Bayesianisches Denken fokussierend – von Seiten der Psychologie, deren Befunde und Folgerungen wechsellvoller nicht sein könnten.

1 Frühe Befunde: „Konservativismus“

Mitte der sechziger Jahre des letzten Jahrhunderts wurde Bayesianisches Denken erstmals systematisch untersucht. Die Motivation von Ward Edwards und anderen Psychologen (William L. Hays, Lawrence D. Phillips) entstand aus der Frage, ob das menschliche Gehirn Informationen (bis zu einem gewissen Komplexitätsgrad) gemäß Bayesianischer Inferenz verarbeitet. Vorher (bei Gestaltpsychologen wie Karl Duncker, George Humphrey etc.) hatten in der experimentellen Denkforschung solche Fragen keine Rolle gespielt, sie waren nicht Gegenstand des Interesses (z.B. Humphrey, 1951). Das Gehirn als „intuitiven Statistiker“ zu beschreiben („man as an intuitive statistician“: Peterson & Beach, 1967), war erst jetzt, da auch Statistik methodisch immer mehr an Bedeutung gewann, eine treibende Idee in der Kognitionspsychologie geworden; in diesem Falle unter der Bezeichnung „human information processing“ (Edwards, 1968).

Nach Experimenten zur Beurteilung von Radarbildern⁶, die sehr komplexe Anforderungen an Versuchspersonen stellten, formulierten Edwards und Kollegen einfachere, typische Situationen, die zum Standard solcher „Bayesian Reasoning“-Experimente wurden, z.B.:

2 Urnen sind mit einer großen Zahl Kugeln gefüllt. Die erste Urne enthält 70% rote und 30% blaue Kugeln. Die zweite enthält 70% blaue und 30% rote Kugeln. Nun wird zufällig eine der beiden Urnen ausgewählt (z.B. durch Münzwurf). Dann werden eine Reihe Kugeln aus der gewählten Urne gezogen, z.B. 8 rote und 4 blaue Kugeln. Geschätzt werden soll die Wahrscheinlichkeit, dass die gewählte Urne die mit den überwiegend roten Kugeln ist. (vgl. Peterson & Beach, 1967, S.32)

Ein generelles Ergebnis war, dass Versuchspersonen die A-priori-Wahrscheinlichkeit (hier: 50%) weniger revidierten, als die normative Lösung (hier: 97%) mit dem Bayesschen Theorem ergeben würde.⁷ Phillips und Edwards (1966; Edwards, 1968) kamen durch ihre Untersuchungen zum Schluss, dass Menschen sehr gut zu intuitivem Bayesianischem Denken fähig sind, wobei sie aber tendenziell neue Informationen „konservativ“, d.h. geringer als nötig, in die Gewichtung einbeziehen (vgl. Gigerenzer et al., 1999, S.239). Insbesondere schien der Einfluss von A-priori-Wahrscheinlichkeiten unter einigen Bedingungen, wie z.B. reale Erfahrung des Zufallsexperimentes oder realitätsbezogenere Problemstellung, eher beachtet zu werden (vgl. Sedlmeier, 1999, S.13). Diese Befunde führten auch zu ersten *Trainingsversuchen*, den festgestellten „Konservativismus“ durch korrekatives Feedback zu überwinden (Peterson et al., 1968, Exp. 2; Schaefer, 1976; Lindeman et al., 1988). Die Trainingseffekte blieben gleichwohl sehr gering: Versuchspersonen zeigten immer noch deutlich konservatives Verhalten (vgl. ausführlicher Sedlmeier, 1999, S.51f.).

„Die Metapher [von Edwards und Kollegen, Anm. d. Verf.], die Denkprozesse als Bayesianische Prozesse beschreibt, löste ein überaus folgenreiches Forschungsprogramm aus – vergleichbar dem, das durch die Analogie zwischen Denken und Computer in Gang gesetzt wurde“ (Gigerenzer et al., 1999, S.238).

⁶ 4 Hypothesen mit A-priori-Wahrscheinlichkeiten (z.B. Feindangriff, Meteoritenhagel etc.) sollten aufgrund von 12 möglichen Beobachtungen auf dem Radarschirm revidiert werden.

⁷ Formal: Die Neubewertung der Hypothese (H) mit neuen Daten (D), also die Schätzung der A-posteriori-Wahrscheinlichkeit $P(H | D)$ bei bekannter A-priori-Wahrscheinlichkeit $P(H)$, liegt also an einer systematischen Unterbewertung des Einflusses der Likelihood-Wahrscheinlichkeit $P(D | H)$.

2 Forschungsprogramm „Heuristics and Biases“

Das Bild vom „intuitiven Statistiker“ blieb nicht lange unbestritten. Von Daniel Kahneman und Amos Tversky wurden um 1970 Versuchspersonen strukturell analoge Problemvariationen wie bei Edwards in realitätsbezogenem Kontext vorgelegt, wie z.B. das bekannte „Taxiproblem“:

Ein Taxi war an einem nächtlichen Verkehrsunfall mit Fahrerflucht beteiligt. In der Stadt, in der der Unfall passierte, gibt es zwei Taxiunternehmen, eines mit grünen und eines mit blauen Taxis. Folgende Daten sind nun gegeben: (i) 85% aller Taxis in der Stadt sind blau, die anderen 15% sind grün. (ii) Ein Zeuge identifizierte das davorfahrende Taxi als „grün“. Das Gericht untersuchte nun die Fähigkeit des Zeugen, die Farbe eines Taxis bei Nacht richtig zu identifizieren. In der Versuchsreihe (die Hälfte der Taxis war blau, die andere Hälfte war grün), konnte der Zeuge beide Farben zu 80% richtig identifizieren, aber zu 20% irrte er.

Wie groß ist die Wahrscheinlichkeit, dass das Unfalltaxi wirklich grün und nicht blau war? (vgl. Kahneman & Tversky, 1973⁸)

Aus ihren Befunden, dass im Mittel Personen die Wahrscheinlichkeit mit 80% schätzen, aber 41% gemäß Bayesregel korrekt wäre, folgerten sie im Grunde das Gegenteil zum „Konservativismus“: 80% entspricht der gegebenen Likelihood-Wahrscheinlichkeit korrekter Identifikation, also scheint die A-priori-Information oder Basisrate von 15% grünen Taxis systematisch nicht beachtet zu werden. Diese Vernachlässigung der Basisrate („*Base-rate neglect*“) bei Bayesianischem Denken wurde mit einem „*Heuristikansatz*“ zu erklären versucht, dessen Credo schließlich weithin akzeptiert war (Kahneman & Tversky, 1973, S.237): „In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics that sometimes yield reasonable judgments and sometimes lead to severe and systematic errors“. 150 Jahre nach Laplace hatte sich die positive Bewertung der Eingangsfrage ins Gegenteil verkehrt. Heuristiken dienten in der Folge als theoretische Erklärung für eine Fülle von identifizierten systematischen Diskrepanzen („*fallacies*“) zwischen der Urteilsbildung und der Wahrscheinlichkeitstheorie (Kahneman, Slovic & Tversky, 1982; für eine Übersicht mit didaktischer Perspektive vgl. Scholz, 1981). Im Folgenden soll vor allem auf Befunde und Folgerungen zur Vernachlässigung der Basisrate bei Bayesianischen Beurteilungssituationen näher eingegangen werden.

2.1 Befunde und Folgerungen zum Bayesianischen Denken

„*Base-rate neglect*“ war wohl das meist untersuchte Phänomen im Zusammenhang mit stochastischem Denken. Das „Taxiproblem“ ist nur ein bekanntes Beispiel, medizinisch-diagnostische Situationen sind andere (z.B. Eddy, 1982):

Eine symptomfreie 45-jährige Frau geht zur Krebsvorsorgeuntersuchung und lässt sich mittels einer Mammografie untersuchen. Man weiß aus der Gesundheitsstatistik:

Die Wahrscheinlichkeit, dass eine symptomfreie Frau zwischen 40 und 50 Jahren Brustkrebs hat, beträgt 1%.

Die Wahrscheinlichkeit, dass diese Krankheit mit einer Mammografie erkannt wird, beträgt 80%.

Die Wahrscheinlichkeit, dass eine Mammografie fälschlich auf das Vorliegen von Brustkrebs hinweist, obwohl die Krankheit gar nicht vorliegt, beträgt 9,6%.

Wie groß ist die Wahrscheinlichkeit, dass eine 45-jährige, symptomfreie Frau tatsächlich Brustkrebs hat, wenn sie einen positiven Mammografiebefund erhalten hat?

⁸ Diese ursprüngliche Version wurde aufgrund von Kritik an sprachlichen Einzelheiten auch mehrfach modifiziert.

In einer derartigen Situation, zumindest wenn sie in dieser Form beschrieben ist, wird von den meisten Menschen und sogar von Ärzten, von denen man richtige Diagnosen erwartet, spontan eine Wahrscheinlichkeit von 70 bis 80% geschätzt (vgl. Eddy, 1982).

Die wahrscheinlichkeitstheoretische Lösung mit der Formel von Bayes ist aber eine völlig andere:

$$P(B | M_+) = \frac{P(M_+ | B) \cdot P(B)}{P(M_+ | B) \cdot P(B) + P(\bar{M}_+ | B) \cdot P(B)} = \frac{80\% \cdot 1\%}{80\% \cdot 1\% + 9,6\% \cdot 99\%} = 7,8\% \quad (1)$$

(B: „Brustkrebs“; M₊: „positiver Mammografiebefund“)

Es finden sich sehr viele Untersuchungen in der Literatur, die eine derartige Diskrepanz dokumentieren und als systematische Urteilsfehler („bias“) beschreiben (vgl. Übersichten bei Bar-Hillel, 1980; Borgida & Brekke, 1982).⁹ Die Folgerungen aus diesen Befunden für menschliche Rationalität, zumindest in Bezug auf die Vision eines intuitiven Statistikers, waren sehr negativ und niederschmetternd. Kahneman und Tversky (1972) kamen zum Schluss, dass „solche Resultate jeden entmutigen müssen, der die Menschen als vernünftige intuitive Statistiker ansehen möchte“. In Bezug auf Bayesianisches Denken wurde z.B. konstatiert: „The genuineness, the robustness and the generality of the base-rate fallacy are matters of established fact“ (Bar-Hillel, 1980, S.215).

Die Frage war aber nun, welche *Erklärungen* für dieses Phänomen „Base-rate neglect“ gegeben werden können? Die Heuristikerklärung von Kahneman und Tversky war zunächst die sogenannte „Repräsentativität“. Demgemäß wird ein Ereignis wahrscheinlicher als ein anderes eingestuft, wenn es repräsentativer für einen Zufallsprozess erscheint („by the degree to which it, [...] reflects the salient features of the process by which it is generated“, Kahneman & Tversky, 1972, S.431). Ein einfaches Beispiel ist ein Verhalten von Roulette-Spielern, die nach einer langen Serie von „rot“ argumentieren, dass nun „schwarz“ mit höherer Wahrscheinlichkeit auftritt, da ein Ausgleich der Auftretenshäufigkeiten beider Farben erfolgen muss („gambler’s fallacy“). „Schwarz“ wird für „repräsentativer“ gehalten. Übertragen z.B. auf das Mammografieproblem würde das bedeuten, dass für die Beurteilung $P(M_+ | B)$ die hervorspringenden Merkmale des Prozesses der Stichprobenziehung besser als $P(B)$ widerspiegelt. Die Frage bleibt aber, warum? Wie Gigerenzer et al. (1999, S.246) deutlich machen, kann so allenfalls das Phänomen neu beschrieben, aber nicht genauer erhellt werden (vgl. zur Kritik ausführlicher 2.2 in diesem Kapitel).

Grundsätzlich wurde „Base-rate neglect“ auch oft auf prinzipielle Schwierigkeiten von Menschen beim Umgang mit bedingten Wahrscheinlichkeiten zurückgeführt, die z.B. in der Verwechslung von Kausalität und Konditionalität, der exakten Definition der Bedingung oder der Verwechslung von bedingendem und bedingtem Ereignis¹⁰ liegen könnten (Bar-Hillel, 1983; Falk, 1988; vgl. für eine Übersicht: Bea, 1995, S.41f.). Bei Eddy (1982) z.B. wird berichtet, dass bei medizinischen Diagnosen Ärzte oft die Wahrscheinlichkeit für eine Krankheit bei positivem Befund mit der Wahrscheinlichkeit für einen positiven Befund bei tatsächlich vorliegender Krankheit verwechselten. Vielfach wird diese „Verwechslung der Bedingung“ auch mit der einleitend erwähnten Hypothesentest-Konfusion in Verbindung gebracht. $P(\text{Daten} | \text{Hypothese})$ bzw. $P(\text{Hypothese} | \text{Daten})$ liegt möglichen Aussagen aus klassischen bzw. Bayesianischen Hypothesentests zugrunde und die Verwechslung führt z.B.

⁹ Allerdings sind Experimente mit dem Befund fehlerhaften stochastischen Denkens nicht viel häufiger als jene mit dem Befund guten stochastischen Denkens, werden aber laut Christensen-Szalanski & Beach (1984) 6-mal häufiger zitiert.

¹⁰ In der Hypothese-Daten-Terminologie: Verwechslung von $P(D|H)$ mit $P(H|D)$

zu Fehlinterpretationen von Signifikanz (vgl. z.B. Falk, 1986; Shaughnessy, 1992). Arbeiten in der didaktischen Literatur zum Problem der Interpretation von Signifikanz beinhalten auch Empfehlungen zur Überwindung von Fehlvorstellungen bei Schülern (z.B. Krauss & Wassner, 2001; vgl. Kap.II, 2.2.2). Jedoch bietet die Verwechslungsthese sicher keine ausreichende Erklärung für das Auftreten von „Base-rate neglect“, wie noch deutlich werden wird.¹¹

Systematische *Trainingsversuche* zur Verbesserung Bayesianischen Denkens wurden im „Heuristik“-Forschungsprogramm nur wenige unternommen. Das liegt nicht zuletzt an den Folgerungen, die immer wieder eine Analogie zwischen visuellen und kognitiven Täuschungen betonten (z.B. Tversky & Kahneman, 1974), wodurch die Unüberwindbarkeit der Fehlurteile durch Training bereits prognostiziert wurde. Einer der wenigen Trainingsversuche arbeitete mit Fokussierung (Fischhoff, Slovic & Lichtenstein, 1979). Man versuchte, die Lernenden stärker für den Einfluss von Basisraten zu sensibilisieren, indem sie innerhalb eines Problems variiert wurden¹². So wurde z.B. beim Taxiproblem, anstatt einer Basisrate von 15% grünen Taxis, die gleiche Aufgabe derselben Person auch mit 2% und 98% Basisrate gestellt. Es gab aber kein korrekatives Feedback. Tatsächlich zeigte sich ein Sensibilisierungseffekt bei den wiederholten Wahrscheinlichkeitsschätzungen zu dieser Aufgabe, jedoch keine generelle Sensibilisierung für den Einfluss der Basisrate bei gleichstrukturierten Bayesproblemen mit anderen Aufgabenkontexten. In folgenden Experimenten (Fischhoff & Bar-Hillel, 1984) wurden weitere Fokussierungstechniken angewandt. Z.B. sollten Versuchspersonen die Lösung von Bayes-Problemen immer wieder schätzen, wobei zuerst nur eine der Informationen in Betracht gezogen werden soll, dann sukzessive die weiteren. Erst danach sollte die Gesamtschätzung der Lösungswahrscheinlichkeit vorgenommen werden („isolation analysis“). Nach Fokussierungstraining zeigte sich tatsächlich eine höhere Beachtung der relevanten Informationen. Allerdings wurden nun auch für die Lösung irrelevante Informationen stärker beachtet. Das lässt darauf schließen, dass keine Erhöhung eines konzeptuellen Verständnisses, sondern lediglich eine mechanisierte Sensibilisierung für jegliche Art von (numerischen) Informationen erfolgte.

Insgesamt waren die Trainingsversuche mit Fokussierung auf relevante Informationen kaum erfolgreich im Hinblick auf eine allgemeine Verbesserung Bayesianischen Denkens. Die Missachtung von Basisrateninformationen im Sinne einer Untergewichtung konnte nicht erfolgreich behoben werden (vgl. für einen Überblick Sedlmeier, 1999, S.54). Diese Befunde schienen aber sehr gut verträglich mit dem theoretischen Konzept. Dawes (1988, S.142) drückte es so aus: „Attempts to train people not to think representatively and not to be influenced by availability or other biases have not been very successful, perhaps because it is impossible for us to think in a way we do not think“. Von diversen Vertretern des Heuristikansatzes wurde eine Effektivität von Training zum Urteilen unter Unsicherheit in ähnlicher Weise negiert (z.B. Gould, 1992; Piatelli-Palmarini, 1994). Diese sehr rigiden Folgerungen für das menschliche Denken und insbesondere für seine Lernfähigkeit boten Anlass zu weitreichender Kritik. In Kapitel III dieser Arbeit werden Trainingsstudien zu Bayesianischem Denken diskutiert, die den Folgerungen des Heuristikansatzes widersprechen.

¹¹ Bei Untersuchungen mit Studenten von Gigerenzer & Hoffrage (1995) trat diese „Verwechslung“ zwar ebenfalls auf, aber nur als *eine von vielen* Strategien (u.a. auch Verwechslung von $P(A|B)$ mit $P(A \cap B)$, vgl. 4.1.1.).

¹² wurde auch später als „Subjective Sensitivity Analysis (SSA)“ bezeichnet.

2.2 Theoretische Kritik am Forschungsprogramm

Das „Heuristics & Biases“-Forschungsprogramm und seine rigiden Folgerungen für das menschliche Denken unter Unsicherheit wurden vielfach sowohl theoretisch kritisiert als auch empirisch widerlegt. Ausführungen zu empirischen Ergebnissen mache ich später noch (vgl. 3.). Doch schon bevor die Empirie Gegenbeweise erbrachte, wurden diverse Schwächen auf methodologischer und normativer Ebene kritisiert, auf die ich im Folgenden ohne Anspruch auf Vollständigkeit kurz eingehen will (vgl. kritische Diskussionen bei Lopes, 1991, Koehler, 1996; mit didaktischem Fokus bei Scholz, 1981).

Bereits angesprochen wurde das Problem, dass die zur Erklärung dienenden Heuristiken, wie z.B. die Repräsentativität, zu vage und unspezifisch sind, um die tatsächlich zugrundeliegenden kognitiven Prozesse zu erklären. Auch fehlen jegliche Anhaltspunkte, wann und unter welchen Bedingungen man einer Heuristik unterliegt. Die Heuristiken wurden vielfach isoliert voneinander und post hoc zur Erklärung herangezogen mit der Folge, dass jeglicher Befund, auch konträrer Art wie „Base-rate neglect“ und Konservatismus, letztlich zu einer Heuristik „passt“ (Wallsten, 1980; Gigerenzer, 1996). Die wenigen Heuristiken sind aber keinesfalls ausreichend zur Analyse der kognitiven Strategien (z.B. Scholz, 1981; Wallsten, 1983). „In place of plausible heuristics that explain everything and nothing – not even the conditions that trigger one heuristic rather than another – we will need models that make surprising (and falsifiable) predictions and that reveal the mental processes that explain both valid and invalid judgment“ (Gigerenzer 1996, S.595).

Eine andere Kritik richtet sich gegen die zu enge Festlegung von „normativen“ Kriterien, die sich ausschließlich an wahrscheinlichkeitstheoretischen Lösungen orientieren. Abweichungen von diesen Kriterien wurden im Sinne von „bias“ als kognitive Defizite betrachtet. Diese engen Bewertungsmaßstäbe berücksichtigen z.B. nicht die Kontextabhängigkeit intuitiver, menschlicher Informationsverarbeitung. Gigerenzer (1996, S.593) führt an: „Content-blind norms are appropriate for textbook problems in probability theory, where the content is only decorative, but they are not appropriate either for evaluating human judgment or as a research tool to uncover the underlying processes“. Um menschliches Denken beurteilen zu können, d.h. welche kognitiven Prozesse zu welchen Urteilen führen, reicht die bloße Betrachtung der Kongruenz mit einer Norm nicht aus. Z.B. können sich bereits durch unterschiedliche Interpretation von Aufgabentexten abweichende Lösungen ergeben, die keine kognitiven Fehlschlüsse bedeuten (Fiedler, 1980). Manchmal treten auch Fehler im analytischen Vorgehen auf, z.B. simple Rechenfehler (Scholz, 1987), die aber kaum im Sinne einer falschen Denkweise zu bewerten sind. Auch können völlig unterschiedliche Denkprozesse zur selben Lösung führen.¹³ Für die Aussagekraft empirischer Studien ist daher eine Untersuchung von *Lösungsprozessen* entscheidend. In solchen Studien zeigte sich deutlich, dass von der Norm abweichende Urteile ganz unterschiedliche kognitive Strategien als Ursache haben können (vgl. empirische Untersuchungen in Kap.I.4 bzw. Kap. III und IV).

Bedeutsam in diesem Zusammenhang sind auch Theorien und Befunde zu einfachen und effizienten Heuristiken („satisficing“), die vom Menschen bewusst als Strategie verwendet werden, um mit möglichst wenig Aufwand zu Urteilen zu gelangen. Diese Vision von Rationalität beachtet z.B. adaptive Ziele, natürliche Kontexte und kognitive Beschränkungen des Denkens (für eine Übersicht z.B. Chase, Hertwig & Gigerenzer, 1998; Gigerenzer, Todd & the ABC Research Group, 1999; Martignon et al., 2003). Dass dabei abhängig von der Struktur des Kontextes erstaunliche Urteilsgenauigkeit verglichen mit der normativen Optimalstrategie erreicht werden kann, ist deutlich gezeigt worden (z.B.

¹³ für eine übersichtliche Zusammenstellung der Kritik vgl. auch Bea, 1995, S.43f.

Martignon & Laskey, 1999; Martignon & Hoffrage, 1999; Martignon, 2001). Es wird darauf hingewiesen, dass die Strategie der „Repräsentativitätsheuristik“ unter gewissen Umständen eine gute Orientierung bei stochastischen Problemen bietet, d.h. nicht unbedingt zu einem fehlerhaften Urteil führt. In Bayesianischen Situationen wäre das etwa bei einer Informationsstruktur der Fall, wenn die Wahrscheinlichkeit der Hypothese $P(H)$ ähnlich hoch wie die Wahrscheinlichkeit der Daten $P(D)$ ist.¹⁴ Tatsächlich zeigten empirische Untersuchungen anhand verschiedener Bayesianischer Probleme, dass die Anwendung der „Repräsentativitätsstrategie“ mit dem Vorliegen einer solchen Informationsstruktur korrelierte (Gigerenzer & Hoffrage, 1995, S.698; vgl. eine detaillierte Darstellung der Studien in Kap.I.4.).

So müssen „falsche“ – oder besser gesagt nicht normative – Ergebnisse bei der Lösung von Textproblemen nicht unbedingt fehlerhaftes Denken oder Intuition bedeuten und sind schon gar nicht auf die Stufe von Befunden über *das menschliche Denken* generalisierbar. Diesem „bias“ unterlagen vielfach die Vertreter des Heuristikprogramms (vgl. z.B. Einhorn & Hogarth, 1981; Gigerenzer & Murray, 1987). In diesem Sinne kommt Fiedler (1980, S.36) zu dem Schluss: „Die kognitiv-psychologische Aussagekraft dieser Befunde ist [...] bei weitem überschätzt worden“.

2.3 Bewertung aus didaktischer Sicht

Generell wurde von Mathematikdidaktikern, die sich mit Stochastik befassten, häufig betont, dass bei der Beschäftigung mit stochastischen Problemstellungen in weit stärkerem Maße als bei anderen Gebieten der Mathematik von der Theorie abweichende Vorstellungen existieren. Freudenthal z.B. unterstreicht: „Seit dem Anfang der Wahrscheinlichkeitsrechnung hat es immer strittige Probleme gegeben, das heißt Aufgaben, die verschieden gelöst wurden, bis eine sorgfältige Analyse zeigte, welche der streitenden Parteien recht hatte. Es waren übrigens durchaus nicht zweitklassige Mathematiker, die da desavouiert wurden; zu dieser Erscheinung gibt es in der reinen Mathematik keine Analogie“ (Freudenthal, 1973, S.532).

Die im „Heuristikprogramm“ identifizierten Schwierigkeiten von Versuchspersonen, z.B. bei bedingter Wahrscheinlichkeit und Bayesianischen Folgerungen, prägten schließlich auch eine didaktische Diskussion. Das Interesse der Mathematikdidaktik richtete sich jedoch – wie auch neuere Trainingsforschung in der Kognitionspsychologie (vgl. Kap.III, z.B. Sedlmeier, 1999) – naturgemäß eher auf die prozessuale Frage, wie Schwierigkeiten besser überwunden werden können bzw. stochastisches Denken gefördert werden kann, als auf die reine Untersuchung der Diskrepanz menschlichen Denkens von normativen Lösungen und der Kategorisierung. Eine Vielzahl von Arbeiten, die auf kritischer didaktischer Analyse und Diskussion der Ergebnisse von Kahneman & Tversky's Heuristiktheorie basierten (z.B. Shaughnessy, 1977; Scholz, 1981; Walter, 1983; Bentz & Borovcnik, 1985; Pollatsek et al., 1987; Falk, 1988), regten zum Weiterdenken in Richtung einer Veränderung der konventionellen Lehr- und Lernmethoden an. Nur um ein Beispiel zu geben, wird die Repräsentativitätsheuristik bei Bentz und Borovcnik (1985) als vornehmlich *positive* Strategie aufgefasst, „[...] nach der man sich in bestimmten stochastischen Problemen gut orientieren kann. Man muss nur aufpassen, wie man ‚repräsentativ‘ im jeweiligen Fall zu verstehen hat. Dieses zu klären, zu schulen, ist eine Aufgabe des Unterrichts“ (S.248). Die didaktische Perspektive reicht dabei teilweise über die rein stoffliche Dimension hinaus und führt zu elementaren Überlegungen über grundlegende Schwierigkeiten bei der Modellbildung und Mathematisierung oder über Grenzen und Wert intuitiver und formaler Lösungen. Ziel ist eine

¹⁴ ergibt sich aus: $P(H|D) \approx P(D|H) \Leftrightarrow P(H \cap D) : P(D) \approx P(H \cap D) : P(H) \Leftrightarrow P(H) \approx P(D)$.

Veränderung des Schulcurriculums in Richtung einer verbesserten Entwicklung von Handlungs- und Entscheidungsfähigkeit der Schüler (Scholz, 1981, S.3f.).

Die Frage der Intuitivität der aus der Theorie abgeleiteten Vorgehensweisen spielt in diversen Arbeiten eine zentrale Rolle (z.B. Borovcnik & Bentz, 1991; Borovcnik, 1992) und basiert auf den entwicklungspsychologischen Untersuchungen von z.B. Piaget & Inhelder (1951/1975), Fischbein (1975; 1987) oder Kapadia (1984). Die Berücksichtigung der grundlegenden Intuitionen der Lernenden und die Weiterentwicklung stehen im Vordergrund, wobei nun nicht mehr nur die Widersprüche, sondern vor allem auch die Komplementarität von intuitiven und formalen Strategien Beachtung finden (Wittmann, 1981). Das zentrale Prinzip des Lernens von Stochastik ist demnach ein dynamisches Wechselspiel aus intuitiven Vorstellungen und z.B. durch systematische Unterweisung neu entstehender Theorien. Eine große Forschungsfrage ist natürlich, wie diese Unterweisung auszusehen hat, um Intuitionen in den Köpfen der Schüler zu adäquatem Wissen werden zu lassen. Diese Forschung impliziert auch die Frage nach geeigneten Bewertungsmaßstäben für konzeptuelles Wissen in Stochastik (Konold, 1995).

Trotz dem Bewusstsein um die besondere Problematik des Lehrens und Lernens von Stochastik ist diese Sichtweise weit optimistischer und konstruktiver als das Credo der „Heuristics and biases“-Literatur und ermutigte nun dazu, über didaktische Möglichkeiten und Veränderungen unter Berücksichtigung dieser intuitiven Vorstellungen nachzudenken: „In stressing the difficulty of altering various student conceptions, I certainly do not mean to discourage attempts to teach probability and statistics concepts. Rather I want to suggest that we do more than we currently do to determine the effects of instruction on whatever concepts we hope our instruction fosters“ (Konold, 1995, S.7). Einige empirische Untersuchungen zeigten bereits Möglichkeiten für diese Förderung stochastischen Wissens auf (z.B. Shaughnessy, 1992). Für den Themenbereich „Bedingte Wahrscheinlichkeit und Bayes“ gibt Kapitel II eine Analyse des Status quo in der Mathematikdidaktik.

Ein meines Erachtens sehr wichtiges Fazit aus der großen Fülle didaktischer und psychologischer Analysen des Denkens unter Unsicherheit, die durch das „Heuristikprogramm“ stimuliert wurden, ist, dass beide Perspektiven sich wechselseitig befruchten können (Scholz, 1991, S.4). Einerseits müssen psychologische Erklärungsansätze für Denkprozesse, insofern sie plausible Aussagen mit Prozesscharakter bieten, natürlich bei den Konstruktionen von Curricula und Lehr-Lern-Szenarien beachtet werden. Andererseits kann die didaktische Analyse manche, aus experimental-psychologischer Sicht manchmal vernachlässigte Probleme bei der empirischen Untersuchung von Denkprozessen aufdecken (vgl. Bentz & Borovcnik, 1985, S.258f.). Auch diese Arbeit soll zur Vernetzung der beiden Disziplinen einen weiteren Beitrag leisten.

3 Forschungsprogramm „Ökologische Rationalität“

Wir fanden bisher bezüglich der Frage der Übereinstimmung Bayesianischen Denkens mit den Gesetzen der Wahrscheinlichkeitstheorie von Kongruenz (zumindest des „*homme éclairé*“) zu Zeiten der Aufklärung, über „tendenzielle“ Kongruenz bei frühen Kognitionsexperimenten bis hin zu weitgehendem Konflikt im Heuristikprogramm einen höchstmöglichen Grad der Kontroverse. Der nun dargestellte Ansatz geht davon aus, dass diese Frage zu eng gefasst ist und einen entscheidenden Aspekt außer Acht lässt: Kognitive Algorithmen sind jeweils an ihre spezifische Umwelt adaptiert und nicht an mathematische Theorie oder Logik. Die Analogie zwischen kognitiven und visuellen Illusionen, die von

Vertretern der Heuristikforschung gerne zur Verdeutlichung der Stabilität von Fehlschlüssen verwendet wurde, kann auch die Idee der „ökologischen Rationalität“ plausibel machen. Aus „optischen Täuschungen“, die unter bestimmten Bedingungen auftreten (z.B. abhängig von Lichtverhältnissen), lässt sich nicht folgern, dass menschliche Wahrnehmung generell defizitär zu bewerten ist. Analog betrachtet, würde auch das menschliche Denken unter Unsicherheit nur unter bestimmten Bedingungen „versagen“.

Gerd Gigerenzer (z.B. 1991, 1993c) und die Forschungsgruppe „Adaptives Verhalten und Kognition“ sahen in der *Repräsentation von Informationen* eine wichtige Bedingung für menschliche Rationalität. Gewissermaßen folgt diese Art der Repräsentation aus der „Umwelt“, an die kognitive Algorithmen mehr oder weniger adaptiert sind. Diese Vision einer „ökologischen Rationalität“, also der Verbindung von menschlicher Rationalität mit den Informationsstrukturen der Umwelt, ist insbesondere von Egon Brunswik (z.B. 1943; 1955) und Herbert Simon (z.B. 1956; 1990) besonders gefördert worden.

3.1 Natürliche Häufigkeiten und „natural sampling“

Im Falle Bayesianischen Denkens war also die entscheidende Frage, an welche Art von Informationsrepräsentation kognitive Algorithmen eher adaptiert sind. Die (numerischen) Informationen in den Aufgabenitems der Studien des „Heuristikprogramms“ waren als Wahrscheinlichkeiten bzw. relative Häufigkeiten formuliert. Dieses Informationsformat entspricht der heutzutage üblichen Art, Unsicherheit zu kommunizieren. Der Umgang mit relativen Häufigkeiten bzw. Wahrscheinlichkeiten oder Prozentwerten entwickelte sich aber erst allmählich im Zuge der Mathematisierung von Unsicherheit und Zufall, also in den letzten Jahrhunderten. Auf welche Informationen hätte der Mensch aber ohne solche mathematischen Objekte seine Beurteilung von Unsicherheit gestützt? Plausibel ist, dass er absolute Häufigkeiten heranzieht, die er beobachtet oder anders erfahren hat. Das Argument, das nun zur obigen Frage zurückführt, ist, dass menschliches Denken aufgrund der Langsamkeit evolutorischer Prozesse noch nicht an die aus der mathematischen Theorie entstandenen numerischen Formate (z.B. „85% aller Taxis in der Stadt sind blau“) angepasst ist, wohl aber natürlicherweise an Häufigkeitsinformationen, die in einer Art sequenziellem Beobachtungs- oder Erfahrungsprozess gebildet werden (z.B. „Von 60 beobachteten Taxis sind 51 blau“). Aufgrund der Überzeugung, dass Menschen intuitive Urteile auf eine solche induktive Weise in Interaktion mit der natürlichen Umwelt fällen, sprechen Gigerenzer und Kollegen von „natürlichen Häufigkeiten“. Der Prozess ihrer Entstehung wird als „*natural sampling*“ bezeichnet (Kleiter, 1994; Gigerenzer & Hoffrage, 1995).

Ich möchte kurz Begriffe und Bedeutung von „natural sampling“ bzw. „natürlichen Häufigkeiten“ näher erläutern und insbesondere von „relativen Häufigkeiten“ abgrenzen, um Missverständnisse zu vermeiden. Ich werde dies in Analogie zur obigen medizinischen Diagnosesituation (vgl. 2.1) deutlich machen.

Stellen wir uns eine Welt vor, in der nichts über das Auftreten von Brustkrebs oder die Verlässlichkeit von Testverfahren bekannt ist. In dem Fall könnte ein Diagnostiker nur aus seiner Erfahrung urteilen.

Nehmen wir an, ein Arzt hätte bisher 1000 symptomfreie 40 bis 50-jährige Frauen untersucht, von denen 10 Brustkrebs (B) hatten. Von den 10 kranken Frauen erhielten 8 ein positives Testergebnis (T_+). Von den 990 Gesunden waren 95 Testergebnisse positiv.

Nun kommt eine 45-jährige symptomfreie Frau zum Arzt und will auch auf Brustkrebs getestet werden. Sie testet positiv. Wie groß ist die Wahrscheinlichkeit, dass sie Brustkrebs hat?

Diese „Übersetzung“ der Diagnosesituation mit relativen Häufigkeiten in natürliche Häufigkeiten könnte das Ergebnis eines „Natural sampling“ – Prozesses sein. Die Wahrscheinlichkeitsfrage sollte wohl richtiger lauten: Wie groß sollte der Arzt aufgrund seines Kenntnisstandes die Wahrscheinlichkeit einschätzen, dass die Frau Brustkrebs hat? Unter Einbeziehung seiner Erfahrungen gibt der Anteil der von ihm untersuchten Frauen, die krank waren und positiv testeten, von allen Frauen, die positiv testeten, eine vernünftige Schätzung:

$$P(B|T_+) = \frac{|B \cap T_+|}{|T_+|} = \frac{|B \cap T_+|}{|B \cap T_+| + |\bar{B} \cap T_+|} = \frac{8}{8+95} = \frac{8}{103} = 7,8\% \quad (2)$$

Wie man sieht, braucht der Arzt bei der Schätzung nur zwei Teilpopulationen aller untersuchten Frauen zu beachten, nämlich die kranken Frauen mit positivem Test ($B \cap T_+$) und die nicht kranken Frauen mit positivem Test (nicht $B \cap T_+$). In diesem Sinne ist „natural sampling“ ein simpler Zählvorgang des Auftretens bestimmter Merkmalskategorien, der zu „natürlichen Häufigkeiten“ führt.

Relative Häufigkeiten entstehen ganz genauso in einem „Sampling-Prozess“, den ich zur Unterscheidung als „*relative sampling*“ bezeichnen will (vgl. Abb.I.1). Die Beobachtung, dass bei einer Untersuchung von symptomfreien 40-50jährigen Frauen (oder allgemein beim ersten von j Experimenten), die 1000-mal wiederholt wird, 10-mal eine Krankheit auftritt, führt zu der mehr oder weniger berechtigten Folgerung, dass die Krankheit auch bei den nächsten 1000 Untersuchungen (unter gleichen Bedingungen) etwa 10-mal – vielleicht ein wenig häufiger oder auch seltener – beobachtet wird. Als Folge dieser Gesetzmäßigkeit ordnet man dem Auftreten der Krankheit einen Wert zu, der ihre Auftretenswahrscheinlichkeit (unter diesen Bedingungen) beschreibt, nämlich $P(B) = 10 / 1000$ oder 1% („relative Häufigkeit“). Nun folgt ein anderes Experiment (z.B. von einem Entwickler für medizinische Tests, Experiment 2): Diesmal werden 40-50jährige Frauen untersucht, die bereits krank sind, z.B. auch 1000. Man stellt nun fest, dass ein spezielles Testverfahren bei 800 Frauen zu einem positiven Ergebnis führte, und folgert analog, dass die Auftretenswahrscheinlichkeit von positiven Testergebnissen bei kranken 40-50jährigen Frauen $P(T_+ | B) = 800 / 1000$ ist oder 80%. Ein analoges „Test-Experiment“ wird auch mit gesunden Frauen durchgeführt.

Der Unterschied zwischen natürlichen und relativen Häufigkeiten ist nun ersichtlich: Während natürliche Häufigkeiten für alle relevanten Ereignisse in *einem einzigen* Experiment bzw. „Sampling-Prozess“ entstehen, werden relative Häufigkeiten für jedes Ereignis in unterschiedlichen Experimenten bzw. „Sampling-Prozessen“ ermittelt (vgl. Abb.I.1). Sie tragen damit aber keine Informationen über die anderen „Sampling-Prozesse“. Also z.B. die relative Häufigkeit für positive Testergebnisse bei Brustkrebs „80%“ beinhaltet keine Information über die Basisrate für Brustkrebs „1%“, die natürliche Häufigkeit „8 (von 10)“ hingegen schon. Gewissermaßen wurde die Basisrate bereits mitverarbeitet im Sinne eines Anteils von einem Anteil („8 von 10 Kranken von 1000 Frauen insgesamt“). Während man aus natürlichen Häufigkeiten jegliche Art von Wahrscheinlichkeitsschätzung, wie z.B. für $P(B | T_+)$, direkt treffen kann, müssen für gewisse Wahrscheinlichkeitsschätzungen aus relativen Häufigkeiten Verknüpfungsregeln angewandt werden.

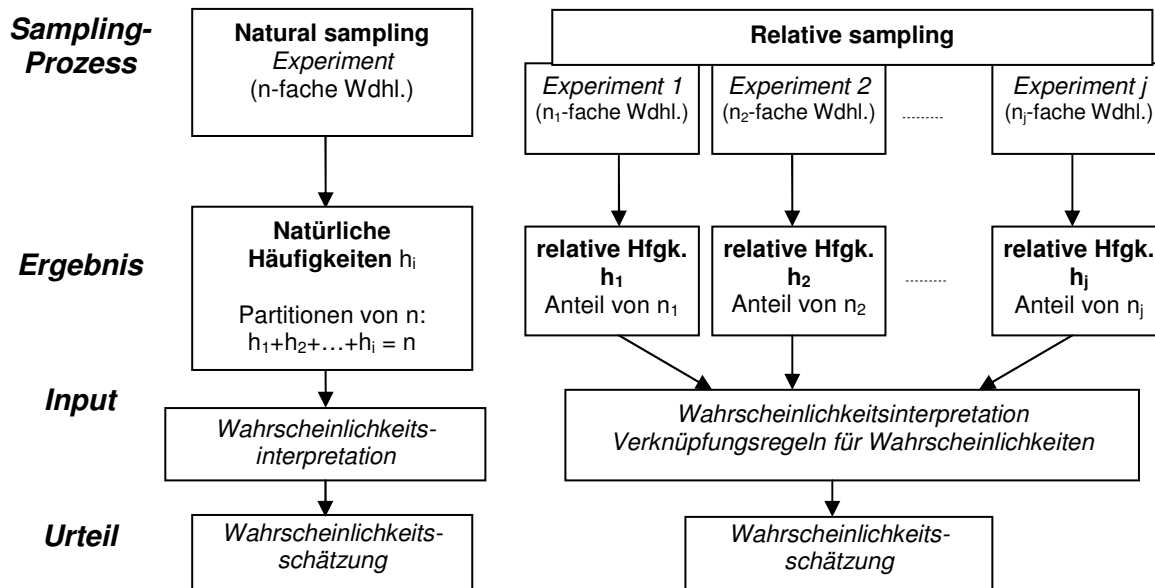


Abb.I.1: „Natural sampling“ und „Relative sampling“ im Vergleich

Bei Bayesianischen Situationen heißt das konkret, dass A-priori- mit Likelihood-Wahrscheinlichkeiten gewichtet und diese gewichteten Wahrscheinlichkeiten über weitere Regeln für bedingte Wahrscheinlichkeit bzw. totale Wahrscheinlichkeit weiterverarbeitet werden müssen, wodurch eine Regel (von Bayes) für eine vernünftige Schätzung der „neuen“ Wahrscheinlichkeit entsteht (vgl. Gleichung 1 in 2.1.).

Zurück zum menschlichen Denken und zur heutigen Welt: Selbstverständlich können Menschen heutzutage sehr häufig auf Statistiken als Ergebnis gesammelter Informationen zurückgreifen und müssen nicht mehr selbst „samplen“. Der Denkprozess setzt in diesem Fall sozusagen erst *nach* dem Sampling-Prozess ein. Die Informationen sind oft lediglich als relative Häufigkeiten verfügbar, wie insbesondere auch in den Aufgabenitems der „Heuristikexperimente“ von Kahneman, Tversky und Kollegen. Die Hypothese, die Gigerenzer (1991) oder Cosmides & Tooby (1996) aufstellten, besagt, dass gerade dieses Format „relative Häufigkeit“ bzw. „Wahrscheinlichkeit“ als Grenzwert relativer Häufigkeiten die mannigfaltig festgestellten Probleme bei Wahrscheinlichkeitsschätzungen erst hervorruft. Weil Menschen ohne Wissen um stochastische Regeln intuitiv „natural sampling“ betreiben würden, nehmen sie an, dass Wahrscheinlichkeitsschätzungen, die auf der Basis natürlicher Häufigkeiten getroffen werden, deutlich besser ausfallen, als solche auf der Basis von relativen Häufigkeiten. Diese Grundhypothese wurde detailliert empirisch untersucht (vgl. 4. in diesem Kapitel).

Die Hypothese, dass mentale Algorithmen zum Umgang mit Unsicherheit auf Basis natürlicher Häufigkeiten sehr gut arbeiten, ist auch konsistent mit Befunden anderer Forschungsgebiete zur menschlichen Kognition, z.B. der *Forschung zur Häufigkeitsverarbeitung* („frequency processing“). Wie Häufigkeiten verarbeitet werden, ist in der aktuellen psychologischen Forschung ein vielfach diskutiertes Thema (für eine Übersicht: Sedlmeier & Betsch, 2002), und war es in Bezug auf Gedächtnisleistung bereits zu den Anfängen der Psychologie (z.B. Ebbinghaus, 1913/ 1885). Unzweifelhaft sind Häufigkeitsinformationen eine der fundamentalen Dimensionen der Erfahrung und verdienen besondere Aufmerksamkeit.

Der Untersuchungsschwerpunkt der „frequency processing“-Forschung ist die Verarbeitung wiederholt auftretender Ereignisse oder Objekte, die entweder direkt beobachtet oder indirekt, z.B. als zusammengefasste Häufigkeitsinformation, enkodiert werden („ich sah fünf junge Leute mit grünen Haaren“). Die Verarbeitung von Häufigkeitsinformationen wurde aufgrund diverser Experimente vielfach als nahezu automatischer Prozess beschrieben, der ohne besondere Aufmerksamkeit und Intention erfolgt (Hasher & Zacks, 1984; Überblick: Zacks & Hasher, 2002). Die Erklärung des Heuristikansatzes für „Fehler“, die Verfügbarkeitsheuristik (Tversky & Kahneman, 1973), ist in der heutigen Sicht nur eine von vielen möglichen Strategien, die bei der Verarbeitung von Häufigkeiten angewandt werden könnte. Dem Erklärungsansatz mangelt es, ähnlich dem Repräsentativitätsansatz, an Aussagekraft, theoretischer Konsistenz und sogar an Falsifizierbarkeit (Betsch & Pohl, 2002). Ein derzeit weitreichender Konsens besteht darin, dass Menschen sehr sensitiv gegenüber Häufigkeitsinformationen sind. Es gibt mittlerweile gute theoretische Simulationsmodelle, die sowohl den generell sehr kompetenten Umgang mit wiederholt auftretenden Ereignissen und darauf basierende Schätzungen von relativen Häufigkeiten oder Wahrscheinlichkeiten, als auch mögliche Störungen und ihre Bedingungen erklären können (vgl. MINERVA-DM: Dougherty et al., 1999; PASS: Sedlmeier, 1999, 2002). Die Ergebnisse dieses Forschungsgebietes der menschlichen Häufigkeitsverarbeitung erhellen auch die Prozesse, die bei der Verarbeitung von natürlichen Häufigkeiten als Input bei probabilistischen Urteilen ablaufen, und besitzen damit einen hohen Erklärungswert für mögliche Befunde.

3.2 Grafisch-visuelle Repräsentation

Es besteht kein Zweifel darüber, dass grafische Darstellungen die Leistungsfähigkeit des Menschen bei der Informationsverarbeitung unterstützen können, insbesondere in stochastischen Bereichen (vgl. Tukey, 1977; Biehler, 1982). Sie sind damit eine weitere wichtige Dimension des didaktischen Interesses (z.B. Biehler, 1985). Eine theoretische Erklärung gibt z.B. die „Dual-Coding Theory“ (z.B. Paivio, 1971, 1978), die von zwei kognitiven, aber interagierenden Subsystemen ausgeht: einem verbalen und einem visuellen, die beide an der Verarbeitung visueller Informationen beteiligt sind. Die Theorie besagt, dass grafisch-visuelle Darstellung, die von beiden Subsystemen simultan verarbeitet wird, besser behalten und reproduziert wird als textuelle, die nur vom verbalen System verarbeitet wird. Der genaue Einfluss der Modalitäten und deren Wechselwirkung im Denken ist jedoch noch wenig erforscht. Insbesondere kann die Theorie keine klärenden Aussagen darüber bieten, warum eine grafische Darstellung einer anderen überlegen sein kann (für einen Überblick vgl. Mandl & Levin, 1989).

Aus der Theorie der „ökologischen Rationalität“ folgt eine differenzierende Hypothese über den Nutzen grafisch-visueller Darstellungen für die Beurteilung probabilistischer Situationen. Grundsätzlich ist sie umso gebräuchlicher für die intuitive Problemlösung je näher sie die in der natürlichen Umgebung sichtbaren Eigenschaften der auftretenden Ereignisse repräsentiert bzw. widerspiegelt. Die These der Adaption des menschlichen Denkens an natürliche Häufigkeiten als Ergebnis eines „Natural sampling“-Prozesses kann also auch auf grafische Darstellungen übertragen werden. Insbesondere folgt aus den Überlegungen zu „Natural sampling“-Prozessen, dass eine Verbesserung der Problemlöseleistung auch erreicht werden kann, wenn der Sampling-Prozess (oder eine Simulation davon) visuell verfolgt werden kann, anstatt nur darüber berichtet wird (Sedlmeier, 1999, S.65). Befunde bei Christensen-Szalanski und Beach (1982) zeigten z.B., dass nach einer sequenziellen Beobachtung von Ereignissen bessere Wahrscheinlich-

keitsschätzungen gegeben werden konnten als nach einer Angabe der relativen Häufigkeiten des Auftretens dieser Ereignisse.¹⁵

Wie müsste eine grafische Darstellung aussehen, die möglichst gut das Ergebnis eines „natural sampling“-Prozesses wiedergibt? Ich halte die folgenden Punkte für entscheidend (vgl. auch Wassner, Krauss & Martignon, 2002):

1. Sie sollte auf *natürlichen Häufigkeiten* basieren, die entweder direkt beobachtet wurden oder als Ergebnis eines Sampling-Prozesses kommuniziert werden.
2. Sie sollte den *prozessualen Charakter* des „natural sampling“ verdeutlichen. D.h. es muss nachvollziehbar sein, wie die natürlichen Häufigkeiten gebildet wurden.
3. Sie sollte *Mehrstufigkeit* veranschaulichen können.
4. Sie sollte *hierarchische Beziehungen* abbilden können.

Punkte 1 und 2 folgen aus dem bisher Dargestellten. Punkt 3 ist erforderlich, da mehr als ein Merkmal in die Beurteilung einbezogen werden muss. Punkt 4 ist sinnvoll, da eine Klassifizierung bzw. Kategorisierung („Teilmengenbildung“) zuerst nach einem Merkmal, dann nach dem nächsten Merkmal u.s.w. erfolgt.

Das *Baumdiagramm* wird benutzt, um mehrstufige Prozesse zu veranschaulichen. Es dient auch, z.B. in der Informatik, zur Darstellung hierarchischer Beziehungen oder Datenstrukturen. Die Baumdiagrammstruktur eignet sich in besonderer Weise als grafische Veranschaulichung des Ergebnisses eines „natural sampling“-Prozesses. Abb. I.2 zeigt die Darstellung der Diagnosesituation (vgl. 3.1.) in einem Baumdiagramm („Häufigkeitsbaum“; vgl. Gigerenzer & Hoffrage, 1995, S.687). Die Darstellung erfüllt insbesondere die 4 oben genannten Anforderungen. Die verschiedenen Informationen werden stufenweise im Baum verarbeitet, wobei jeweils fehlende Werte über Summenbildung ergänzbar sind.

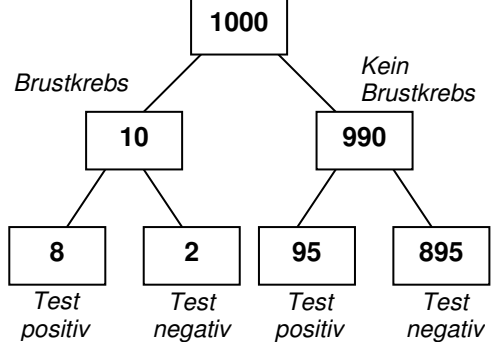
„Natural sampling“	Darstellung im Häufigkeitsbaum
<p>Information: Nehmen wir an, ein Arzt hätte bisher 1000 symptomfreie 40 bis 50-jährige Frauen untersucht, von denen 10 Brustkrebs (B) hatten.</p> <p>Von den 10 kranken Frauen erhielten 8 ein positives Testergebnis (T₊). Von den 990 Gesunden waren 95 Testergebnisse positiv.</p> <p>Problemstellung: Nun kommt eine 45-jährige symptomfreie Frau zum Arzt und will auch auf Brustkrebs getestet werden. Sie testet positiv. Wie groß ist die Wahrscheinlichkeit, dass sie tatsächlich Brustkrebs hat?</p>	<p>Grafisch-visuelle Darstellung der Information:</p>  <pre> graph TD A[1000] -- "Brustkrebs" --> B[10] A -- "Kein Brustkrebs" --> C[990] B -- "Test positiv" --> D[8] B -- "Test negativ" --> E[2] C -- "Test positiv" --> F[95] C -- "Test negativ" --> G[895] </pre> <p>Beurteilung: Nur noch die positiven Testfälle sind relevant. Alle positiven Testfälle: 8 + 95 = 103, davon tatsächlich Brustkrebs: 8 Wahrscheinlichkeitsschätzung: 8 / 103 = 7,8%</p>

Abb. I.2: „Natural sampling“ und korrespondierende grafische Darstellung im „Häufigkeitsbaum“

¹⁵ Sequentielle Darbietung der Informationen über Patienten erfolgte mittels 100 Dias mit jeweiligen Angaben, ob ein Patient Lungenkrebs hat oder nicht und ob ein Test positiv oder negativ ausfiel. Die Versuchspersonen schätzten danach im Mittel P(Lungenkrebs | positiver Test) auf 0,22 (tatsächliche relative Häufigkeit war 0,24); allerdings war eine hohe Varianz der Schätzungen festzustellen.

Das Baumdiagramm beinhaltet alle relevanten Ergebnisse des „Sampling“ Prozesses, die als Input für die Beurteilung herangezogen werden müssen. Die Urteilsbildung geschieht auf Basis weiterer sehr intuitiver Überlegungen, die – wie bereits erwähnt – ohne Verknüpfungsregeln für Wahrscheinlichkeiten auskommen (vgl. auch Abb.I.1).

Es gibt weitere grafische Darstellungsformen, die das Ergebnis eines „Natural sampling“- Prozesses veranschaulichen können, z.B. *Tabellen* (vgl. Abb.I.3) oder *Flächen- bzw. Rasterdiagramme* (vgl. Abb.I.6 bzw. Abb.II.12; Gigerenzer, 1993c, S.106).¹⁶

	<i>gesamt</i>	<i>Brustkrebs</i>	<i>Kein Brustkrebs</i>
<i>gesamt</i>	1000	10	990
<i>Test positiv</i>	103	8	95
<i>Test negativ</i>	897	2	895

Abb. I.3: Grafische Darstellung in einer Tabelle mit natürlichen Häufigkeiten („Vierfeldertafel mit Randhäufigkeiten“)

Bezüglich des Informationsgehalts bzw. der -menge können diese Darstellungen analog sein. Fraglich ist allerdings, ob sie auch ähnlich gut den Urteilsprozess unterstützen. Einige Befunde dazu werden in 4.2. dargestellt.

Es sei an dieser Stelle auf theoretische Vorteile von Baumdiagrammen gegenüber anderen Darstellungsformen hingewiesen. Die oben aufgeführten Punkte 2, 3, und 4 sind nach meiner Einschätzung bei Baumdiagrammen am besten verwirklicht. Die Kategorisierung (Bildung der möglichen Teilmengen) wird im Baum durch die hierarchische Struktur am klarsten veranschaulicht, d.h. die Entstehung der natürlichen Häufigkeiten kann am einfachsten nachvollzogen werden. Tatsächlich kann aber *nur* das Baumdiagramm Mehrstufigkeit veranschaulichen, wenn *mehr als zwei* Merkmale in die Beurteilung einbezogen werden sollen. Auch für diese komplexeren Bayesianischen Situationen gibt es empirische Befunde im Hinblick auf eine Urteilsverbesserung durch Häufigkeiten (vgl. 4.1.3).

4 Empirische Ergebnisse des Forschungsprogramms „Ökologische Rationalität“ zum Bayesianischen Denken

In einer Reihe von empirischen Studien wurden aus den theoretischen Überlegungen, wie im vorherigen Abschnitt erläutert, generierte Thesen zum Einfluss der Darbietungsform¹⁷ auf Performanz *und* Prozesse bei Bayesianischem Denken untersucht. Ich fasse einige grundlegende theoretische Überlegungen zu Vorteilen natürlicher Häufigkeiten als „Input“ für Bayesianisches Denken im Vergleich zu relativen Häufigkeiten (bzw. Wahrscheinlichkeiten) nochmals zusammen (vgl. auch Gigerenzer & Hoffrage, 1995):

- Wenn der Mensch über adäquate intuitive mentale Algorithmen verfügt, arbeiten diese mit Häufigkeitsinformationen („*evolutionary argument*“, ebenda S.686).

¹⁶ Auf diese und weitere Darstellungsformen wird auch detailliert in einer didaktischen Analyse (Kapitel II.2) eingegangen.

¹⁷ Ich verwende im Folgenden „Darbietungsform“ als Überbegriff für die Darstellung einer Problemsituation i.w.S. hinsichtlich numerischer Informationsformate und -struktur, Elementen der Situationsbeschreibung, Art der Frage- bzw. Problemstellung, Einbeziehung grafischer Darstellungen etc.

Daraus ergeben sich direkte prozedurale Konsequenzen:

- Natürliche Häufigkeiten werden in *einem* sequentiellen Prozess gebildet, in dem insbesondere Basisraten in jeder Stufe berücksichtigt werden. „Base-rate Neglect“ wird insbesondere unterbunden.¹⁸
- Die Notwendigkeit von Verknüpfungsregeln für Wahrscheinlichkeiten ist bei natürlichen Häufigkeiten nicht gegeben.
- Der „Input“ der mathematischen Verarbeitung sind natürliche Zahlen an Stelle von rationalen Zahlen im Intervall [0;1] (bzw. Prozentwerte).
- Insgesamt sind weniger mathematische Operationen nötig (vgl. Gleichung 1 und 2).

Im Folgenden gebe ich einen Überblick über einige Studien und wichtige empirische Ergebnisse zum Bayesianischen Denken. Dabei gehe ich zunächst auf Laborvergleichsstudien ein, in denen das numerische Format der Information bzw. die Struktur der Situationen variiert, aber *keine vorherige Instruktion* bzw. grafischen Darstellungshilfen¹⁹ gegeben wurden. Anschließend gehe ich auf Studien ein, in denen auch der Einfluss grafischer Darstellung (ohne Instruktion und Training) untersucht wurde. Die Verwendung dieser Ergebnisse für *Trainingsstudien* wird in *Kapitel III* gesondert behandelt.

4.1 Einfluss des Informationsformates auf Bayesianisches Denken

Das Ziel der Studien war – gleichsam als „Antwort“ auf das Heuristikforschungsprogramm – ein Performanzvergleich zwischen Problemsituationen in der „Standardformulierung“ mit relativen Häufigkeiten bzw. Wahrscheinlichkeiten und „Übersetzungen“ dieser Situationen in natürliche Häufigkeiten. Nicht nur die Urteilsperformanz, sondern auch die Urteilsprozesse wurden untersucht.

Eine Wahrscheinlichkeitsversion der Mammografie-Aufgabe (vgl. 2.1.) wurde z.B. folgendermaßen in eine Häufigkeitsversion „übersetzt“:

Wahrscheinlichkeitsversion (Standard)	Häufigkeitsübersetzung
<i>Die Wahrscheinlichkeit, dass eine 40-jährige, symptomfreie Frau Brustkrebs hat, die an einer Vorsorgeuntersuchung teilnimmt, beträgt 1%.</i>	<i>10 von 1000 symptomfreien, 40-jährigen Frauen, die an einer Vorsorgeuntersuchung teilnehmen, haben Brustkrebs.</i>
<i>Wenn Brustkrebs vorliegt, ist die Wahrscheinlichkeit 80%, dass die Mammografie positiv ausfällt.</i>	<i>8 von 10 Frauen mit Brustkrebs erhalten eine positive Mammografie.</i>
<i>Wenn kein Brustkrebs vorliegt, ist die Wahrscheinlichkeit 9,6%, dass die Mammografie trotzdem positiv ausfällt.</i>	<i>95 von 990 Frauen ohne Brustkrebs erhalten trotzdem eine positive Mammografie.</i>
<i>Eine 40-jährige, symptomfreie Frau erhält eine positive Mammografie in einer Routineuntersuchung.</i>	<i>Stellen Sie sich eine Menge zufällig ausgewählter, symptomfreier Frauen mit 40 Jahren vor, die eine positive Mammografie in einer Routineuntersuchung erhalten haben.</i>
<i>Wie groß ist die Wahrscheinlichkeit, dass sie tatsächlich Brustkrebs hat? _____%</i>	<i>Wie viele dieser Frauen haben nach Ihrer Einschätzung tatsächlich Brustkrebs? _____ von</i>

Tab.I.1: Mammografie-Aufgabe in der Wahrscheinlichkeitsversion und in der Häufigkeitsversion (englische Originalversion: Gigerenzer & Hoffrage, 1995, S.688)

¹⁸ Im Diagnosebeispiel: Im Prozess der Bildung der natürlichen Häufigkeit für kranke Frauen mit positivem Test „8 (von 10 von 1000)“ sind Basisrateninformation „10 von 1000“ und Likelihoodinformation „8 von 10“ enthalten.

¹⁹ Man beachte, dass die Versuchspersonen evtl. im eigenen Lösungsprozess grafische Darstellungen benutzen konnten.

Ich möchte eingangs, bevor ich näher auf die Studien eingehe, auf einen klärenden Punkt hinweisen, der diese Übersetzung betrifft. Unter wahrscheinlichkeitstheoretischen Gesichtspunkten betrachtet stellt die Häufigkeitsversion ein gedankliches, idealisiertes, mehrstufiges Zufallsexperiment dar. *Gedanklich* deswegen, weil es der Urteilende nicht selbst durchführt, also der Sampling-Prozess nicht in der eigenen Erfahrung des Urteilenden liegt, sondern ihm nur das Ergebnis bekannt ist. *Idealisiert* deswegen, weil das Ergebnis genau der in den Wahrscheinlichkeiten ausgedrückten Erwartung entspricht, insofern ein „ideales“ Ergebnis darstellt, das im Allgemeinen nicht unbedingt so auftritt, wenn das Experiment durchgeführt wird. *Mehrstufig* deswegen, weil zwei Merkmale in Betracht gezogen werden. Die Wahrscheinlichkeitsversion weist im Übrigen auf ein Zufallsexperiment mit genau denselben Eigenschaften hin. Wenn man „Wahrscheinlichkeit x %“ als „x von 100“ interpretiert, wird dies ersichtlich, wobei aber – wie bereits ausgeführt – keine „Natural sampling“-Situation entsteht.

Es sei betont, dass der Übergang von Häufigkeiten zu Wahrscheinlichkeiten eine fundamentale Idee darstellt, die sich im empirischen Gesetz der großen Zahlen widerspiegelt. Die Bemühung um das Verständnis dieser Idee verdient freilich unter didaktischen Gesichtspunkten besondere Beachtung und ist als solche grundsätzlich umfassend zu erhellen. Einer möglichen Fehlinterpretation soll frühzeitig Vorschub geleistet werden: Übersetzungen in natürliche Häufigkeiten sind *nicht als Ersatz* der Idee „Wahrscheinlichkeit“ zu bewerten, sondern als „Hilfestellungen“ für die Modellierung stochastischer Probleme und zur Veranschaulichung wahrscheinlichkeitstheoretischer Zusammenhänge.

Im Folgenden gehe ich auf die Studien von Gigerenzer & Hoffrage (1995) und Cosmides & Tooby (1996) zur detaillierten, experimentellen Untersuchung des Nutzens von Häufigkeitsformaten ein. Anschließend werden darauf aufbauende Studien mit komplexeren Situationen von Krauss et al. (1999; eingereicht) und mit 10 bis 12-jährigen Schülern von Zhu & Gigerenzer (eingereicht) besprochen.

4.1.1. Variation des Informationsformats in verschiedenen Bayesianischen Kontexten: *Studien von Gigerenzer & Hoffrage (1995)*

In dieser Studie wurden Studenten (60 Teilnehmer) ohne vorherige Instruktion jeweils Wahrscheinlichkeits- bzw. Häufigkeitsversionen in Standardstruktur 15 verschiedener Bayesianischer Situationen, u.a. auch der Mammografie-Aufgabe (vgl. Tab.I.1) und des Taxi-Problems (vgl. Kap.I.2.), zur Bearbeitung vorgelegt. Außerdem wurden die Aufgaben auch jeweils verkürzt gestellt („short menu“) mit bereits zusammengefassten Informationen. Das heißt z.B. in der Mammografie-Aufgabe, dass bereits die konjunktive Wahrscheinlichkeit $P(M_+ \cap B) = 0,8\%$ und die totale Wahrscheinlichkeit $P(M_+) = 10,3\%$ gegeben sind, analog in der Häufigkeitsversion $|M_+ \cap B| = 8$ und $|M_+| = 103$. Insofern sind die „Short menu“-Aufgaben eigentlich keine Bayesianischen Probleme, sondern Aufgaben zur bedingten Wahrscheinlichkeit.²⁰

Eine Gruppe erhielt jeweils die Standard-Wahrscheinlichkeitsversionen, eine andere Gruppe die Standard-Häufigkeitsversionen (für Designdetails vgl. Gigerenzer & Hoffrage, 1995, S.693). Erhoben wurden jeweils die Ergebnisse als Wahrscheinlichkeits- bzw. Häufigkeitsschätzungen *und* die Prozesse als schriftliche Aufzeichnungen der Lösungsüberlegungen („Write aloud“-Protokolle). Die Lösung wurde nur dann als „korrekt“ gewertet, wenn Schätzwert und Lösungsprozess eindeutig darauf hinwiesen („Bayesian

²⁰ Ich gehe auf die Ergebnisse im „short menu“ im Rahmen dieser Arbeit nicht näher ein.

algorithms“). Auf diese Weise wurden auch „unkorrekte“ Lösungsprozesse identifiziert („non Bayesian algorithms“).

Ergebnisse

Die Versuchspersonen konnten im Durchschnitt die Standard-Häufigkeitsversionen zu 46% (50% im „short menu“) im obigen Sinne korrekt lösen, die Wahrscheinlichkeitsversionen nur zu 16% (28% im „short menu“).²¹ Darüber hinaus wurden auch die nicht-Bayesianischen Algorithmen identifiziert, die öfter als *Lösungsstrategie* auftraten (vgl. Tab.I.2; Gigerenzer & Hoffrage, 1995, S.694).

Cognitive algorithm	Formal equivalent	Information format and menu				Total	% of total
		Probability		Frequency			
		Standard	Short	Standard	Short		
Bayesian	$p(H D)$	69	126	204	221	620	34.9
Joint occurrence	$p(H \& D)$	39	97	20	97	253	14.3
Adjusted joint occurrence	$p(H \& D) \pm .05$		64		55	119	6.7
Fisherian	$p(D H)$	67		36		103	5.8
Adjusted Fisherian	$p(D H) \pm .05$	32		19		51	2.9
Multiply all	$p(D)p(H \& D)$		79		12	91	5.1
Likelihood subtraction	$p(D H) - p(D - H)$	30		4		34	1.9
Base rate only	$p(H)$	6		13		19	1.1
Less frequent algorithms (<1% of total)		71	32	60	29	192	10.8
Not identified		119	52	89	32	292	16.5
Total		433	450	445	446	1,774 ^a	100.0

Note. Numbers are absolute frequencies.

^a The sum of total answers in Table 3 is 1,774 rather than 1,800 (60 participants times 30 tasks) because of some participants' refusals to answer and a few missing data.

Tab.I.2: Lösungsstrategien bei Bayesianischem Denken, Gigerenzer & Hoffrage (1995, S.695)

- „*Joint occurrence*“: Es wird die konjunktive Wahrscheinlichkeit (bzw. ein äquivalentes Häufigkeitsverhältnis) errechnet und angegeben. Teilweise wurde das Ergebnis „adjustiert“ angegeben (um max. $\pm 5\%$).
Z.B. In der Mammografie-Aufgabe: $P(M_+ \cap B) = P(M_+ | B) \cdot P(B) = 80\% \cdot 1\% = 0,8\%$ bzw. äquivalent als Häufigkeitsverhältnis, z.B. 8 von 1000.
- „*Fisherian*“²²: Es wird die zur gesuchten Wahrscheinlichkeit inverse bedingte (Likelihood-) Wahrscheinlichkeit (bzw. ein äquivalentes Häufigkeitsverhältnis) angegeben. Teilweise wurde das Ergebnis „adjustiert“ angegeben (um max. $\pm 5\%$).
Z.B. In der Mammografie-Aufgabe: $P(M_+ | B) = 80\%$ bzw. äquivalent als Häufigkeitsverhältnis, z.B. 8 von 10.
- „*Likelihood-Subtraction*“: Die beiden bekannten Likelihood-Wahrscheinlichkeiten werden subtrahiert (bzw. äquivalente Häufigkeitsverhältnisse).
Z.B. In der Mammografie-Aufgabe: $P(M_+ | B) - P(M_+ | \text{nicht } B) = 80\% - 9,6\% = 70,4\%$ bzw. äquivalent als Häufigkeitsverhältnis, z.B. 7 von 10.

²¹ Z.B. die Mammografieaufgabe wurde in der Standard-Häufigkeitsversion zu 42%, in der Wahrscheinlichkeitsversion zu 18% korrekt gelöst. Zum Vergleich: Bei Eddy (1982) gaben mindestens 95% der befragten Ärzte bei der Mammografie-diagnose im Wahrscheinlichkeitsformat deutlich falsche Schätzungen ab.

²² diese Bezeichnung wird bei Gigerenzer & Hoffrage (1995) verwendet, um auf die eingangs erwähnte analoge Inversität des Signifikanzniveaus bei Fischer's Nullhypothesentests zu Ergebnissen bei Bayesianischen Testverfahren hinzuweisen.

- „*Base-rate only*“: Es wird nur die bekannte Basisrate angegeben (bzw. ein äquivalentes Häufigkeitsverhältnis).

Z.B. In der Mammografie-Aufgabe: $P(B) = 1\%$ bzw. äquivalent als Häufigkeitsverhältnis, z.B. 10 von 1000.

Wie bereits in 2.2. erwähnt, bietet z.B. der „Fisherian-Algorithmus“, der im Ergebnis der Verwendung der Repräsentativitätsheuristik im Kahnemanschen Sinne gleichkommt, gute Schätzungen, wenn $P(H) = P(D)$ ist. Ein Vergleich über die 15 verschiedenen Bayesianischen Aufgaben ergab, dass die Anwendung des „Fisherian-Algorithmus“ mit dem Vorliegen einer solchen Informationsstruktur in der Aufgabe korrelierte.²³

Diskussion

Die These wurde voll bestätigt, dass allein die Veränderung der Informationsformate in Bayesianischen Aufgabentexten von relativen Häufigkeiten bzw. Wahrscheinlichkeiten in natürliche Häufigkeiten eine deutliche Verbesserung Bayesianischer Wahrscheinlichkeits-schätzungen bewirkt. Die Ergebnisse warfen ein neues Licht auf die Frage, wie Menschen mit Bayesianischen Situationen intuitiv umgehen. Weder die positive Grundeinstellung zu Zeiten der Aufklärung noch die negativen Befunde der Heuristikforschung lieferten eine ausreichende Analyse. Sie stützten ihre Folgerungen auf die Güte kognitiver Algorithmen, ohne zu berücksichtigen, auf welchen Informationsformaten diese Algorithmen arbeiten. Eine gute Metapher ist der einfache Taschenrechner, dessen Algorithmen mit Zahlen im Dezimalsystem als Input perfekt funktionieren, jedoch bei binären Zahlen kläglich versagen (z.B. Gigerenzer, 1993c, S.104). Natürlich wäre es schlichtweg falsch, zu behaupten, der Taschenrechner könne keine Grundrechenarten durchführen aufgrund der Feststellung, dass er das mit binären Zahlen nicht kann.

Die Studien von Gigerenzer und Hoffrage (1995) schafften aus einem evolutions-theoretischen Argument und prozessualen Folgerungen einen neuen Rahmen zur Erforschung von Denken unter Unsicherheit. Man beachte, dass die Ergebnisse der Heuristikforschung nicht widerlegt wurden, sondern ihre Folgerungen. Das betrifft insbesondere ganz entscheidend didaktische Implikationen, wie ich noch ausführlich erläutern werde (vgl. Kap.I.5 und II.3). Darüber hinaus wurde durch das Untersuchungsdesign ermöglicht, prozessuale Analysen durchzuführen, um Gründe für Performanzunterschiede beleuchten zu können. Es wurden *mehrere* häufiger auftretende Algorithmen identifiziert, die zu „Base Rate Neglect“ führen können (u.a. „Fisherian“ und „Likelihood subtraction“), so dass gezeigt wurde, dass die Repräsentativitätsheuristik als einziger Erklärungsansatz nicht ausreicht. Deutlich wurde auch, dass „Base-rate Neglect“ bei natürlichen Häufigkeiten weit weniger oft auftritt als bei Wahrscheinlichkeiten bzw. relativen Häufigkeiten. Eine weitere Folgerung der Prozessanalysen war, dass „Base-rate Neglect“, der bisher als *der* typische „Fehler“ beim Bayesianischen Denken eingestuft wurde, nicht häufiger vorkommt als z.B. die Missachtung der „Falsch-Alarm Rate“ (im Mammografiebeispiel: $P(M_+ | \text{nicht } B)$).

Obwohl hier nicht alle Folgerungen aus den Ergebnissen von Gigerenzer & Hoffrage (1995) dargestellt sind, wurde bereits deutlich, dass der Rahmen zur Erforschung Bayesianischen Denkens viel weiter zu stecken ist, als dies in bisherigen Forschungsprogrammen getan worden war. Auf diese Studien aufbauend konnten eine Vielzahl weiterer Versuche unternommen werden, noch detaillierter Performanz- und Prozessvariable zu untersuchen, von denen ich im Folgenden einige wichtige Ergebnisse

²³ Genauer: Verwendungshäufigkeit und $|1 - p(H)/p(D)|$ sind negativ korreliert ($r = -0,49$).

erläutern werde. Auf die praktischen Konsequenzen, z.B. für Risikokommunikation oder didaktische Überlegungen, gehe ich abschließend in diesem Kapitel ein.

4.1.2. Variation der Informationselemente innerhalb der Bayesianischen Situation: *Studien von Cosmides & Tooby (1996)*

Auch Cosmides und Tooby vertraten Positionen und Grundgedanken der „ökologischen Rationalität“ und damit zusammenhängende evolutorische Argumente zum Denken unter Unsicherheit: „The assumption of severe processing limitations has forestalled many researchers from seriously considering or investigating a contrasting possibility: that our minds come equipped with very sophisticated intuitive statistical competences that are well engineered solutions to the problems humans normally encountered in natural environments“ (Tooby & Cosmides, 1992, S.9; Cosmides, 1989). Sie stützten ihre Untersuchungen ebenfalls auf die „Häufigkeitshypothese“ und versuchten in einer Serie von Experimenten mit Studenten eine Erhöhung der Performanz mit Häufigkeitsversionen gegenüber der originalen Wahrscheinlichkeitsversion bei der Beurteilung einer Bayesianischen Situation zu erreichen (*Teil I*). Darüber hinaus untersuchten sie in weiteren Experimenten, ob bereits eine Spezifizierung der Wahrscheinlichkeitsversion diese positive Wirkung hervorrufen kann (*Teil II*) und insbesondere welche Elemente der Häufigkeitsversion für die positiven Effekte verantwortlich sind (*Teil III*).

Ausgangspunkt war ein Diagnoseproblem, das bereits von Casscells et al. (1978; Cosmides & Tooby, 1996, S.23) verwendet wurde, um medizinische Experten im Bayesianischen Denken zu testen:

Wenn ein Test für eine Krankheit, deren Prävalenz [= Basisrate, Anm. d. Verf.] 1/1000 beträgt, eine falsch-positiv Rate von 5% aufweist, wie hoch ist die Wahrscheinlichkeit, dass eine Person mit einem positiven Testergebnis tatsächlich unter der Krankheit leidet, unter der Annahme, dass keinerlei Symptome oder sonstige Anzeichen festzustellen sind? _____%

45% der Teilnehmer (Ärzte und Medizinstudenten) antworteten mit „95%“, 18% antworteten mit „2%“.²⁴ Eine Folgerung war, dass auch bei medizinischen Experten „Base-rate Neglect“ weitverbreitet auftritt (Tversky & Kahneman, 1982, S.154). Cosmides & Tooby „übersetzten“ das Diagnoseproblem in folgende Häufigkeitsversion (vgl. Cosmides & Tooby, 1996, S.24):

1 von je 1000 Amerikanern hat die Krankheit X. Ein Test wurde entwickelt, um festzustellen, ob eine Person die Krankheit X hat. Immer, wenn der Test bei einer kranken Person durchgeführt wird, ist das Ergebnis positiv (i.e. eine „richtig-positiv“ Rate von 100%). Manchmal wird das Testergebnis auch positiv, wenn eine Person völlig gesund ist. Genauer, von je 1000 gesunden Personen erhalten 50 ein positives Ergebnis (i.e. eine „falsch-positiv“ Rate von 5%). Stellen Sie sich eine zufällige Stichprobe von 1000 Amerikanern vor, die getestet werden sollen. Sie wurden in einer Lotterie ausgewählt, ohne dass die Durchführenden Informationen über den Gesundheitszustand der Personen hatten. Beurteilen Sie mit obigen Informationen, wie viele tatsächlich die Krankheit haben von denen, die ein positives Testergebnis erhalten? _____ von _____

Diese Häufigkeitsversion enthält, außer der Übersetzung in Häufigkeitsverhältnisse, eine Reihe von Änderungen bzw. Spezifizierungen zur Originalversion:

- a. Die Sensitivität des Tests wird explizit angegeben: „Immer, wenn der Test bei einer kranken Person ...“

²⁴ Man beachte, dass die Aufgabenstellung unterdeterminiert ist, da keine Richtig-positiv-Rate (Sensitivität des Tests) angegeben ist. Unter der Annahme, dass diese 100% beträgt und dass diese Basisrate von 1/1000 für die betreffende Person angemessen ist, wäre „2%“ eine richtige Antwort.

- b. Der Begriff „falsch-positiv“ wird näher erläutert: „Manchmal wird das Testergebnis auch positiv, wenn ...“
- c. Auf die zufällige Auswahl der Stichprobe wird hingewiesen: „Stellen Sie sich eine zufällige Stichprobe ... vor. Sie wurden in einer Lotterie ausgewählt, ...“
- d. Eine Stichprobengröße als konkrete Referenzpopulation wird angegeben: „... von 1000 Amerikanern...“
- e. Es wird nach einem Häufigkeitsverhältnis gefragt: „Beurteilen Sie ..., wie viele ... von...“

Die Vorgehensweise war nun kurz gesagt die, dass diese Häufigkeitsversion schrittweise durch Hinzufügen, Abändern oder Streichen von Informationen modifiziert wurde, um eine differenziertere Analyse zu ermöglichen. Verschiedene Versionen wurden jeweils ca. 25 Studenten zur Bearbeitung vorgelegt.

Ergebnisse

Teil I. „Optimierung“ der Häufigkeitsversion

Zunächst konnten die Ergebnisse von Casscells et al. (1978) mit der originalen Wahrscheinlichkeitsversion repliziert werden, genauer gesagt: 56% der Studenten antworteten mit „95%“ und 12% mit „2%“. Allerdings waren auch z.B. 12% „konservativ“ und beachteten nur die Basisrate („Base rate only“). Mit der Häufigkeitsversion urteilten nun nur noch 4% mit „95%“ („950 von 1000“) und 56% korrekt mit „2%“ (also z.B. „1 von 51“).

Wenn vorher einige auf die Lösung führende Zusatzfragen gestellt wurden, wie z.B. „Wie viele der 1000 Personen haben die Krankheit UND testen positiv?“ oder „Wie viele der 1000 Personen testen insgesamt positiv, egal ob sie die Krankheit haben oder nicht?“ konnten die richtigen Urteile auf 76% erhöht werden.

Ohne die Angabe von relativen Häufigkeiten (in den Klammern) stieg die Urteilsperformanz auch auf 72% (ohne Zusatzfragen). Zusätzliche Informationen in relativen Häufigkeiten wirkten sich also negativ auf die Performanz aus.²⁵

Unter Zuhilfenahme einer grafischen Darstellung konnte die Performanz sogar auf bis zu 92% korrekte Urteile gesteigert werden. Dieses Experiment wird in 4.2.1 genauer erläutert.

Teil II. „Optimierung“ der Wahrscheinlichkeitsversion

Wenn in der Wahrscheinlichkeitsversion die Sensitivität explizit angegeben und der Begriff „Falsch-positiv-Rate“ näher erläutert wurde (= Änderungen a. und b., siehe oben), stieg der Anteil korrekter Lösungen („2%“) auf 36% und die Antwort „95%“ fiel auf 32%.

Der Hinweis auf die Zufälligkeit der Stichprobe könnte bei der Beurteilung eine große Rolle spielen, wenn die Prävalenz-Information für die Krankheit in einer Population nicht als „a-priori-gültig“ für einen Einzelfall gesehen wird. Eine begründete Annahme ist, dass Urteilende Populationsbasisraten nur dann in die Wahrscheinlichkeitsschätzung eines Einzelfalles miteinbeziehen, wenn sie von der Zufälligkeit überzeugt sind (vgl. Gigerenzer, Hell & Blank, 1988). Die Einbeziehung des Hinweises auf die zufällige Auswahl über ein „Lotterieverfahren“ (= Änderung c. ohne Angabe von d.) zusätzlich zu den Änderungen a. und b. in der Wahrscheinlichkeitsversion konnte keine Erhöhung korrekter Antworten bewirken, jedoch die Antwort „95%“ fiel weiter auf 16%.

Teil III. Elemente des „Häufigkeitseffektes“

Wenn in der Wahrscheinlichkeitsversion mit Änderungen a. und b. anstatt der Wahrscheinlichkeitsfrage eine Häufigkeitsfrage gestellt wurde (= Änderung e.), stieg der Anteil korrekter

²⁵ Alle Unterschiedshypothesen wurden auf Signifikanz überprüft ($p < 0.05$), genaue Angaben bei Cosmides & Tooby (1996).

Lösungen („2%“) weiter auf 64% (vorher 36%, s. Teil II). Durch den zusätzlichen Hinweis auf die Zufälligkeit der Stichprobe (= Änderung c.) konnte der Anteil nicht weiter erhöht werden. Das galt auch für eine zusätzliche Konkretisierung der Stichprobengröße von „1000“ (= Änderung d.).

Andererseits, wenn in der Häufigkeitsversion (ohne zusätzliche Angaben von relativen Häufigkeiten) *nur* die Häufigkeitsfrage durch eine Wahrscheinlichkeitsfrage ersetzt wurde (Wegfall der Änderung e.), fiel der Anteil korrekter Lösungen deutlich auf 56% (vorher 72%). Wenn in der Häufigkeitsversion (ohne zusätzliche Angaben von relativen Häufigkeiten) *nur* die Konkretisierung der Stichprobengröße von „1000“ weggelassen wurde (Wegfall der Änderung d.), verminderte sich der Anteil richtiger Lösungen praktisch nicht (68%).

Diskussion

Neben weiterer klarer Evidenz, dass eine substantielle Verbesserung Bayesianischer Wahrscheinlichkeitsurteile durch eine Übersetzung der Informationen von Wahrscheinlichkeits- in Häufigkeitsformate erreicht werden kann, konnten Cosmides und Tooby aus ihrer Experimentserie differenzierte Aussagen über die Einflussfaktoren treffen (zumindest für den einen Aufgabentyp). Es wurde deutlich, dass die Variation der Darbietungsform einer Problemstellung eine sehr weite Bandbreite an Performanz hervorrufen kann (von 12% bis 92% korrekter Urteile!).

Teil II der Studie zeigte insbesondere, dass die niedrigen Raten „richtiger“ Antworten bei der originalen Wahrscheinlichkeitsversion nur teilweise an der Ambiguität bzw. Underdeterminiertheit des Aufgabentextes liegen. Durch Vervollständigung und weitere Erklärung der enthaltenen Informationen waren aber Steigerungen von 12% auf 36% erreichbar. Teil III spezifizierte, dass sowohl das Häufigkeitsformat der Informationen als auch die Fragestellung nach einem Häufigkeitsverhältnis maßgeblich für die weitere Erhöhung der Performanz im Sinne des „Häufigkeitseffektes“ verantwortlich waren. Beide Veränderungen der hinsichtlich der Originalversion von Casscells et al. „optimierten“ Wahrscheinlichkeitsversion erhöhten einzeln die Performanz jeweils um ca. 20 Prozentpunkte, in Verbindung um ca. 36 Prozentpunkte. Kaum Einfluss hat offensichtlich der Hinweis auf die Zufälligkeit der Situation bzw. die Angabe einer konkreten Größe einer Referenzpopulation.

Ein generelles Ergebnis dieser Experimente war, dass die Performanz bei Bayesianischen Urteilen umso mehr anstieg, je mehr Häufigkeitsüberlegungen „initiiert“ wurden. Cosmides und Tooby (1996) konstatierten, dass die Befunde als Belege dafür anzusehen sind, dass Häufigkeitsrepräsentationen kognitive Mechanismen aktivieren, die vernünftiges Bayesianisches Denken produzieren. Sie sehen gemäß Marr's (1982, S.24) Untersuchungen zu „visual information processing“ auch verschiedene Ebenen dieses Denkens. Die Frage, ob der Mensch gemäß der wahrscheinlichkeitstheoretischen Regeln denkt, zielt auf das Vorhandensein von kognitiven Mechanismen ab, die eine solche Verarbeitung durchführen. Es ist zu bezweifeln, dass diese grundsätzlich zur Verfügung stehen. Aus den Befunden der Experimente folgt, dass aber wenigstens ein Mechanismus vorhanden zu sein scheint, der Häufigkeitsinformationen (a-priori) mit weiteren Häufigkeitsinformationen verknüpfen kann, so dass als Ergebnis eine vernünftige Wahrscheinlichkeitsschätzung erfolgen kann.

4.1.3. Verallgemeinerung der Bayesianischen Basissituation:

Studien von Krauss, Martignon, Hoffrage und Gigerenzer (eingereicht bzw. 1999)

Die bisherigen Experimente beschränkten die Untersuchung Bayesianischen Denkens auf den „einfachsten Fall“ (Basissituation): Eine Hypothese mit zwei Alternativen (dichotom) wird anhand eines Befundes (Datums) bewertet, der entweder vorliegt oder nicht vorliegt (dichotom). Diese einfachste Bayesianische Situation kann in dreierlei Hinsicht komplizierter werden:

- a. „*Polychotome Hypothese*“: In realer Anwendung könnte z.B. in der medizinischen Diagnosesituation auf mehrere verschiedene (sich ausschließende) Krankheiten getestet werden.
- b. „*Polychotomes Datum*“: Auch der Testbefund kann weitere Abstufungen aufweisen.
- c. „*Mehrfache Daten*“: Weitere Tests können durchgeführt werden und die Krankheitshypothese wird anhand weiterer Befunde neu beurteilt. Wahrscheinlichkeitstheoretisch erfolgt die mehrfache Anwendung der Bayesregel.

Bei der Beurteilung „*komplexerer*“ *Bayesianischer Situationen* mittels Bayesregel²⁶ nehmen sowohl Anzahl der zu verarbeitenden Informationen als auch Komplexität des Algorithmus zu, so dass der Schwierigkeitsgrad solcher Beurteilungsaufgaben offensichtlich anwächst.

In *Studie 1* von Krauss et al. (eingereicht) wurden Medizinstudenten vier verschiedene Diagnoseaufgaben mit erhöhter „Komplexität“ zur Bearbeitung vorgelegt:

1. Diagnoseaufgabe mit trichotomem Datum (Brustkrebs ja/nein, Mammografie positiv/unklar/negativ)
2. Diagnoseaufgabe mit trichotomer Hypothese (Hepatitis A/Hepatitis B/keine der beiden Krankheiten, Test positiv/negativ)
3. Diagnoseaufgabe mit zwei dichotomen Testbefunden (Brustkrebs ja/nein, Mammografie und Ultraschalluntersuchung, jeweils positiv/negativ)
4. Diagnoseaufgabe mit drei dichotomen Testbefunden (Krankheit X ja/nein, Test 1, Test 2 und Test 3, jeweils positiv/negativ)

Einer Gruppe wurden die vier Aufgaben mit Wahrscheinlichkeitsangaben vorgelegt, der anderen Gruppe als Häufigkeitsversionen analog zu Gigerenzer & Hoffrage (1995).

In *Studie 2* erhielten die Versuchspersonen jeweils vorher eine kurze (schriftliche) Demonstration, wie man die Mammografieaufgabe in der Basissituation löst, die auch mit Hilfe grafischer Darstellung erfolgte. Gruppe 1 wurde die Lösung einer Wahrscheinlichkeitsversion mit der Bayesregel gezeigt; Gruppe 2 die Lösung einer Wahrscheinlichkeitsversion mit Übersetzung in natürliche Häufigkeiten und Darstellung in einen Häufigkeitsbaum; Gruppe 3 die Lösung einer Häufigkeitsversion mit Darstellung in einem Häufigkeitsbaum. Gruppe 1 und 2 wurden daraufhin die Aufgaben 1 und 3 (siehe oben) in der Wahrscheinlichkeitsversion, Gruppe 3 in der Häufigkeitsversion gestellt.

Ergebnisse

Die Performanz beider Gruppen in Studie 1 bei den verschiedenen Aufgaben geht aus Abbildung I.4 hervor.

²⁶ Der Beweis von Bayes (1763) beinhaltet die Verallgemeinerung der „dichotomen“ Situation.

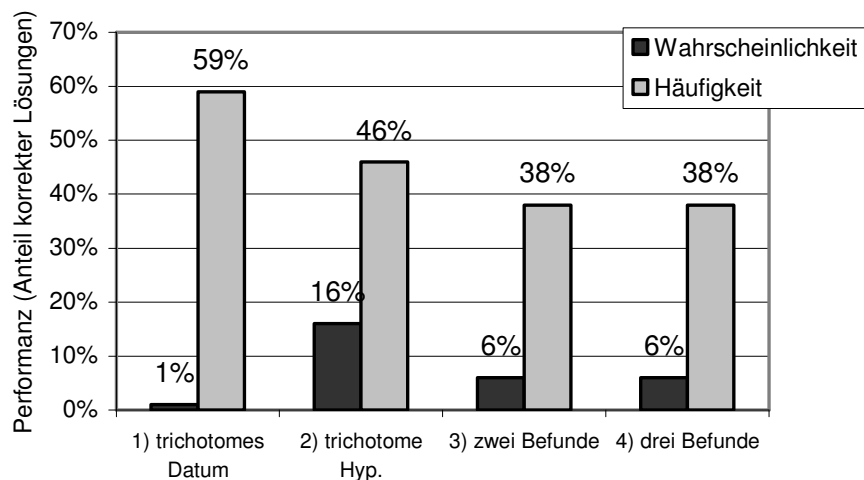


Abb.I.4: Performanz bei komplexeren Bayesianischen Situationen

Bei allen komplexeren Aufgaben ermöglichten die Häufigkeitsversionen deutlich bessere Urteilsperformanz als die Wahrscheinlichkeitsversionen (über alle Aufgaben 45% korrekte Urteile im Vergleich zu 7%). Am besten in der Häufigkeitsversion wurde Aufgabe 1 mit trichotomem Datum gelöst (59%) und der Unterschied zur Wahrscheinlichkeitsversion war am größten (58 Prozentpunkte). Am besten in der Wahrscheinlichkeitsversion wurde Aufgabe 2 mit trichotomer Hypothese gelöst (16%). Erstaunlicherweise war weder in der Wahrscheinlichkeits- noch in der Häufigkeitsversion ein Unterschied zwischen Aufgabe 3 und 4 festzustellen, obwohl die Komplexität ansteigt und der Kontext abstrakter wird.

In Studie 2 bewirkte bei Aufgabe 1 mit trichotomem Datum die Demonstration der Lösung der Basissituation mit der Regel (Gruppe 1) eine Performanzsteigerung auf 18%, Demonstration mit „Übersetzung“ von Wahrscheinlichkeiten in einen Häufigkeitsbaum (Gruppe 2) auf 40%, Demonstration von vornherein mit Häufigkeiten und Baumdiagramm (Gruppe 3) auf 73%. Für Aufgabe 3 mit der „Doppel-Testsituation“ ergaben sich entsprechende Steigerungen in den Gruppen 1 bis 3 auf 22%, 40% und 81%.

Diskussion

Die Ergebnisse zeigen die Anwendbarkeit und den Nutzen des „Häufigkeitskonzeptes“ im Vergleich zu Wahrscheinlichkeitsformaten auch in komplexeren Bayesianischen Situationen. Es wurde in Studie 2 auch deutlich, dass eine vorherige Demonstration der Lösung der Bayesianischen Basissituation sich bereits förderlich auf die Performanz auswirken kann. Insbesondere wurde festgestellt, dass auch *grafische Darstellung* bei dieser Demonstration weiter performanzsteigernd wirkt. Insofern geben die Studien von Krauss et al. (eingereicht) erste Hinweise auf den Nutzen von *Instruktion* mit Hilfe grafischer Darstellungen bei Bayesianischem Denken. Ich habe sie hier mitdargestellt, da sie im Zusammenhang mit komplexeren Bayesianischen Aufgaben erfolgten. In separaten Abschnitten dieser Arbeit werden weitere Ergebnisse insbesondere zum Nutzen grafischer Darstellung und Instruktion fokussiert (vgl. 4.2 und Kap.III).

Eine Anmerkung für den Leser soll noch zu Möglichkeiten der grafischen *Darstellung* dieser komplexeren Situationen gemacht werden, die bei Krauss et al. (eingereicht) zur Verständlichmachung der Aufgabenstruktur eingesetzt wurden, deren Nutzung aber nicht empirisch untersucht wurde. Die in 3.2 bereits erwähnte Flexibilität der Baumstruktur ist v.a. für die Modellierung der Aufgaben 3 (vgl. Abb.I.5) und 4 entscheidend. Sie wäre mit anderen

grafischen Darstellungen, wie etwa der Tabelle oder Flächendarstellungen, nicht übersichtlich möglich. Zu einigen empirischen Befunden über die Verwendung grafischer Darstellungen in komplexeren Bayesianischen Situationen (v.a. bei Strukturen wie in Aufgabe 2) werde ich im Rahmen der eigenen Trainingsstudien mit Schülern kommen (vgl. Kap. III, 3).

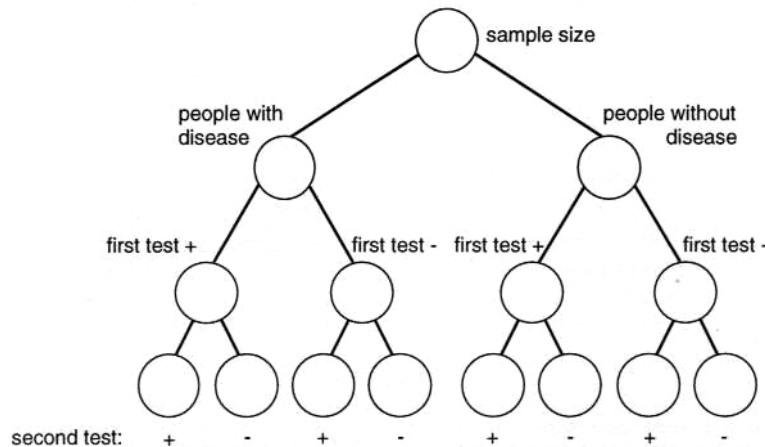


Abb.I.5: Baumstruktur zu einer Diagnoseaufgabe mit zwei dichotomen Testbefunden (Krauss et al., eingereicht)

4.1.4. Bayesianisches Denken bei Kindern: Studie von Zhu & Gigerenzer (eingereicht)

Die Annahme der Existenz eines Mechanismus zu häufigkeitsbasiertem, probabilistischem Denken führte auch zur Frage, ob und wann dieser in der Entwicklung des Menschen entsteht. Aufbauend auf Arbeiten von Piaget und Inhelder (1975/1951), die mit Interviews zeigten, dass Kinder ab einer gewissen Stufe intuitiv gemäß Grundideen des Zufalls denken können, ist die Hypothese naheliegend, dass bereits Kinder über intuitive häufigkeitsbasierte Bayesianische Kompetenzen verfügen. Befunde zum frühen stochastischen Verständnis zeigten, dass die Handlungskompetenz einer verbalen Artikulationskompetenz voraussetzt (z.B. Yost et al., 1962; Davies, 1965). Wollring (z.B. 1994) stellte mit Hilfe von Spielinterviews (Wollring, 1993) und videobasierter Handlungsdokumentation fest, dass bei Grundschulkindern (8-9 Jahre) Elemente frequentistischer Kompetenzen zu beobachten sind, die sich auf das zufallsbedingte Entstehen von Ergebnissen, auf mögliche Verteilungen und auch basale Vorstellungen zu Mittelwerten und Streumaßen beziehen. Sie beruhen meist auf Zählstrategien. Die „stochastische Kompetenz von Grundschulkindern besteht im wesentlichen in ihrer Fähigkeit, bei geeigneter Artikulationsunterstützung aufgrund von Versuchserfahrungen stochastische Konzepte zu konstruieren“ (Wollring, 1994, S.176).

In der Studie von Zhu und Gigerenzer wurden querschnittlich chinesischen Schülern der 4., 5. und 6. Klasse (durchschnittliches Alter: 10-12 Jahre) und einer Kontrollgruppe von Erwachsenen (MBA-Studenten, durchschnittliches Alter: 29 Jahre) frequentistische Textversionen „Bayesianischer Situationen“ vorgelegt, die weitgehend an die Lebenserfahrung von Kindern dieses Alters angepasst waren:

Eine Kekspackung beinhaltet salzige und süße Kekse mit verschiedenen Formen. In der Packung sind von 100 Keksen 20 salzig. Von den 20 salzigen Keksen sind 14 rund. Von den anderen 80 süßen Keksen sind 16 rund. Stell Dir vor, Du nimmst eine Handvoll runde Kekse aus der Packung. Wie viele davon sind salzig? ___ von ___

Ergebnisse

Jeder Schüler bekam 7 derartige Aufgaben vorgelegt. Tabelle I.3 gibt an, wie viele Aufgaben jeweils von wie vielen Versuchspersonen richtig gelöst wurden. Kriterien zur Wertung einer Lösung als „korrekt“ wurden von Gigerenzer & Hoffrage (1995; vgl. 4.1.1.) übernommen. Im Falle der obigen Aufgabe würde z.B. „20 von 30“ als eine korrekte Antwort gelten.

Schüler der (Altersdurchschnitt)	Anzahl korrekter Antworten							Gesamt- zahl	
	0	1	2	3	4	5	6		7
4.Klasse (9;9)	6	2	7	1	0	0	0	0	16
5.Klasse (10;8)	6	2	1	1	2	2	1	0	15
6.Klasse (11;9)	3	0	0	1	0	0	4	6	14
Erwachsene (29)	4	1	0	0	1	0	3	14	23

Tab.I.3: Performanz der Schüler nach Klassenstufen und Erwachsener (Zhu & Gigerenzer, eingetrag.)

Die Viertklässler gaben in 17% aller Aufgaben korrekte Antworten. 6 Schüler gaben keine richtige Antwort und der beste Schüler gab 3 richtige Lösungen an. Bereits bei den Fünftklässlern war eine deutlich höhere Performanz (30% richtige Lösungen) zu erkennen, der beste Schüler konnte 6 richtige Lösungen angeben, wobei eine hohe Varianz erkennbar war. Die Sechstklässler erreichten bereits eine sehr hohe Performanz (70% richtige Lösungen), der größte Teil der Schüler (71,5%) konnte 6 oder 7 Aufgaben richtig lösen. Es ist eine starke Polarisierung am oberen und unteren Ende des Performanzspektrums festzustellen. Bei der Kontrollgruppe mit Studenten zeigte sich kaum höhere Performanz (75% richtige Lösungen) und eine ähnlich polarisierte Performanzstruktur wie bereits bei den Sechstklässlern. Ähnliche Ergebnisse wurden in einer Replikation dieser Studie mit deutschen Schülern der Klassenstufen 5 bis 7 gefunden (Lücking, 2003).

Diskussion

Dieses Ergebnis sehen Zhu und Gigerenzer als Beleg, dass bereits in einem Alter von etwa 11 Jahren Kinder über intuitive Formen Bayesianischen Denkens verfügen: „Bayesian Reasoning [...] can be achieved by children when information is presented in natural frequencies“. Allerdings natürlich nur in einer Form, die ihrer Alltagserfahrung entspricht. Der Umgang (Zählen und Rechnen) mit natürlichen Zahlen und Mengen und einfache logische Folgerungen werden in dieser Entwicklungsstufe ausgebaut (vgl. Inhelder & Piaget, 1969). In dieser Aufgabenformulierung mit natürlichen Zahlen gibt es eine Handlung, anhand derer ein konkretes Ergebnis gewonnen werden kann. Dieses konkrete Operieren mit natürlichen Zahlen und Mengen scheint bereits im Alter von 12 Jahren bei vielen Schülern gut entwickelt zu sein. Es wird gefolgert, dass Häufigkeitsschätzungen bei Bayesianischen Situationen ähnlich gut erfolgen wie bei Erwachsenen.

Durch die schriftliche Angabe von Lösungswegen (ähnlich der „Write aloud“-Protokolle bei Gigerenzer & Hoffrage, 1995, vgl. 4.1.1) konnten auch die „nicht-Bayesianischen“ Strategien genauer identifiziert werden, die bei den Lösungsprozessen auftraten. Es waren ähnliche Strategien wie die bei Gigerenzer und Hoffrage identifizierten zu beobachten. Ein weiterer Befund war, die deutliche Abnahme „nicht-Bayesianischer“ Strategien mit höherem Alter und die Durchsetzung adäquaten Bayesianischen Denkens.

4.2 Einfluss grafischer Darstellung auf Bayesianisches Denken

Die bisherigen Studien zeigten unter verschiedenen Gesichtspunkten, dass die Performanz des Bayesianischen Denkens und zugrunde liegende kognitive Prozesse vom numerischen Format der Informationen abhängen. Neben der Bedeutung des Informationsformates wurde bereits

der Nutzen von grafisch-visueller Repräsentation betont (vgl. 3.2.). In Publikationen zum Bayesianischen Denken wurden diverse Vorschläge für häufigkeitsbasierte, grafische Darstellungen gegeben, jedoch oft nur zur Veranschaulichung für den Leser verwendet (z.B. Gigerenzer & Hoffrage, 1995; Fiedler, 2000). Die Hauptfragestellung, die den im Folgenden dargestellten Studien zugrunde liegt, ist, inwieweit die positiven Wirkungen „verbal-numerischer“ Häufigkeitsübersetzungen noch gesteigert werden können durch „grafisch-visuelle“ Darstellung der relevanten Informationen. Man könnte es auch so formulieren: Es geht um die Frage, inwieweit die Darbietungsform „optimiert“ werden kann, um Bayesianisches Denken zu unterstützen.

Zunächst werden empirische Befunde für die passive und aktive Verwendung von Häufigkeitsrasterdarstellungen (Cosmides & Tooby, 1996) gezeigt. Es gibt – soweit mir bekannt – kaum Ergebnisse, die den Nutzen verschiedener grafischer Darstellungen für intuitive Wahrscheinlichkeitsurteile (ohne Training²⁷) systematisch vergleichen. In Experimenten, die ich in Kooperation mit Silke Atmaca und Stefan Krauss durchführte, wurde diese Frage untersucht (vgl. 4.2.2.).

4.2.1. Aktive und passive Verwendung des Häufigkeitsrasters

Wie bereits in 4.1.2. dargestellt, haben Cosmides & Tooby (1996) durch Übersetzung der Wahrscheinlichkeitsversion einer medizinischen Diagnoseaufgabe und der Fragestellung in eine Häufigkeitsversion die Performanz bei Bayesianischen Urteilen erheblich verbessern können. In einem weiteren Experiment (ebenda, S.33, Exp.4) wurde zusätzlich als *aktives* bzw. *passives* Hilfsmittel eine grafische Rasterdarstellung verwendet. Die Häufigkeitsversion (inkl. alle Änderungen a. bis e., vgl. 4.1.2) wurde auf eine Referenzpopulation von 100 bezogen und im unteren Teil folgendermaßen abgeändert („aktive“ grafische Version):

[...] Stellen Sie sich eine zufällige Stichprobe von 100 Amerikanern vor, die getestet werden sollen. Sie wurden in einer Lotterie ausgewählt, ohne dass die Durchführenden Informationen über den Gesundheitszustand der Personen hatten. Die 100 unten abgebildeten Kästchen stellen die zufällige Stichprobe von 100 Amerikanern dar, jedes Kästchen eine Person.

Unter Verwendung der Kästchen, stellen Sie bitte nun die gegebenen Informationen dar. Um anzuzeigen, dass eine Person tatsächlich krank ist, setzen Sie einen Kreis um das Kästchen, das die Person darstellt. Um anzuzeigen, dass eine Person ein positives Testergebnis erhalten hat, füllen Sie das Kästchen aus, das die Person darstellt.

Gehen Sie also nun so vor:

(1) Setzen Sie Kreise für alle Personen, die krank sind.

(2) Füllen Sie die Kästchen aus, die Personen mit positivem Test darstellen.

*(3) Beurteilen Sie, wie viele die Krankheit haben von denen, die ein positives Testergebnis erhalten?
_____ von _____*

Dem Text folgte ein leeres Raster mit 100 Kästchen (10 Zeilen und 10 Spalten).

In einer weiteren Version („passive“ grafische Version) wurde die Aufforderung und Anleitung zur selbstständigen Erstellung ersetzt mit: „*Wir haben die gegebenen Informationen in den unten abgebildeten Kästchen dargestellt....*“. Die Rasterdarstellung wurde entsprechend vorgegeben (vgl. Abb.I.6) und erklärt: „*Wir haben Kreise um das Kästchen, das die Person darstellt, gesetzt, wenn sie tatsächlich krank ist. Wir haben es ausgefüllt, wenn die Person ein positives Testergebnis erhalten hat...*“.

²⁷ Von Bea (1995) und Sedlmeier (1999) wurden verschiedene grafische Darstellungen im Rahmen von Trainingsstudien verglichen. Diese werden in Kapitel III ausführlich behandelt.

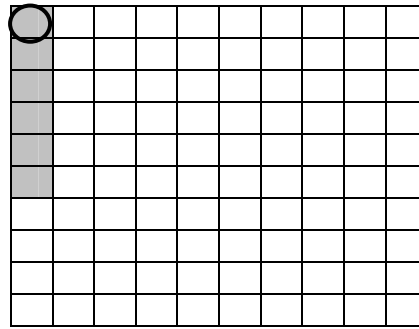


Abb. I.6: Grafische Darstellung der Diagnoseaufgabe in einem Rasterdiagramm

Ergebnisse

Die „aktive“ grafische Version, welche die Versuchspersonen zur selbstständigen Konstruktion einer Häufigkeitsdarstellung der Information zwang, erhöhte die Performanz der Urteile auf 92%. Die „passive“ grafische Version brachte jedoch kaum Steigerung zur reinen Textversion (76%; ohne grafische Darstellung: 72%). Cosmides und Tooby interpretierten ihre Befunde so, dass angebotene grafische Darstellungen nicht unbedingt benutzt werden, obwohl sie hilfreich wären. Wenn sie aber gezwungenermaßen selbstständig (unter Anleitung) erzeugt werden, können sie zu weiteren Steigerungen der Urteilsperformanz führen.

4.2.2. Vergleich verschiedener grafischer Häufigkeitsdarstellungen

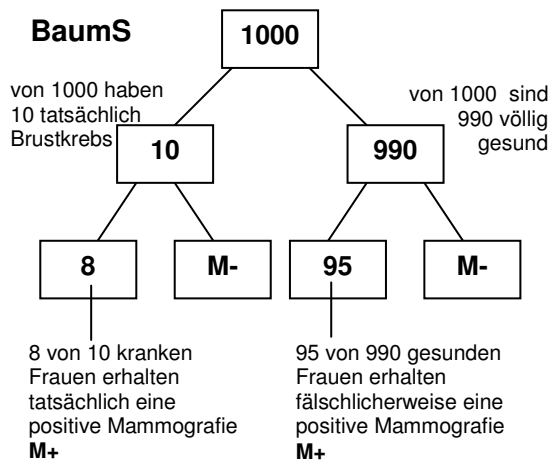
Ausgehend von den Ergebnissen von Gigerenzer und Hoffrage (1995; vgl. 4.1.1.), die im Durchschnitt eine Performanzerhöhung auf 46% (von 16%) durch die verbal-numerische Häufigkeitsübersetzung aufzeigten, habe ich in Zusammenarbeit mit Silke Atmaca und Stefan Krauss systematisch verschiedene grafische Darstellungen zur Visualisierung u.a. der Mammografieaufgabe in der Häufigkeitsversion untersucht.²⁸ Wir verbanden jeweils verschiedene grafische Darstellungsformen (Baumdiagramm, Tabelle, Flächendiagramm) mit dem Aufgabentext in Häufigkeiten und der Häufigkeitsfragestellung und legten die verschiedenen Versionen jeweils ca. 40 Studenten zur Bearbeitung vor.

Alle Versionen begannen zunächst mit einem einleitenden Text, der die Situation beschreibt. Für die Mammografieaufgabe wurde diese „cover story“ erzählt:

Stellen Sie sich bitte vor, Sie sind ReporterIn einer Frauenzeitschrift und möchten einen Artikel über Brustkrebs schreiben. Sie recherchieren auch darüber, was von dem Test zu halten ist, der im Rahmen von Routineuntersuchungen eingesetzt wird, um Brustkrebs zu entdecken. Ihr besonderes Interesse gilt der Frage, was es bedeutet, wenn eine Frau bei einem solchen Test (in der Regel eine Mammografie) ein positives Ergebnis (welches Brustkrebs anzeigt) erhält. Ein Arzt erklärt Ihnen die Situation mit folgender Zeichnung anhand einer imaginären Stichprobe von 1000 symptomfreien Frauen zwischen 40 und 50:

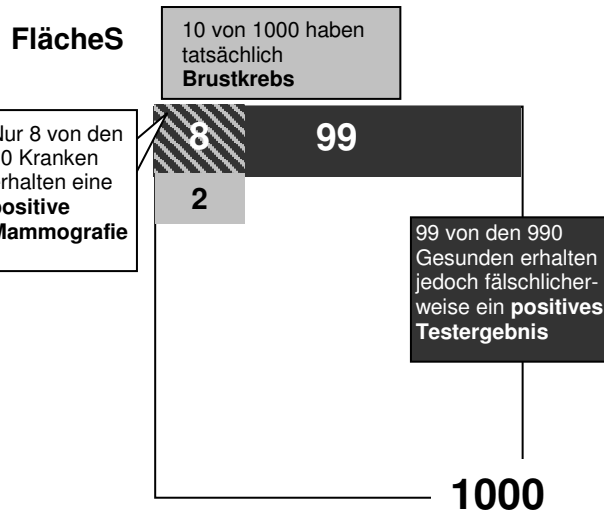
²⁸ Das Mammografieproblem im Standard-Häufigkeitsformat wurde zu ca. 42% korrekt gelöst (Gigerenzer & Hoffrage, 1995, S.694).

Folgende Darstellungsformen wurden zur weiteren Situationsbeschreibung verwendet (vgl. Abb.I.7-11):



TabS

Von 1000 Frauen	haben tatsächlich Brustkrebs	sind völlig gesund
zwischen 40 und 50 ohne Symptome	10	990
davon erhalten jeweils eine positive Mammografie	8	95



TabK

Von 1000 Frauen	haben tatsächlich Brustkrebs	sind völlig gesund
erhalten eine positive Mammografie	8	95
erhalten eine negative Mammografie	2	895

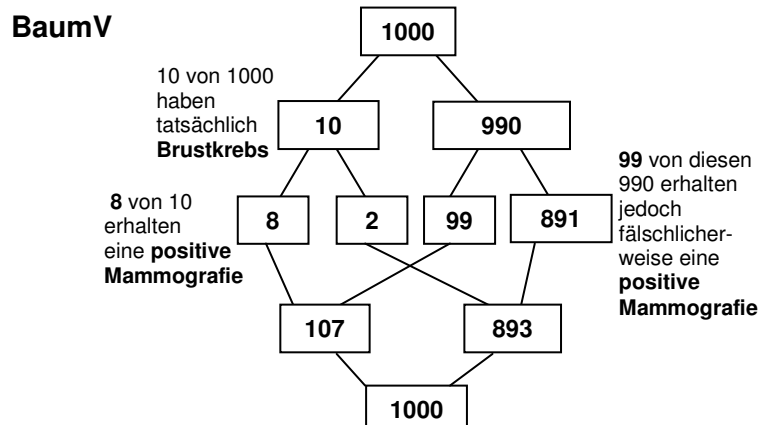


Abb.I.7 bis 11: Verschiedene untersuchte Häufigkeitsdarstellungen des Mammografieproblems

- *Häufigkeitsbaum mit Standard-Information (BaumS):*
Dieser Baum enthält genau dieselben Häufigkeitsinformationen wie die Textversion von Gigerenzer und Hoffrage (1995; vgl. 3.2.) für die Grundgesamtheit (GG), das Vorliegen der Hypothese bzw. nicht (H bzw. $\neg H$), das Vorliegen des Datums und der Hypothese bzw. des Datums und nicht der Hypothese ($D \wedge H$ bzw. $D \wedge \neg H$).
- *Häufigkeitstabelle mit Standard-Information (TabS):*
Es handelt sich um eine Tabelle mit identischer Informationsmenge²⁹ wie bei BaumS.
- *Häufigkeitsflächendiagramm mit Standard-Information (FlächeS):*
In diesem Flächendiagramm werden die numerischen Relationen zusätzlich durch die Größenverhältnisse der Flächen repräsentiert. Die Informationsmenge ist {GG, H, $\neg H$, $D \wedge H$, $D \wedge \neg H$ }, also identisch zum Standard-Informationsmenu.
- *Häufigkeitstabelle mit konjunktiver Information (TabK):*
Diese Tabelle wird auch als Vierfeldertafel bezeichnet (ohne Randhäufigkeiten). Im Unterschied zur Standard-Information enthält sie eine andere Informationsmenge: {GG, $D \wedge H$, $D \wedge \neg H$, $\neg D \wedge H$, $\neg D \wedge \neg H$ } (vgl. Fiedler, 2000).
- *Häufigkeitsbaum mit vollständiger Information (BaumV):*
Dieser „vollständige Baum“, der eine Erweiterung des Standard-Häufigkeitsbaums darstellt (vgl. Wassner & Martignon, 2002), enthält eine maximale Menge der abzuleitenden Häufigkeitsinformationen in der dichotomen Struktur: {GG, H, $\neg H$, $D \wedge H$, $D \wedge \neg H$, $\neg D \wedge H$, $\neg D \wedge \neg H$, D, $\neg D$ }. Die Informationsmenge ist identisch zu einer Vierfeldertafel mit Randhäufigkeiten (vgl. Abb. I.3).

Folgendes Häufigkeitsverhältnis und eine Wahrscheinlichkeitsschätzung sollten daraufhin jeweils angegeben werden:

Wie viele der Frauen, die ein positives Testergebnis bei der Mammografie erhalten, haben also tatsächlich auch Brustkrebs? _____ von _____

Was bedeutet dies nun für eine symptomfreie Frau zwischen 40 und 50, die bei einer Routineuntersuchung einen positiven Mammographiebefund erhalten hat?

Die Wahrscheinlichkeit, dass diese Frau tatsächlich Brustkrebs hat, beträgt: _____ %

Ergebnisse

Abb.I.12 zeigt die Performanz der Versuchspersonen bei der Häufigkeitslösung bei den verschiedenen Darstellungsversionen der Mammografieaufgabe. Die Darstellungsformen BaumS, TabS und TabK erhöhten die Performanz gegenüber der Textversion um 19 bis 26 Prozentpunkte. Die Darstellungsform FlächeS um 13 Prozentpunkte. Das vollständige Baumdiagramm bewirkte eine Performanzerhöhung um 39 Prozentpunkte auf eine fast doppelt so hohe Performanzrate wie ohne grafische Darstellung.

Eine korrekte Antwort auf die im Anschluss gestellte Frage nach einer entsprechenden Wahrscheinlichkeit, die durch Division der Häufigkeiten in der ersten Antwort erhalten wird, konnten durchschnittlich 26 Prozentpunkte weniger Versuchspersonen geben.

²⁹ Mit „Informationsmenge“ bezeichne ich die tatsächlich gegebenen numerischen Informationen. Man beachte: Der „Informationsgehalt“ ist in jeder Version analog, da durch die gegebenen Informationen jeweils alle weiteren ableitbar sind. (insbesondere ist keine Version unterdeterminiert, vgl. Diagnoseaufgabe bei Casscells et al., 1978, vgl. 4.1.2)

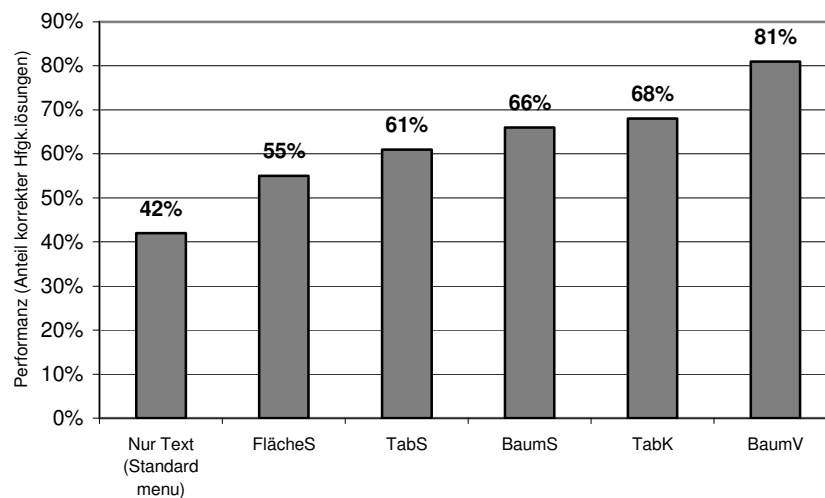


Abb.I.12: Performanz bei verschiedenen Darstellungsversionen der Mammografieaufgabe

Diskussion

Die Befunde bestätigten die Vermutung, dass eine Visualisierung in Verbindung mit dem sprachlich-numerischen Häufigkeitsformat weitere deutlich positive Performanzeffekte bei Bayesianischem Denken hervorrufen kann. Die grafischen Darstellungen mit zur Textversion identischer Informationsmenge (BaumS, TabS und FlächeS) unterschieden sich nur geringfügig: Der Baum ist dabei tendenziell am performanzsteigerndsten zu bewerten, dann die Tabelle und das Flächendiagramm. Dennoch lassen sich aus diesen Befunden in diesem speziellen Aufgabenkontext keine Folgerungen für eine generelle Überlegenheit einer Darstellungsform ableiten. Eine Variation der Informationsmenge bei gleichbleibender Anzahl der Informationen in der Tabellendarstellung in der Form, dass nur noch „konjunktive“ Häufigkeiten enthalten sind (TabK), führte zu keiner entscheidenden Verbesserung. Erst eine Erhöhung der Informationsmenge im Baumdiagramm bewirkte eine weitere deutliche Steigerung der Performanz. Neben Häufigkeitsformat und grafischer Darstellung ist offensichtlich das Bayesianische Denken auch von der Menge der direkt zugänglichen Informationen abhängig.

Dieser Befund ist nicht so trivial, wie er zunächst erscheinen mag. Es gibt auch gute Gründe anzunehmen, dass mehr Information die Verarbeitung erschweren kann, v.a. wenn der Urteilende keine intuitive Verarbeitungsstrategie besitzt. Dem scheint aber nicht so zu sein. Allerdings muss beachtet werden, dass der vollständige Baum (BaumV) bereits eine für die Beurteilung wichtige Information enthält, die in allen anderen untersuchten Darstellungen erst als Summe zu berechnen ist: Die Summe aller positiven Befunde (107). In diesem Sinne spiegelt er das vollständige Ergebnis des „Natural sampling“-Prozesses wider. So würde es gewissermaßen bei optimaler Aufzeichnung bzw. Erinnerung vorliegen.

5 Fazit: Anwendungen der Ergebnisse des Forschungsprogramms „Ökologische Rationalität“ zum Bayesianischen Denken

Wenn man sich unter Berücksichtigung der einschlägigen Befunde des Forschungsprogramms „Ökologische Rationalität“³⁰ noch einmal der pragmatischen Ausgangsfrage zuwendet: „Haben Menschen kognitive Mechanismen, um gut mit Bayesianischen Situationen umgehen zu können?“, kann weder die Haltung der Aufklärung noch die des „Heuristikprogramms“ in der kognitiven Psychologie unterstützt werden. Einerseits scheint es völlig richtig, dass der Mensch kein guter „Wahrscheinlichkeitstheoretiker“ ist. Er kann Wahrscheinlichkeiten tatsächlich nicht gut verarbeiten bzw. auf ihrer Grundlage Bayesianische Situationen beurteilen. Andererseits haben die Befunde gezeigt, dass Menschen gute *intuitive Wahrscheinlichkeitsschätzungen* auf Basis von Häufigkeitsinformationen abgeben können, ohne mit der Wahrscheinlichkeitstheorie vertraut zu sein. Insofern ist die Ausgangsfrage mit einem klaren „Ja“ zu beantworten, wenn Häufigkeiten die Grundlage der Beurteilung von Unsicherheit sind.

Es wird deutlich, dass die Frage verschiedene Beantwortung nach sich zieht, je nachdem wie „Bayesianisches Denken“ definiert wird, etwa als „Gebrauch mathematischer Werkzeuge“ oder als „Gebrauch intuitiver Werkzeuge“ zum Umgang mit entsprechenden Situationen. Beides kann ein didaktisches Ziel sein. Jedoch gibt es (mindestens) zwei Argumente dafür, dass das Zurückgreifen auf „intuitive Werkzeuge“ eine stärkere Betonung, v.a. bei stochastischen Laien, erfahren sollte. Sie können erstens bereits ohne didaktische Bemühungen gut „funktionieren“ und zweitens können sie eine treibende Kraft für die Entwicklung des verständigen Gebrauchs mathematischer Werkzeuge sein. In diesem Sinne soll es im Lernprozess zu einem dynamischen „Wechselspiel zwischen Intuitionen und Mathematik“ kommen (vgl. Fischbein, 1975; Borovcnik, 1992). Im weiteren Verlauf dieser Arbeit werde ich hauptsächlich auf die Möglichkeiten eines didaktischen Zugangs mit dem „Häufigkeitskonzept“ in der *Stochastikausbildung* eingehen und weitere empirische Ergebnisse dazu präsentieren.

Zuvor möchte ich noch kurz auf Anwendungen der Forschungsergebnisse in Bereichen der *Risikokommunikation* eingehen. Die Folgerung, die Vertreter des Häufigkeitskonzeptes zogen, war: Wenn Menschen auf Grundlage von natürlichen Häufigkeiten intuitiv besser in der Lage sind, Wahrscheinlichkeitsschätzungen abzugeben, dann können sie auch Risiken, die mit Häufigkeiten kommuniziert werden, besser verstehen. Das gilt sowohl für Experten wie auch Laien (vgl. für einen Überblick Hoffrage, Lindsey, Hertwig & Gigerenzer, 2000)

Ein großer Anwendungsbereich für Bayesianisches Denken mit potenziell hoher Tragweite für den Einzelnen ist, wie in der Einführung bereits beispielhaft angedeutet, die *medizinische Diagnostik*. Hoffrage & Gigerenzer (1998) legten 48 Ärzten (mittl. Berufserfahrung von 14 Jahren) diagnostische Situationen zur Beurteilung vor, die spezielle medizinische Testverfahren betrafen (u.a. auch Mammografie). Es ergab sich auch das in den bisherigen Studien erhaltene Bild der Performanzabhängigkeit vom verwendeten Informationsformat. Bei Wahrscheinlichkeitsversionen der Mammografieaufgabe lagen die Urteile von 2/3 der Ärzte völlig falsch (über 50%; korrekter Wert: 8%), bei Häufigkeitsversionen nur noch von 21% (vgl. Abb.I.13). Hoffrage und Gigerenzer stießen meist auf negative Grundhaltung der medizinischen Experten gegenüber Stochastik bzw. eigener mathematischer Fähigkeiten. Sie waren meist nicht bereit, sich solchen Fragen zu

³⁰ im Folgenden meist als „Häufigkeitskonzept“ bezeichnet.

öffnen. Diese negative Grundhaltung und Selbstbewertung konnten durch Häufigkeitsdarstellungen oft überwunden werden.

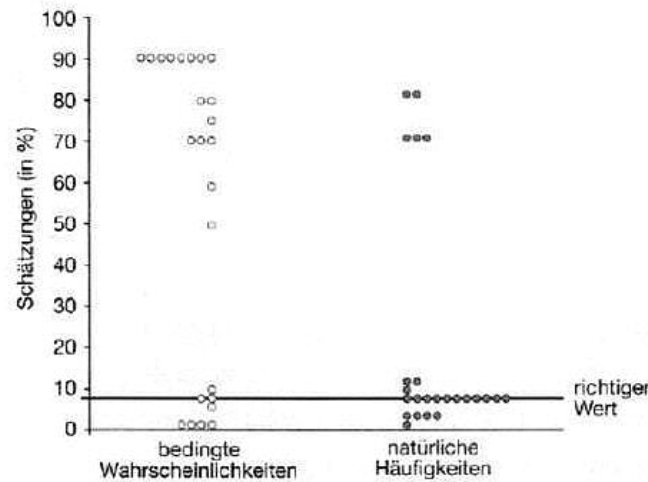


Abb.I.13: Wahrscheinlichkeitsschätzungen befragter Ärzte zur Aussagekraft eines Mammografiebefundes aufgrund von Wahrscheinlichkeits- bzw. Häufigkeitsinformationen (vgl. Gigerenzer, 2002)

Eine Studie von Gigerenzer, Hoffrage & Ebert (1998) zeigte, dass in so entscheidenden Fragen, wie der Aussagekraft positiver HIV-Tests, auch professionelle AIDS-Berater versagten. Ein studentischer Mitarbeiter unterzog sich freiwillig 20 AIDS-Tests mit anschließendem Beratungsgespräch und stellte u.a. die Frage nach der Sicherheit einer HIV-Infektion nach positivem Testergebnis. 10 Berater sprachen von 100%iger Sicherheit, weitere 5 von „99,9% oder höher“. Alle weiteren Angaben lagen über 90%. Die Mehrheit erklärte, es könne keine falsch-positiven Diagnosen geben, weil immer ein Bestätigungstest erfolgt. Für Personen mit einem niedrigen Risikoverhalten sind aber diese Aussagen nicht richtig (vgl. Gigerenzer et al. 1998, S.200). Nur ein Berater hat überhaupt erklärt, dass die Aussagekraft des Tests von der Prävalenz für eine HIV-Infektion abhängt, also von der entsprechenden A-priori-Risikoeinschätzung. Keiner der Berater hatte versucht, die Informationen auf Basis von Häufigkeiten zu erklären, die diesen Zusammenhang sehr gut hätten verständlich machen können. Diese Studien belegten große Defizite in der Arzt-Patient-Kommunikation, die mit dem „Häufigkeitskonzept“ verbesserbar sind (vgl. auch Hoffrage, Kurzenhäuser & Gigerenzer, 2000).

Juristische Urteilsbildung hat ähnlich der medizinischen eine besondere Tragweite für einzelne Personen. Die Einschätzung und Kommunikation von Unsicherheit sind von entscheidender Bedeutung. Auch in diesem Bereich kann die Verwendung von Häufigkeitsdarstellung wertvolle Hilfe leisten für eine verbesserte Urteilsbildung. Lindsey, Hertwig und Gigerenzer (im Druck) legten angehenden Richtern und Jurastudenten detaillierte Gerichtsfälle vor, in denen jeweils DNA-Proben des Angeklagten mit am Tatort gefundenen Spuren in einer DNA-Analyse übereinstimmten. Die Fälle waren so konstruiert, dass neben dem Ergebnis der DNA-Analyse kaum verwertbare Indizien vorhanden waren. In der Fallbeschreibung war auch ein Sachverständigengutachten mit statistischen Bewertungen enthalten, z.B.:

Auf der Basis des DNA-Profiles der am Tatort an der Leiche gefundenen Blutspur erstellte der DNA-Sachverständige ein Gutachten. Dort gab er zu Protokoll, dass das DNA-Profil der DNA-Probe des Angeklagten mit dem DNA-Profil der am Tatort gefundenen Spur übereinstimmt. Zusätzlich führte er aus, dass in einer Stadt von der Größe Berlins ungefähr eine Million Männer als mögliche Täter in Frage kommen. Im Falle der Schuld des Angeklagten zeigt die verwendete DNA-Analyse mit an Sicherheit grenzender Wahrscheinlichkeit eine Übereinstimmung des DNA-Profiles des Angeklagten

mit dem in der Blutprobe an der Leiche sichergestellten genetischen Fingerabdruck. Die Wahrscheinlichkeit, dass der genetische Fingerabdruck eines Unschuldigen rein zufällig mit diesem DNA-Profil übereinstimmt oder dass aufgrund von Laborfehlern eine DNA-Übereinstimmung fälschlich festgestellt wurde, beträgt 0,3%.³¹

Diese statistischen Angaben lassen eine Abschätzung der Wahrscheinlichkeit zu, dass der Angeklagte allein aufgrund der DNA-Übereinstimmung ohne weitere Indizienberücksichtigung der Verursacher der Spur ist. Sie beträgt aber lediglich ca. 0,03% (sic!³²). Deutlicher wird diese erstaunlich geringe Wahrscheinlichkeit, wenn eine Häufigkeitsüberlegung angestellt wird (vgl. Abb.I.14).

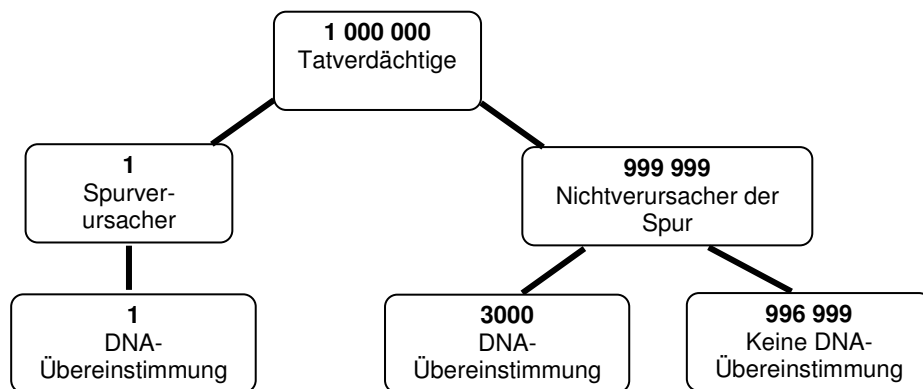


Abb.I.14: Häufigkeitsdarstellung zur Aussagekraft eines DNA-Gutachtens

In der obigen Form mit Wahrscheinlichkeitsangaben schätzte lediglich unter 1% der Studenten und 13% der angehenden Richter die Wahrscheinlichkeit richtig ein. War das Gutachten mit Häufigkeiten formuliert (ohne die grafische Baumdarstellung), wurde von 44% der Studenten und 68% der Richter die Wahrscheinlichkeit richtig angegeben. Interessant ist auch, dass mit Wahrscheinlichkeitsangaben 55% der Studenten und 45% der Richter den Angeklagten verurteilt hätten, mit Häufigkeitsangaben jedoch nur 33% bzw. 32%.

Grundsätzlich wird die DNA-Evidenz als sehr aussagekräftiges Indiz gesehen und es besteht die Gefahr einer Überbewertung bei der Gesamtbeurteilung einer juristischen Fragestellung. Die Häufigkeitsdarstellung kann helfen, Überbewertungen zu vermeiden und eine transparente Kommunikation (z.B. Gutachter - Richter) von statistischer Information bei juristischen Fragen zu ermöglichen (vgl. Krauss & Hertwig, 2000).

Gigerenzer und Kollegen zogen aus diesen Ergebnissen praktische Implikationen: Erstens sollte sich die Ausbildung von Medizinern und Juristen bzw. von allen, die mit Risiken umgehen müssen, die Vorteile von Häufigkeitsdarstellungen zu Nutze machen, z.B. durch entsprechende Instruktion und Training (vgl. Kurzenhäuser & Hoffrage, 2002; Sedlmeier, 1999; vgl. Kap.III dieser Arbeit). Zweitens sollte auch die Risikokommunikation zwischen Ärzten und Patienten bzw. vor Gericht, allgemein zwischen Experten und Laien, auf Häufigkeitsebene erfolgen. In diesem Falle ist es ganz entscheidend, dass die einfachste und verständlichste Art der Kommunikation gewählt wird, um die tatsächlichen Risiken transparent machen zu können. Einige Fortschritte in dieser Richtung wurden bereits gemacht.³³

³¹ Diese Fehlerrate basiert auf Angaben bei Koehler, Chia & Lindsey (1995).

³² Das liegt an der sehr hohen Zahl möglicher Tatverdächtiger von 1 Million, die in praxi natürlich oft erheblich kleiner ist.

³³ Tatsächlich gibt es z.B. entsprechende Empfehlungen im englischen Rechtssystem (vgl. Redmayne, 1998).

Gigerenzer gibt in dem Buch „Das Einmaleins der Skepsis“ (2002) einen umfassenden Überblick über den alltäglichen Umgang mit Risiken. Sein „didaktisches Programm“ besteht aus drei Schritten: Weise erstens die Leute auf die vielen Formen der Unsicherheit hin, insbesondere auf die, die sie gar nicht als solche erkennen. Lehre sie zweitens, diese Unsicherheit als Risiken zu beurteilen. Zeige ihnen drittens verschiedene Repräsentationen, um ihnen den verständlichsten Weg zu einer vernünftigen Risikobeurteilung und Kommunikation zu ermöglichen.

Der in diesem Kapitel gegebene, auf Bayesianische Situationen fokussierte Überblick über die wechselvolle Kontroverse zur untrennbaren Verbindung menschlichen Denkens und dem Phänomen der Unsicherheit führte durch den Aspekt der „ökologischen Rationalität“ zu einer unzweifelhaften Antwort: *Repräsentation* ist der Schlüssel zum Denken unter Unsicherheit. Repräsentation ist auch einer der Eckpfeiler für didaktische Prozesse. Es ist deshalb dringend geboten und das Ziel dieser Arbeit, die wichtigen Erkenntnisse der Kognitionspsychologie auch in der Didaktik bzw. Lehr-Lernforschung anzuwenden und zu etablieren (vgl. Martignon & Wassner, 2001).

KAPITEL II

DIDAKTISCHE UND CURRICULARE ANALYSEN

ZUM THEMENGEBIET „BEDINGTE WAHRSCHEINLICHKEIT UND BAYESSCHE REGEL“

Im vorausgehenden Kapitel wurden Ansätze zur Untersuchung und Erklärung kognitiver Prozesse des Menschen im Zusammenhang mit Bayesianischem Denken diskutiert. Aus der Kontroverse über diese Frage wurden – wie bereits ausgeführt – unterschiedliche Folgerungen über didaktische Möglichkeiten gezogen. In der Mathematikdidaktik hat man trotz oder gerade wegen eines Bewusstseins für besondere Hürden in diesem Bereich konstruktiv didaktische Analysen und Vorschläge gemacht, wobei allerdings kaum empirisch vorgegangen wurde. Eine vielfach vertretene Einschätzung, die oft mit Ergebnissen des „Heuristikprogramms“ (Kap.I.2) untermauert wurde, war: “A conceptual grasp of Bayesian methods rests on an understanding of conditional probability, a notoriously difficult idea. [...] There is, of course, no easy path to understanding inference” (Moore, 1997). Die Relevanz des Bayesianischen Denkens im Hinblick auf ihren Bildungsgehalt bzw. -wert für die Schüler wurde – im Gegensatz etwa zur Relevanz in der Kognitionspsychologie – auch kontroverser eingeschätzt.

Im Folgenden werden Zugänge von Seiten der Mathematikdidaktik zu Themen um „Bedingte Wahrscheinlichkeit und dem Satz von Bayes“³⁴ besprochen. Zunächst wird kurz dargestellt, welche Bedeutung diesen Inhalten in derzeitigen deutschen Gymnasiallehrplänen und curricularen Empfehlungen zugesprochen wird, und kritisch bewertet. Im zweiten Abschnitt des Kapitels werden einige didaktische Ansätze näher erläutert. Dabei werden sowohl bereits in Schulbüchern umgesetzte Konzepte, als auch auf Artikelbasis existierende Vorschläge dargelegt. Der dritte Teil stellt einen Ansatz auf dem Konzept der „natürlichen Häufigkeiten“ vor, dessen Besonderheit die Integration der Resultate des kognitionspsychologischen Forschungsprogramms „Ökologische Rationalität“ (Kap.I.3) ist.

1 Curriculare Diskussion

1.1 Aktuelle Einbettung in den gymnasialen Lehrplänen³⁵

In den Lehrplänen und Richtlinien der Bundesländer des gymnasialen Bildungsgangs sind Themen der Wahrscheinlichkeitsrechnung zu unterschiedlichen Zeitpunkten und mit unterschiedlicher Inhaltsgewichtung enthalten. Während in vielen Bundesländern Ziele einer Allgemeinbildung in Wahrscheinlichkeitsrechnung bereits Bestandteil der Pläne der frühen Sekundarstufe I sind, werden sie in manchen Ländern überhaupt erst in der Sekundarstufe II berücksichtigt. Eine detaillierte Analyse aller Lehrpläne und Lehrplanentwürfe der Bundesländer würde den Rahmen dieser Arbeit sprengen. Im Folgenden soll nur kurz auf einige Aspekte, die im Zusammenhang mit dem Problemfeld „Bedingte Wahrscheinlichkeit und Satz von Bayes“ stehen, eingegangen werden.

³⁴ Diese Bezeichnung wird i.w.S. auch verwendet, wenn bedingte Wahrscheinlichkeit und der Satz von Bayes nicht formal entwickelt werden (vgl. Abb.II.1). In Kapitel I behandelte Situationen zum Bayesianischen Denken sind entsprechende Beispiele.

³⁵ Diese Bezeichnung wird für die gesetzlichen Planungsgrundlagen in den einzelnen Bundesländern verwendet stellvertretend für Richtlinien, Rahmenplan, Bildungsplan, etc.

In der *Sekundarstufe I* sind einige typische Stichworte in den Lehrplänen zur Wahrscheinlichkeitsrechnung „Zufallsexperiment“, „relative Häufigkeit“, „Wahrscheinlichkeit“, „Urnenmodell“, „Baumdiagramm“, „Pfadregeln“ (z.B. Kultusministerium NRW, 1993). Der Begriff der „bedingten Wahrscheinlichkeit“ wird im Allgemeinen nicht explizit entwickelt, sondern – wenn überhaupt – erfolgt eine implizite Behandlung von bedingter Wahrscheinlichkeit im Kontext von „mehrstufigen Zufallsversuchen“ und Pfadregeln anhand von Baumdiagrammen. Anwendungen des Satzes von Bayes werden in der Sekundarstufe I normalerweise ausgeklammert. Ein Gegenbeispiel ist der Lehrplan Mathematik für die Sekundarstufe I aus NRW (1993), der dieses Themengebiet (optional) für die gymnasiale Doppeljahrgangsstufe 9/10 explizit enthält (vgl. Abb.II.1). Auch in der neuen Entwurfsfassung vom Oktober 2003 ist die Anwendung der Bayesregel an authentischen Beispielen als Kompetenzerwartung am Ende der Jahrgangsstufe 10 enthalten: „Schülerinnen und Schüler nutzen die Kenntnisse über zweistufige Zufallsexperimente, um statistische Aussagen in authentischen Texten zu interpretieren und zu beurteilen (z.B. Gesundheitstests)“.³⁶ Neben dem üblichen Aufbau der Wahrscheinlichkeitsrechnung gemäß *frequentistisch-klassischer* Wahrscheinlichkeitsinterpretation werden auch Elemente des *subjektivistischen Ansatzes* eingebaut (Abb.II.1, letzter Satz). Diese Tendenz findet sich auch in neuen Entwürfen der niedersächsischen Rahmenrichtlinien für die Sekundarstufe I. Der subjektivistische Ansatz wird in Klasse 9 im Zuge der *Bayes-Statistik* entwickelt. Der Bayes-Statistik wird in Klasse 10 der Alternativtest anhand der Binomialverteilung gegenübergestellt (vgl. Eichler, 2003; Cukrowicz & Zimmermann, 2001).

Alternative (a)

Probleme aus dem Themenkreis der Bayesschen Regel werden mit Hilfe von Baumdiagrammen bearbeitet. Dabei soll die Bayessche Regel weder formal behandelt noch bewiesen werden (das ist der gymnasialen Oberstufe vorbehalten). Hier geht es um ein inhaltliches Verständnis der Zusammenhänge, wobei der Begriff der bedingten Wahrscheinlichkeit intuitiv verwendet wird. Für statistisches Denken wird die Regel relevant, wenn man studiert, wie sich die Wahrscheinlichkeiten verschiedener Alternativen auf Grund von Beobachtungen ändern.

Abb.II.1: Der Themenkreis Bayesregel in der Sekundarstufe I (vgl. Kultusministerium NRW, 1993, S. 57)

Einige weitere Lehr- und Rahmenpläne zur *gymnasialen Sekundarstufe I* verschiedener Bundesländer seien als repräsentativer Querschnitt ohne Anspruch auf Vollständigkeit im Hinblick auf die Entwicklung von Inhalten der Wahrscheinlichkeitsrechnung und insbesondere der Grundlagen für Themen zu „bedingter Wahrscheinlichkeit und Anwendung der Bayesregel“ kurz umrissen.

In der gymnasialen Sekundarstufe I in Bayern wurde bisher völlig auf Wahrscheinlichkeitsrechnung verzichtet. Im neuen Rahmenplan für das Gymnasium für Mathematik in Bayern (Kultusministerium Bayern, 2003) sind Inhalte für die Klassenstufe 10 zu „zusammengesetzten Zufallsversuchen“: Anwenden der Pfadregeln, Definition des Begriffs „bedingte Wahrscheinlichkeit“ und Einfluss von zusätzlichen Informationen oder Bedingungen auf die Wahrscheinlichkeit eines Ereignisses. Themen um bedingte Wahrscheinlichkeit sollen in Grund- und Leistungskursen der 12. Klasse wieder aufgegriffen werden, allerdings wird der Satz von Bayes nicht explizit erwähnt:

„Sie [die Schüler, Anm.d.Verf.] entwickeln bei der Diskussion von Themen wie z.B. der Sicherheit von Krankheitsdiagnosen ihr Urteilsvermögen weiter. Simulationen beispielsweise mit Zufallszahlen helfen, Vermutungen aufzustellen und überraschende Rechenergebnisse zu überprüfen“ (ebenda, M 12.3.2).

³⁶ vgl. NRW, Kernlehrplan Mathematikunterricht Gymnasium Sekundarstufe I, Entwurfsfassung Oktober 2003, S.21.

Im Rahmenplan Mathematik aus Brandenburg (Kultusministerium Brandenburg, 2002) finden sich für die Doppelstufe 9/10 zum Themenfeld Zufall nur sehr ungenaue Angaben:

„Das Berechnen von Wahrscheinlichkeiten für mehrstufige Zufallsversuche soll gefestigt und vertieft werden“ (ebenda, S.58).

Im Mathematik Rahmenplan aus Berlin (Kultus Berlin, 1987) für die gymnasiale Sekundarstufe I (Niveau II) findet sich keine Stochastik.

In der neuen Fassung des Hamburger Rahmenplans für das neunstufige Gymnasium (Kultus Hamburg, 2003) gehört der Umgang mit mehrstufigen Zufallsexperimenten, Baumdiagrammen und Pfadregeln zu verbindlichen Inhalten bis zur Stufe 8, wobei bereits in möglichen Lernsituationen der „Bayesianische Testgedanke“ vorkommen kann. Weitergehende Vertiefung erfolgt in der Sekundarstufe II.

In einer neuen Erprobungsfassung des Mathematik Rahmenplans für Mecklenburg-Vorpommern (Kultusministerium Mecklenburg-Vorpommern, 2002) werden als Ziele in Stufe 8 („Wahrscheinlichkeit I“) genannt:

„Die Schüler können zunehmend Aufgaben mit komplexem Charakter lösen. Sie berechnen Wahrscheinlichkeiten mit Hilfe von Baumdiagrammen und kennen Zählverfahren zur Anzahlbestimmung bei einfachen kombinatorischen Problemen“ (ebenda, S.25).

In Stufe 10 erfolgt hier bereits eine (klassische) Einführung in Elemente der beurteilenden Statistik („Wahrscheinlichkeit II“).

Im Lehrplan Mathematik für das Gymnasium in Sachsen (Kultusministerium Sachsen, 2001) befinden sich die Inhalte zu mehrstufigen Zufallsversuchen und der Benutzung von Baumdiagrammen und Pfadregeln auch in Klasse 8. Erst in der Sekundarstufe II ist bedingte Wahrscheinlichkeit Lehrplanthema (GK12: bedingte Wahrscheinlichkeit als Zusatzthema, LK12: bedingte Wahrscheinlichkeit und Satz von Bayes, aber keine Bayes-Statistik).

In den Plänen der *Sekundarstufe II* sind Inhalte des Themenbereichs „Bedingte Wahrscheinlichkeit und Satz von Bayes“ im Allgemeinen enthalten. Oftmals unterscheiden sich jedoch die Inhalte von Grund- und Leistungskursen dahingehend, dass im Leistungskursplan der „Satz von Bayes“ und seine Anwendung explizit Erwähnung finden, im Grundkursplan aber nicht. Diese wohl sehr willkürliche Abgrenzung ist um so unverständlicher, da gerade dieser Themenbereich viele interessante Anwendungs- und Alltagsbezüge eröffnet. Gerade diese Auseinandersetzung mit der Wahrscheinlichkeitsrechnung in realistischen Kontexten sollte aber, wie auch in Richtlinien oft deutlich gemacht wird, besonders ein „Markenzeichen“ der Grundkurse sein, um gerade bei mathematisch uninteressierteren Schülern eine positivere Einstellung zur Mathematik zu fördern (z.B. Kultusministerium NRW, 1999, S.26)

Ein weiterer wichtiger Aspekt des Themenbereiches ist die Anwendung im Bereich der Statistik zur Beurteilung von Hypothesen (vgl. auch Einführung). Der wachsenden Relevanz dieser statistischen Konzepte und der damit verbundenen grundlegenden Rolle der Bayesschen Denkweise wird in derzeitigen Lehrplänen allerdings meist kaum Rechnung getragen. Eine Ausnahme ist z.B. der Lehrplan aus NRW, der nicht ein Verfahren für das Testen von Hypothesen explizit vorschreibt, sondern klassische oder Bayesianische Verfahren gleichwertig gegenüberstellt. Diese offene Grundhaltung zeigt sich bereits ab Stufe 9/10 (vgl. Abb.II.1). Eine vergleichende Gegenüberstellung wird dabei für „sehr lehrreich“ gehalten (Kultusministerium NRW, 1999, S.26). Mögliche ausführlichere Vorschläge finden sich in Lehrbüchern, die im Zuge dieser Lehrpläne entwickelt wurden (vgl. Schmid & Weidig, 1996; ausführlicher in 2.2.2.).

1.2 Aktuelle Empfehlungen aus der Mathematikdidaktik

Die Diskussion um Inhalte, Ziele und Gestaltung des Stochastikunterrichts wurde im letzten Jahrzehnt auf nationaler und internationaler Ebene verstärkt geführt. Eine neue bildungspolitische Relevanz erhielten stochastische Disziplinen aus mancherlei Gründen. Hauptursachen liegen in einer Reaktion auf den rasanten Wandel der modernen Gesellschaft, die immer mehr vom Austausch von Informationen geprägt ist. Die Analyse statistischer Daten und der Einsatz stochastischer Modelle zur Entscheidungsfindung gewinnt für den mündigen „Normalbürger“, erst recht aber für „Experten“ ständig an Bedeutung, nicht zuletzt da Fehlinterpretation und die Gefahr von bewusstem Missbrauch zunehmen (vgl. z.B. Krämer, 1995; Gigerenzer, 2002). Zum anderen ist Statistik das an deutschen Universitäten meist unterrichtetste Fach, mit dem schätzungsweise jeder dritte Studierende in Berührung kommt (vgl. AK Stochastik, 2002, S.75). Ein Reformhintergrund ist auch die Verfügbarkeit neuerer Technologien, die neue Möglichkeiten für kreative Lernhandlungen der Schüler bieten (vgl. z.B. Biehler, 1991, 1995).

Aus diesen Gründen ist es eine dringende Aufgabe, zur besonderen Rolle der Stochastik in der schulischen Mathematikausbildung Stellung zu nehmen (vgl. auch Baumert, Gigerenzer & Martignon, 2004). Exemplarisch gehe ich im Folgenden auf zwei aktuelle Empfehlungen ein: Ein zur Diskussion gestelltes Papier des deutschen Arbeitskreises „Stochastik in der Schule“ der Gesellschaft für Didaktik der Mathematik (AK Stochastik, 2002) und anschließend auf die entsprechenden Teile der US-amerikanischen „Principles and Standards for school mathematics“ des National Council for Teaching Mathematics – kurz als „NCTM-Standards“ bezeichnet (NCTM, 2000).

1.2.1 Empfehlungen des Arbeitskreises „Stochastik in der Schule“

Die Empfehlungen des Arbeitskreises (AK Stochastik, 2002) betreffen eine grundlegende, sich möglichst durch alle Stufen (1.-12./13. Klasse) entfaltende stochastische Bildung, die verbindlicher Bestandteil der Lehrpläne *aller* Bundesländer werden sollte. Grundsätzlich muss dem Stochastikunterricht ein größerer Beitrag zur Realisierung allgemeiner Ziele des Mathematikunterrichts zugemessen werden. Das betrifft auch eine stärkere Vernetzung aller Disziplinen des Mathematikunterrichts mit stochastischen Denk- und Vorgehensweisen und auch die Verbindung von wahrscheinlichkeitstheoretischen und statistischen Inhalten.³⁷ Anforderungen des Stochastikunterrichts über die allgemeinen Prinzipien des Mathematikunterrichts hinaus werden spezifiziert: So besteht z.B. ein erhöhter Bedarf an kritischer Reflexion der Methoden und Ergebnisse, da oftmals mehrere Vorgehensweisen bzw. Schlussfolgerungen möglich sind oder im gesellschaftlichen Leben sogar bewusst fehlerhaft erfolgen. Herausgestellt wird auch die hohe Bedeutung des Anwendungsbezuges, so dass bereits von Anfang an die Anwendung der Inhalte im Vordergrund stehen muss. Der mathematische Kalkül und das Begriffsgefüge sollen sehr „vorsichtig“ entwickelt werden, also auf dem minimal nötigen Niveau. Zum Abschlussniveau von Primarstufe, Sekundarstufe I und Sekundarstufe II werden detaillierte Empfehlungen gegeben, sowie Mindestziele des Stochastikunterrichts und mögliche Erweiterungen und Vertiefungen³⁸ formuliert.

Speziell für eine Hinführung zu Inhalten zur „bedingten Wahrscheinlichkeit“ und Anwendung der Bayesschen Regel wird für die gymnasiale *Sekundarstufe I* nur die Modellierung mehrstufiger Zufallsvorgänge mithilfe von Baumdiagrammen und Pfadregeln

³⁷ Diese Grundforderungen waren Gegenstand der Arbeitskreistagung 2003 in Kassel, deren Vorträge und Diskussionen in einem Tagungsband im Franzbecker Verlag erscheinen werden.

³⁸ Diese Themen werden im Gegensatz zu den obligatorischen Mindestzielen hinsichtlich Realisierungsmöglichkeiten als abhängig von den Abiturstufenplänen der jeweiligen Bundesländern eingestuft.

genannt. Erst in *Sekundarstufe II* soll dieses Niveau zu einem inhaltlichen Verständnis für den Begriff „bedingte Wahrscheinlichkeit“ und für die Bayessche Regel ausgebaut werden. Zu den Mindestzielen, d.h. bereits in Grundkursen, soll gehören, dass die Schüler ein inhaltliches Verständnis für den Begriff „bedingte Wahrscheinlichkeit“ und für die Bayessche Formel erwerben und sie in Sachsituationen verständlich anwenden können. Als Erweiterung und Vertiefung (z.B. Leistungskursinhalt) sollen die Schüler das Konzept der Bayes-Analyse kennen lernen, das die Beurteilung von Hypothesen erlaubt. Dabei soll die grundlegende Rolle der Bayesschen Regel erkannt werden. Ziel ist, dass sie eine Bayes-Analyse für einen diskreten Parameterraum durchführen können (vgl. AK Stochastik, 2002, S.83).

1.2.2 Empfehlungen aus den „NCTM Standards“

Die „NCTM-Standards“ (NCTM, 2000)³⁹ bieten eine umfassende Gesamtsicht und Diskussion der gegenwärtigen Mathematikdidaktik. Die curricularen Empfehlungen basieren auf der langjährigen Diskussion von US-amerikanischen Mathematikdidaktikern und -lehrern. In dem umfassenden Dokument werden Grundpositionen und didaktische Prinzipien formuliert, sowie inhaltliche bzw. prozessuale Standards über alle Klassenstufen hinweg angegeben und mit anschaulichen Beispielen illustriert. Einer der fünf inhaltsorientierten Standards befasst sich mit „Datenanalyse und Wahrscheinlichkeit“.⁴⁰ Grundsätzlich werden für alle Jahrgangsstufen (von der Vorschule bis zur Klasse 12) Lerninhalte angeboten. Kennzeichen ist eine hohe „Daten- und Handlungsorientierung“. In alle typischen Schritte einer statistischen Untersuchung sollen die Schüler miteinbezogen werden: Planung und Erhebung, Darstellung, Zusammenfassung, Interpretation von Daten und schließlich Entscheidungen und Inferenz basierend auf Daten. Dabei sollen grundlegende Konzepte der Wahrscheinlichkeitsrechnung verstanden und angewendet werden.

Bereits in früheren „Standards“ (NCTM, 1989) wurde die Wichtigkeit eines sich über alle Klassenstufen erstreckenden Stranges zur Datenanalyse betont. Die Konkretisierung erfolgte in Curriculumprojekten, wie etwa „Data-driven mathematics“ (vgl. Burrill, 1996). Ein wichtiges didaktisches Prinzip ist die Verwendung realer Daten, die für die Schüler möglichst interessant und bedeutungsvoll sein sollen. Reale Daten geben dem Stochastikunterricht Legitimität und Bedeutung. In der Handhabung sind sie oft unangenehmer als konstruierte und gerundete Daten. Da aber die Schüler kaum Erfahrung mit Zahlen aus der „realen Welt“ im sonstigen Mathematikunterricht bekommen, sollte dieses Prinzip im Stochastikunterricht verwirklicht werden (vgl. für eine Übersicht Engel, 2001).

Inhaltlicher Zugang zu bedingter Wahrscheinlichkeit erfolgt zunächst über mehrstufige Zufallsversuche unter Verwendung von Listen, Baumdiagrammen und Flächenmodellen als Bestandteil der Klassen 6-8. Darauf aufbauend wird in den Klassen 9-12 das Konzept der bedingten Wahrscheinlichkeit und der Unabhängigkeit als Thema genannt. Als Beispiel wird eine medizinische Diagnosesituation verwendet, in dem Wahrscheinlichkeitsinformationen als Tabellen mit konkreten Häufigkeitswerten dargestellt werden sollen; jedoch ist nicht genauer ausgeführt, für welche Urteilssituationen diese Darstellungen verwendet werden können. Zur Frage statistischer Inferenzverfahren, die auf der Bayesregel beruhen, gibt es keine Aussagen.

³⁹ Im Internet verfügbar unter <http://standards-e.nctm.org>

⁴⁰ Eine deutsche Übersetzung findet sich in Borovcnik, Engel & Wickmann, 2001, S.11-42.

1.3 Bewertung

Grundsätzlich gehören mehrstufige Zufallssituationen, die die Grundlage für Probleme mit bedingter Wahrscheinlichkeit und Bayesianwendungen sind, zu den Mindestinhalten der Wahrscheinlichkeitsrechnung in der Sekundarstufe I. Wenn man diese Zufallssituationen mit realen Anwendungen bzw. realen Daten, die für Schüler interessant und bedeutungsvoll sein können (also ohne übliche Würfel- und Urnenexperimente), verknüpfen will, befindet man sich natürlicherweise beim Themenfeld „Bedingte Wahrscheinlichkeit und Satz von Bayes“. Die Probleme passen deshalb gut in die Sekundarstufe I und könnten dort auch ohne Entwicklung des mathematischen Kalküls und mit nur minimaler Erweiterung des Begriffsgefüges behandelt werden, wie in dieser Arbeit deutlich werden wird (vgl. Kap.II.3 und Kap.IV).

Ein nicht unerheblicher Grund für die Diskrepanz in der Bedeutungsbewertung von bedingter Wahrscheinlichkeit und dem Satz von Bayes liegt in der theoretischen Kontroverse („klassisch-frequentistischer vs. Bayesianisch-subjektivistischer Zugang“), die sich je nach Überzeugung in einer Betonung oder Vernachlässigung dieser Inhalte äußert (vgl. Eichler, 2003). In der heutigen statistischen Praxis werden längst beide – klassische und Bayesianische Vorgehensweisen – gebraucht (vgl. auch Einführung). In der bildungspolitischen und curricularen Diskussion hat diese seit Jahrzehnten geführte Kontroverse erst in letzter Zeit und nur partiell Eingang gefunden, womöglich auch aufgrund der wachsenden Relevanz „Bayesianischer Verfahren“. In der Mathematikdidaktik wurde hierzu ausführlich Stellung genommen (z.B. Borovcnik, 1984, 1992; Riemer, 1985, 1991; Wickmann, 1990; Buth, 1991; Götz, 1997). Die Grundsatzfrage ist, ob aufgrund eines wohl nie zu beendenden Methodenstreits der Bayesianische Zugang vernachlässigt oder die didaktischen Chancen in der *Methodenvielfalt* gesehen werden sollten.⁴¹ Letzteres bedeutet, beide Ansätze in das Stochastik-Curriculum zu integrieren und in der Verschiedenheit der Aussagen und Interpretationen Einsicht zu ermöglichen. Entsprechend wäre dies mit einer Bedeutungserhöhung des Bayesianischen Denkens in der Sekundarstufe I vorzubereiten. Konkrete Vorschläge werden im Folgenden gezeigt. Im Übrigen liegen Inferenz und Entscheidung völlig natürlich am Ende des Prozesses des Umgangs mit Daten, der als Leitidee für den Aufbau eines modernen, anwendungsbezogenen Stochastikunterrichts gelten soll („making sense of data“; vgl. NCTM, 2000).

2 Mathematikdidaktische Analyse

Schwierigkeiten, die im Themenbereich der bedingten Wahrscheinlichkeit und der Anwendung der Bayesregel auftreten können, sind als mögliche „Hürden“ für das Lehren und Lernen in der Mathematikdidaktik seit vielen Jahren Ernst genommen worden. Im Folgenden werden einige Zugänge, die mögliche Herangehensweise an diese Inhalte aufzeigen, detaillierter diskutiert. Sie wurden teilweise in Schulbüchern umgesetzt bzw. sind in Fachartikeln zu finden. Ich möchte ausgehend von dem, was ich als „Grundform“ oder traditionellen Zugang bezeichnen möchte, einige sich durch „innovative“ Elemente auszeichnende Vorschläge besprechen.

⁴¹ Eichler (2003) identifizierte Formen eines solchen Aufbaus des Curriculums bei Lehrenden und bezeichnet sie als „Universalisten“.

2.1 Traditionelle Zugänge

Seit den 70er Jahren des letzten Jahrhunderts hatte die Wahrscheinlichkeitsrechnung Eingang in deutsche Schulbücher gefunden. Auf Basis der inhaltlichen Festlegung durch einheitliche Prüfungsanforderungen etablierte sich eine „Grundform“ eines an deutschen Gymnasien üblichen Lehrgangs in Wahrscheinlichkeitsrechnung, zumindest für die Sekundarstufe II (vgl. Schupp, 1984), insbesondere für die Einbettung von Themen zur bedingten Wahrscheinlichkeit und der Bayesregel. Diese Grundform bestimmte im Wesentlichen die Vorgehensweise in den meisten Schulbüchern zur Stochastik. Deshalb möchte ich, obwohl aktuellere Schulbücher natürlich in der Ausgestaltung erheblich weiterentwickelt wurden, kurz auf traditionelle Formen eingehen.

Traditionelle Einstiege in die „Welt des Zufalls“ in der *Sekundarstufe I* (z.B. Athen, Griesel & Postel, 1985; Behrens et al., 1986) beinhalten Zufallsexperimente, den Laplaceschen Wahrscheinlichkeitsbegriff, relative Häufigkeit und den frequentistischen Wahrscheinlichkeitsbegriff. Je nach Überzeugung dient die frequentistische Interpretation als empirische Begründung des zentralen Laplace'schen Modells der Gleichwahrscheinlichkeit oder rückt selber ins Zentrum der Betrachtung, wobei die Laplace'sche Wahrscheinlichkeit eine Sonderform darstellt.⁴² Eine wichtige Erweiterung erfolgt zu mehrstufigen Zufallsexperimenten. Spätestens hier wird als nützliches Veranschaulichungsmittel das Baumdiagramm mit Wahrscheinlichkeiten eingeführt.⁴³ Einerseits kann leichter die Ergebnismenge bestimmt werden (z.B. bei mehrfachem Ziehen aus einer Urne), andererseits können die Pfadregeln für die Berechnung von Wahrscheinlichkeiten in mehrstufigen Experimenten anschaulicher hergeleitet werden.

In Schulbüchern für Leistungskurse der *Sekundarstufe II* (z.B. Althoff, 1985; Athen et al., 1984; Barth & Haller, 1983; Feuerpfeil & Heigl, 1997/1975; Glaser et al., 1983) wurde die bedingte Wahrscheinlichkeit meist über Häufigkeitsbetrachtungen definiert und der Satz von Bayes (über Multiplikationssatz und Satz von der totalen Wahrscheinlichkeit) formal hergeleitet.⁴⁴ Als grafische Veranschaulichungen dienten im Allgemeinen Baum- und Mengendiagramme, selten Flächendiagramme („Unit-Square“).⁴⁵ Die Inhalte der Schulbücher für Grundkurse klammerten – im Gegensatz zu Leistungskursbüchern – meist den Satz von Bayes aus, teilweise sogar den Begriff der bedingten Wahrscheinlichkeit.⁴⁶

Um verständlicher darstellen zu können, an welchen Stellen weiter unten besprochene „Innovationen“ ansetzen und warum, wird ein bekanntes Anwendungsbeispiel zur medizinischen Diagnose verwendet.

Typisches Diagnoseproblem

Bei Frauen ab 40 Jahren werden Routineuntersuchungen auf Brustkrebs durchgeführt. Das Untersuchungsverfahren ist die sogenannte „Mammografie“. Aus der Literatur ist folgendes bekannt: Die Wahrscheinlichkeit, dass eine Frau der Altersgruppe zwischen 40 und 50 Jahren Brustkrebs (B) hat, beträgt 0,8%. Die Wahrscheinlichkeit, dass diese Krankheit mit einer Mammografie erkannt wird (M_+), wenn sie vorliegt, beträgt 91,5%. Die Wahrscheinlichkeit, dass eine Mammografie fälschlicherweise auf Brustkrebs hinweist, obwohl die Krankheit gar nicht vorliegt, beträgt 2,6%.
Wie groß ist die Wahrscheinlichkeit, dass eine Frau dieser Altersgruppe tatsächlich Brustkrebs hat, wenn sie einen positiven Mammografiebefund erhalten hat?

⁴² Bei Eichler (2003) werden erstere als „Traditionalisten“ und letztere als „Praktiker“ bezeichnet.

⁴³ Vgl. eine Übersicht von Eulner (1993), der 6 Schulbücher für die Sek.I und 16 für die Sek.II zu Inhalten der Wahrscheinlichkeitsrechnung untersucht hat. In allen 6 Büchern für die Sek.I aus den Jahren 1973 bis 1986 wurden Baumdiagramme verwendet, in 2 auch Mengendiagramme (S.88).

⁴⁴ In 7 der 8 von Eulner (1993, S.89/90) untersuchten Schulbüchern für Leistungskurse wird die Bayesregel formal aufgestellt.

⁴⁵ Baumdiagramme werden in diesem Themenbereich in 6 der 8 untersuchten LK-Bücher verwendet (ebenda).

⁴⁶ In nur einem der 6 untersuchten GK-Bücher wird der Satz von Bayes behandelt, in 3 der 6 bedingte Wahrscheinlichkeit.

Auf dieses Beispiel, das in Kapitel I bereits vorgestellt wurde, werde ich im Folgenden immer wieder zurückgreifen, um die vergleichende Diskussion anhand derselben Problemsituation führen zu können.

Der „übliche“ Lösungsprozess benötigt als Mindestvoraussetzung, dass Schüler mehrstufige Zufallsexperimente modellieren können. Sie lernen dabei meist in der Sekundarstufe I das Baumdiagramm mit Pfadwahrscheinlichkeiten als Darstellungshilfe mehrstufiger Zufallsexperimente kennen. Mit diesen Mitteln könnten Schüler die Diagnoseaufgabe wie folgt lösen.

Zunächst muss erkannt werden, dass man die Situation als ein zweistufiges Zufallsexperiment mithilfe eines Baumdiagrammes darstellen kann. Beim zufälligen „Ziehen“ einer 40 bis 50-jährigen Frau hat man in der ersten Stufe die möglichen Ergebnisse Brustkrebs liegt vor (B) / Brustkrebs liegt nicht vor (nicht B), in der zweiten Stufe jeweils mögliche Untersuchungsergebnisse Mammografie positiv (M+) / Mammografie negativ (M-). Das ergibt etwa folgende Baumdarstellung:

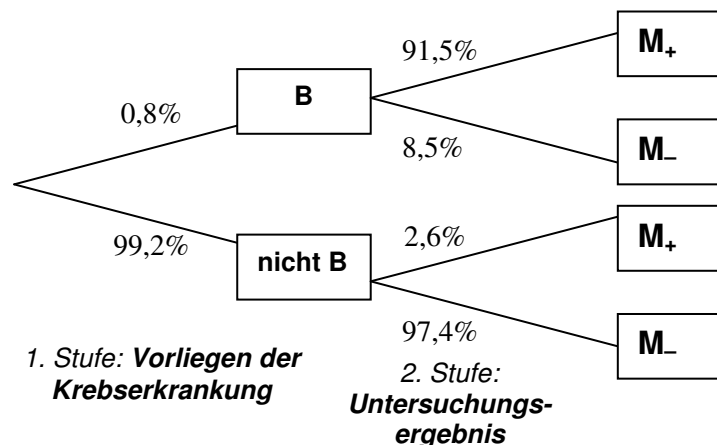


Abb.II.2: Darstellung des Diagnoseproblems im „normalen“ Baumdiagramm.

Ausgerechnet werden können nun die Wahrscheinlichkeiten der 4 vorkommenden Ergebnisse über die Pfadmultiplikationsregel: „Die Wahrscheinlichkeit eines Pfades ergibt sich durch Multiplikation aller Wahrscheinlichkeiten längs des Pfades“. Die „Wahrscheinlichkeit des obersten Pfades“ ist dann z.B. $0,008 \cdot 0,915 = 0,0073 = 0,73\%$. Für eine sachgerechte Interpretation müssen die Lernenden die Beschriftungen mit Wahrscheinlichkeiten Ereignissen zuordnen können. Die erste Wahrscheinlichkeit ist die für Brustkrebs. Die zweite ist die für eine positive Mammografie, wenn Brustkrebs vorliegt. Der Gesamtpfad steht für das Ergebnis „Brustkrebs und positive Mammografie“. Der „Multiplikationssatz“ $P(B \cap M_+) = P(B) \cdot P(M_+ | B)$, der „dahintersteckt“, muss nicht explizit aufgestellt werden. Man kann es zunächst, wie meist in der Sekundarstufe I üblich, bei umgangssprachlichen Formulierungen belassen.

Wie wird die Multiplikationsregel begründet? Die Begründung erfolgt durch Rückgriff auf Häufigkeiten. Hätte man obiges Beispiel zur Einführung verwendet, so wäre etwa zu argumentieren:

„Stelle Dir vor, man wiederholt diesen Untersuchungsprozess 100000mal, dann wäre bei (ungefähr) 0,8% davon, also $100000 \cdot 0,008 = 800$ Frauen Brustkrebs festgestellt worden. Davon hätten (ungefähr) 91,5% eine positive Mammografie erhalten, also ergeben sich $800 \cdot 0,915 = 732$ Frauen mit Brustkrebs und positiver Mammografie. Diese Häufigkeit berechnet sich also als $h(B \cap M_+) = 100000 \cdot 0,008 \cdot 0,915$. Mit der bekannten frequentistischen Definition der Wahrscheinlichkeit gilt

$P(B \cap M_+) = h(B \cap M_+) / 100000$. Nach Kürzen der Grundgesamtheit bleibt die Multiplikation der einzelnen Pfadwahrscheinlichkeiten $0,008 \cdot 0,915$ als Lösung.“

Die Motivation der allgemeinen Regel basiert also wesentlich auf einer Häufigkeitsinterpretation und den Regeln zur Bruchmultiplikation („ein Anteil von einem Anteil von ...“). Die begründete Regel wird von nun an auf Wahrscheinlichkeiten angewendet und eingeübt. Auf Häufigkeiten wird bei Lösungsprozessen aber praktisch nicht mehr Bezug genommen.

Die Hürde, die bei Bayesianischen Problemen entsteht, ist die Berechnung der „inversen“ Wahrscheinlichkeit für Brustkrebs unter der Bedingung einer positiven Mammografie. Bisher sind nur Wahrscheinlichkeiten unter der Bedingung „Brustkrebs“ bzw. „Nicht Brustkrebs“ bekannt. Man benötigt eine Wahrscheinlichkeit für „positive Mammografie“. Hierzu müssen alle Pfade betrachtet werden, die zu positiver Mammografie führen, und die Wahrscheinlichkeiten addiert werden. Im obigen Beispiel: $P(M_+) = 0,008 \cdot 0,915 + 0,992 \cdot 0,026 = 0,033$.

Diese neue Pfadregel (meist als 2. Pfadregel oder Pfadadditionsregel bezeichnet), die formal dem Satz von der totalen Wahrscheinlichkeit entspricht, wird meist nicht mehr durch Häufigkeitsbetrachtungen begründet, sondern mit einem Mengendiagramm und dem Additionsaxiom:

$$P(M_+) = P[(M_+ \cap B) \cup (M_+ \cap \bar{B})] = P(M_+ \cap B) + P(M_+ \cap \bar{B})$$

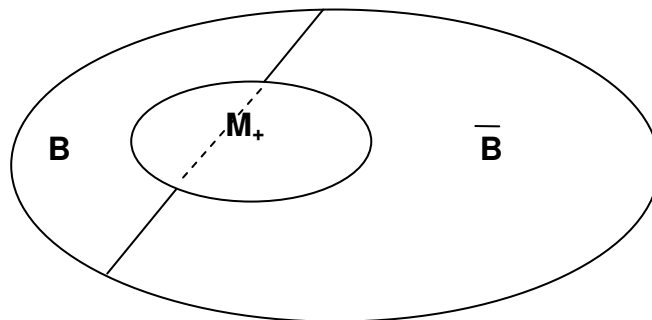


Abb.II.3: Mengendiagramm zur Veranschaulichung der totalen Wahrscheinlichkeit

Da diese mengentheoretische Begründung für die Sekundarstufe I oft als zu schwer und zu abstrakt angesehen wird, wurden in traditionellen Zugängen Bayesianische Probleme meist erst in der Sekundarstufe II behandelt.

Allgemeinere Mängel traditioneller Zugänge zur Stochastik in der Sekundarstufe I werden von Schupp (1984) konkretisiert und so ein „Leitfaden“ für sinnvollen Stochastikunterricht gegeben: Es erfolgen oft zu fachwissenschaftlich orientierte Darstellungen der Stochastik, die insbesondere die Verbindungen zur realen Welt (Modellbildung und Interpretation im Regelkreis des mathematischen Modellierens) ausblenden. Dem Leitziel des sinnvollen Verhaltens und zweckmäßigen Handelns in stochastischen Situationen wird zu wenig Rechnung getragen. Kontexte bleiben oft sehr artifiziell aus dem Bereich der „Münzen-Würfel-Urnen“-Probleme („Würfelbudenmathematik“, ebenda S.236). Auf Primärintuitionen der Schüler ist in höherem Maße Rücksicht zu nehmen. Aktivitäten, wie das Gewinnen von Daten, Durchführen von Experimenten oder Veranschaulichen und Darstellen müssen in allen Stufen ermöglicht werden. Insbesondere sollte der spiralige Aufbau so erfolgen, dass er vorläufige, aber tragfähige Abschlüsse schaffen kann. Zentrale Konzepte sollen so früh wie möglich eingeführt werden, aber auch „vorläufig, phänomenologisch, situationsgebunden,

vereinfacht, ungefähr, um dann immer wieder aufgegriffen, vertieft, präzisiert und verallgemeinert zu werden“ (Schupp, 1984, S.238; vgl. auch von Harten & Steinbring, 1984).

2.2 Innovative Elemente

Bemühungen um realitätsbezogenen Mathematikunterricht waren eine Ursache dafür, dem Satz von Bayes als Themengebiet mit hohem Anwendungsbezug einen „prominenten“ Platz zumindest in der Sekundarstufe II zu verschaffen. Die Rolle des Bayesianischen Denkens für einen allgemeinbildenden Stochastikunterricht und damit auch der Satz von Bayes fanden in der fachdidaktischen Diskussion seit ca. 20 Jahren Beachtung. Der didaktischen Diskussion und innovativen Vorschlägen auf diesem Gebiet wurde auch in einigen Schulbüchern Rechnung getragen. Ich gehe im Folgenden genauer auf einige interessante Ansätze ein, die sich durch besondere Darstellungsmittel, veränderten theoretischen Aufbau bzw. Betonung des Realitätsbezugs auszeichnen.

2.2.1 Darstellungsmittel

a) Vierfeldertafeln und umgekehrte Baumdiagramme

Bei Strick (1998; 1999; 2000) finden sich Vorschläge zur Erweiterung des Darstellungsrepertoires. Dieser Lehrgang von Strick wurde auch in der Schulbuchreihe „Elemente der Mathematik“ von Griesel & Postel (z.B. 1995, 2000) umgesetzt. Neben dem üblichen „normalen“ Baumdiagramm mit Pfadwahrscheinlichkeiten (Abb.II.2) werden Vierfeldertafeln (Tab.II.1a,b) und umgekehrte Baumdiagramme (Abb.II.4) bei Bayes-Aufgaben benutzt. In der Unterrichtspraxis soll in folgenden Schritten vorgegegangen werden, um komplexere Aufgaben bearbeiten zu können (Strick, 1999, S.51):

1. Daten aus dem Aufgabentext (in Wahrscheinlichkeitsformaten) in ein Baumdiagramm übertragen und daraus eine Vierfeldertafel erstellen.
2. Aus einer Vierfeldertafel beide Baumdiagramme („normal“ und „umgekehrt“) erstellen.
3. Wie in 1., anschließend aus der Vierfeldertafel das umgekehrte Baumdiagramm erstellen.

Schritt 3 stellt die intendierte Vorgehensweise dar, wie die Schüler an die Bayes-Aufgaben herangehen sollen. In Strick (2000) wird die Vorgehensweise am Diagnose-Problem zur Mammografie erläutert. Es wird realitätsbezogen vorgegangen. Alle Informationen liegen zunächst als absolute Häufigkeiten in einer Vierfeldertafel vor⁴⁷ (Tab.II.1a). Diese werden in einer Vierfeldertafel in relative Häufigkeiten (Tab.II.1b) übersetzt und gemäß Schritt 2 beide Baumdiagramme mit Wahrscheinlichkeiten erstellt.

		Vorliegen der Krebserkrankung		gesamt
		ja (B)	nein (nicht B)	
Untersuchungs- ergebnis	auffällig (M₊)	118	412	530
	ohne Befund (M₋)	11	15207	15218
gesamt		129	15619	15748

Tab.II.1a: Darstellung des Diagnosebeispiels in einer Vierfeldertafel (absolute Häufigkeiten)

⁴⁷ aus einem tatsächlichen schwedischen Modellversuch, bei dem getestet wurde, wie oft das Mammografie-Screening (Röntgenreihenuntersuchung) zu falschen bzw. richtigen Diagnosen führt. Die Befunde aus der Mammografie wurden mithilfe weiterer Verfahren (Ultraschall etc.) auf Richtigkeit überprüft.

		Vorliegen der Krebserkrankung		gesamt
		ja (B)	nein (nicht B)	
Untersuchungs- ergebnis	auffällig (M ₊)	0,75%	2,62%	3,37%
	ohne Befund (M ₋)	0,07%	96,56%	96,63%
gesamt		0,82%	99,18%	100%

Tab.II.1b: Darstellung des Diagnosebeispiels in einer Vierfeldertafel (mit relativen Häufigkeiten)

Das Baumdiagramm mit Wahrscheinlichkeiten in der „normalen“ Anordnung (Abb.II.2) korrespondiert mit diesen Daten (Pfadwahrscheinlichkeiten gerundet). Im umgekehrten Baumdiagramm werden die Daten so visualisiert (Abb.II.4):

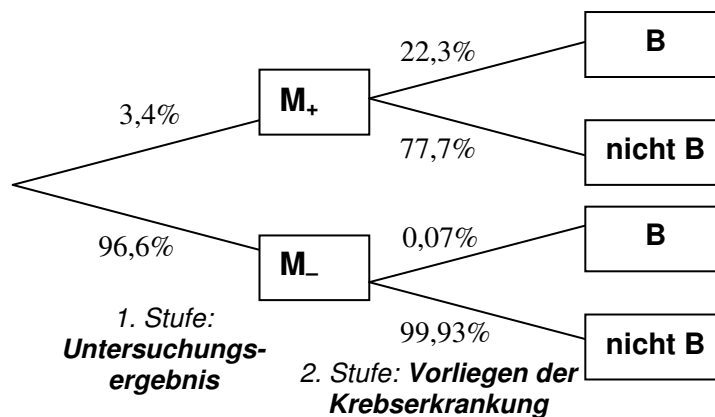


Abb.II.4: Darstellung des Diagnosebeispiels im umgekehrten Baumdiagramm (Strick, 2000, S.248).

Unzweifelhaft eignen sich gerade umgekehrte Baumdiagramme besonders gut, um die Wirkungsweise der Bayesformel, nämlich eine Umkehrung der Sichtweise, zu demonstrieren: Was gegebenermaßen die 1. Stufe des Baumes bildete, ist nun die 2. Stufe und umgekehrt. Insbesondere ist nun Bedingung, was vorher bedingtes Ereignis war (im Bsp: M₊ bzw. M₋).⁴⁸

Die Lösungswahrscheinlichkeit für ein Vorliegen der Krebserkrankung, wenn das Ergebnis positiv war, $P(B | M_+) = 22,3\%$ muss als Quotient von $P(B \text{ und } M_+) = 0,75\%$ durch $P(M_+) = 3,37\%$ aus der Vierfeldertafel berechnet werden. Strick kommentiert hier: „Dass beim Übergang von der Vierfeldertafel zum 2. Baumdiagramm Quotienten gebildet werden müssen, erscheint den Schülerinnen und Schülern *natürlich*, wenn sie vorher genügend Beispiele mit absoluten Häufigkeiten gerechnet haben“ (1999, S.51; Hervorhebung im Original). Es ist allerdings nicht grundsätzlich davon auszugehen, dass den Schülern verständlich ist, warum auch mit relativen Häufigkeiten Quotienten zu bilden sind. Selbst wenn dem so ist, liegt die Frage nahe, was mit der Übersetzung in relative Häufigkeiten bezweckt wird. Dieser „Umweg“ stellt meines Erachtens eine Verkomplizierung dar, die für ein verständiges Lösen nicht nötig wäre (siehe Kap.I u.II.3).

Der Lehrgang von Strick zeichnet sich neben dieser speziellen Darstellungsweise durch anwendungsbezogenes Vorgehen und geringeren Formalisierungsgrad als traditionelle Zugänge aus. Er beinhaltet auch die Anwendung des Satz von Bayes zur Abschätzung der Chancen für die Richtigkeit von Hypothesen (z.B. in Griesel & Postel, 2000, S.385f.), die zentrales Element der Lehrgänge in Abschnitt 2.2.2 ist.

⁴⁸ Man beachte: Der Begriff „umgekehrtes Baumdiagramm“ wird in der Literatur nicht einheitlich verwendet. Ich verwende im Sinne von Strick im folgenden den Begriff „umgekehrtes Baumdiagramm“ nur, wenn eine Umkehrung der Stufen enthalten ist.

b) Flächendiagramme („Unit-Square“)

Bei Flächendiagrammen oder „Unit-Squares“ werden Wahrscheinlichkeiten als proportionale Flächen in einem Rechteck visualisiert. Aus Trainingsvergleichen von Bea (1995) ging z.B. das Einheitsquadrat als „erfolgversprechendste“ Darstellungsform bei Problemen mit bedingter Wahrscheinlichkeit und Basisratenproblemen hervor (genauere Beschreibung der Trainingsstudie in Kap.III.1.3). Didaktisch wird gefolgert, dass es als generelles Darstellungsmittel für diesen Themenbereich völlig ausreicht: „Angesichts dieser Tatsache [teilweise empirische Überlegenheit des Einheitsquadrates, Anm. des Verf.] erscheint es nicht notwendig, *situative* Empfehlungen über den Einsatz der Graphiken abzugeben. [...] Vielmehr bietet es sich an, das Einheitsquadrat durchgängig zu verwenden, um nicht Gefahr zu laufen, durch mehrere verschiedenartige Modelle zusätzliche Verwirrung bei den Lernenden zu stiften“ (Bea, 1995, S.195; Hervorhebung im Original). Flächendiagramme finden sich in Schulbüchern jedoch bisher nur ergänzend (z.B. Feuerpfeil & Heigl, 1997).⁴⁹

Das Diagnosebeispiel kann etwa wie folgt mit einem Einheitsquadrat visualisiert werden (Abb.II.5). Das Einheitsquadrat mit Inhalt 1 repräsentiert einen Wahrscheinlichkeitsraum Ω . Die vertikale Aufteilung entspricht der Wahrscheinlichkeitsverteilung des dichotomen Merkmals „Brustkrebs“ (0,008 bzw. 0,992). Die horizontale Aufteilung bezieht sich auf die jeweiligen Anteile bei Brustkrebs bzw. nicht Brustkrebs mit positiver bzw. negativer Mammografie, also die bedingten Wahrscheinlichkeiten.

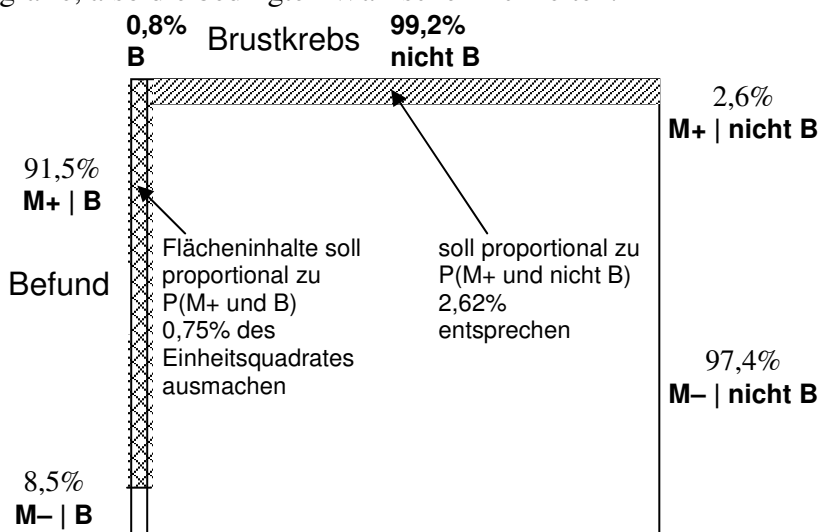


Abb.II.5: Darstellung des Diagnosebeispiels im Einheitsquadrat (vgl. Bea, 1995, S.154)

Gemäß einer Flächeninterpretation des Satzes von Bayes lautet die Lösungsvorschrift:

$$P(B | M_+) = \text{Inhalt der Fläche „\\\\\\\\“} / \text{Inhalt der Flächen „/////“}$$

$$= 0,75\% / (0,75\% + 2,62\%) = 22,3\%$$

Das „Unit-Square“ ermöglicht durch die Proportionalität von Flächen und Wahrscheinlichkeit eine Veranschaulichung der Produktregel: „Inhalte der Flächen“ entsprechen den konjunktiven Wahrscheinlichkeiten und sind das Produkt der jeweiligen Seitenlängen.

⁴⁹ In der Schulbuchuntersuchung von Eulner (1993, S.89/90) für die Sekundarstufe II wird in 4 von 16 Büchern ein Flächendiagramm verwendet.

Also etwa:

$$\text{Flächeninhalt (B und } M_+) = \text{Seitenlänge B} \cdot \text{Seitenlänge (} M_+ | \text{B)} = 0,008 \cdot 0,918 = 0,0075.$$

Auch der Satz von der totalen Wahrscheinlichkeit kann wie in einer Mengendarstellung bei Flächendiagrammen einfach begründet werden:

$$\text{Flächeninhalt } M_+ = \text{Flächeninhalt (B und } M_+) + \text{Flächeninhalt (nicht B und } M_+)$$

Das Einheitsquadrat ist zweifellos zur Einführung besonders geeignet, da es eine visualisierte Begründung der formalen Bayesregel ermöglicht. Es hilft wohl Fehlintonationen, wie z.B. die Vernachlässigung der Basisrate zu vermeiden, da Größenverhältnisse von Wahrscheinlichkeiten mit abgebildet sind. Allerdings werden Flächendarstellungen bei komplizierteren Daten oder Strukturen schnell unübersichtlich und scheinen daher zur aktiven grafischen Modellierung des Lösungsprozesses durch den Lernenden weitgehend ungeeignet.⁵⁰ Eigene Befunde (vgl. Kap.I.4.2.2) sprachen eher gegen eine generelle Favorisierung von Flächendarstellungen.

c) Verbindung von Darstellungsmitteln: Baum-Mengen-Diagramme

Bei Tomlinson und Quinn (1997) wird versucht Mengendarstellungen (sogenannte Venn-Diagramme) und Baumdiagramme zu einem einzigen Problemlösewerkzeug zu verschmelzen. Während als Vorteil der Mengendarstellung die Veranschaulichung von konjunktiven Ereignissen gesehen wird, gibt es auch wesentlichen Nachteile: Es können nur unklar zugrundeliegende Beziehungen der stochastischen (Un-)Abhängigkeit dargestellt werden, die aber wesentlich für die Lösung von Aufgaben mit bedingten Wahrscheinlichkeiten sind. Diese stochastischen Beziehungen, so unterstreichen auch Tomlinson und Quinn (ebenda, S.5), werden am besten mit Baumdiagrammen repräsentiert.

Um beide Vorteile auszunutzen, werden Venn- und Baumdiagramme kombiniert verwendet (vgl. Abb.II.6). Die weitere Besonderheit ist eine *Verschmelzung von normalem und umgekehrtem Baumdiagramm*. Da man sowohl in der normalen als auch in der umgekehrten Sichtweise (entspricht normalem und umgekehrtem Baumdiagramm) zu denselben konjunktiven Wahrscheinlichkeiten gelangt (vgl. Produktsatz), ist dies leicht möglich. Die *kombinierte Baum-Mengen-Darstellung* kann aber nur bei einfachen Beispielen (z.B. einfacher Würfelwurf) eingesetzt werden. Sie ist wie die Flächendarstellung als aktives Modellierungsinstrument anwendungsorientierter Probleme schnell überfordert. Es wird auch vorgeschlagen, Mengendiagramme wieder wegzulassen und durch Wahrscheinlichkeiten zu ersetzen, wenn die Lernenden mit der Darstellung vertraut sind.

Meines Erachtens ist v.a. die Verbindung zu einem „vollständigen Baumdiagramm“ didaktisch wertvoll, worauf bereits eigene Befunde in Kap.I.4.2.2 hindeuteten. Tomlinson und Quinn geben an, dass ihre Schüler eine abstrakte Form der Bayesregel selbstständig aus dem vollständigen Baumdiagramm ableiten konnten. Bei aller Unklarheit über ihre empirische Vorgehensweise diesbezüglich, wäre dieser Befund natürlich ein deutliches Indiz für den Vorteil der Integration der umgekehrten Sichtweise in eine einzige Darstellung, gerade bei Problemen zur Regel von Bayes. In meinen eigenen Untersuchungen wurden auch vollständige Baumdiagramme verwendet, allerdings in Verbindung mit Häufigkeiten (vgl. auch Kap.II.3.2).

⁵⁰ Die Ergebnisse der Trainingsstudie von Bea (1995) bieten zur aktiven Verwendung keine Befunde (vgl. Kap.III.1.3).

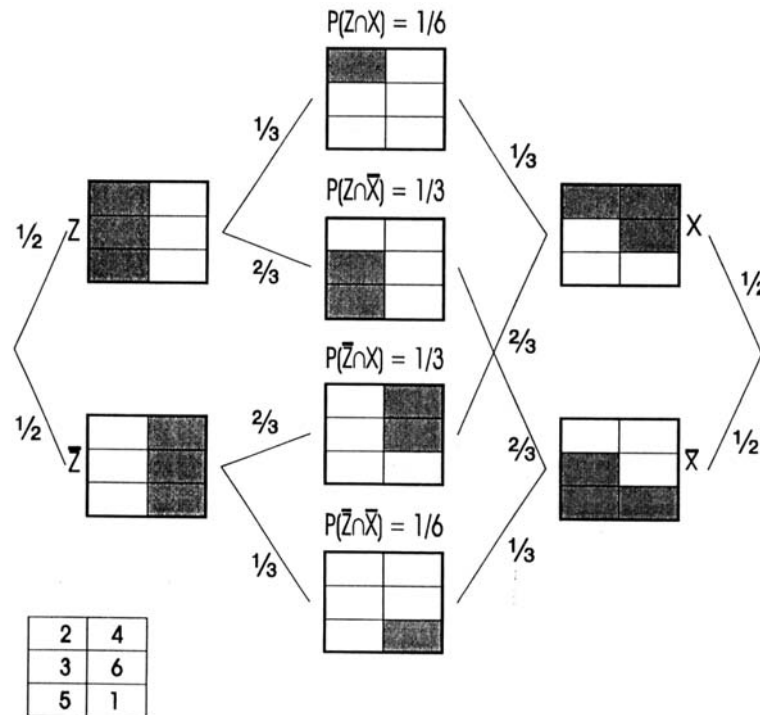


Abb.II.6: Darstellung eines Würfelwurfexperiments im Baumdiagramm mit integrierter Mengendarstellung (X: „Die geworfene Zahl ist gerade“; Z: „Die geworfene Zahl ist eine Primzahl“)

2.2.2 Bayesianischer Ansatz

Riemer (vgl. 1985; 1986) benennt zwei Hauptversäumnisse des üblichen Zugangs zur Stochastik:

- Wahrscheinlichkeiten werden zunächst wie üblich in Mathematik lediglich exakt berechnet, wodurch der hypothetische Charakter der Wahrscheinlichkeitsrechnung kaum deutlich wird.
- Der Begriff der subjektiven Wahrscheinlichkeit für eine bestimmte Hypothese wird nicht entwickelt. Die klassische Theorie kann solche Wahrscheinlichkeiten nicht liefern.

Gerade in Letzterem sieht Buth (1991) sogar eine „Behinderung des gesunden Menschenverstandes“. Intuitiv verspricht sich der Schüler z.B. beim Testen von Hypothesen eine Aussage über die Wahrscheinlichkeit des Zutreffens der Hypothese (bei vorliegendem Ausgang des Experiments) treffen zu können. Eher belanglos wirkt da die Wahrscheinlichkeit, mit der man einen Irrtum riskiert, wenn man sich gegen eine Hypothese ausspricht, die in Wirklichkeit stimmt (ebenda, S.21). Dieser Unterschied wird als erklärungsbedürftig im Unterricht angesehen.

Die Logik der Signifikanzmethode ist alles andere als naheliegend. Das wurde auch in Studien mit Studenten und sogar Statistiklehrenden eindrucksvoll bestätigt. Die Bedeutung eines signifikanten Ergebnisses wird oft falsch interpretiert (z.B. Oakes, 1986; Krauss & Wassner, 2001). Selbst in Lehrbüchern zur Statistik finden sich solche Fehlinterpretationen (Gigerenzer, 1993a). Der gesunde Menschenverstand scheint dazu zu tendieren, irgendeine Aussage über die Wahrscheinlichkeit der Hypothesen treffen zu wollen, auch wenn sie durch Signifikanztesten nicht möglich ist. Cohen (1994, S.997) drückte es so aus: „What’s wrong with significance testing? Well, among many other things, it does not tell us what we want to

know, and we so much want to know what we want to know that, out of desperation, we nevertheless believe in that it does!" Diese Beobachtung machen auch Lehrende oft bei ihren Schülern, etwa wenn gefolgert wird: „Wir haben die Hypothese auf dem 5% Signifikanzniveau verworfen, also gilt die Alternative mit 95%iger Sicherheit“ (z.B. Birnbaum, 1982).

Riemer (z.B. 1986, S.4) propagierte deshalb die Entwicklung des Bayesianischen Zugangs zum Testen „auf unteren Curriculumstufen (Klasse 8)“, indem bereits hier von (subjektiven) *Wahrscheinlichkeiten* für das Zutreffen von *Hypothesen* gesprochen wird. Zur Vermeidung resultierender Fehlvorstellungen kann die iterierte Anwendung der Bayesschen Regel dienen, die es – im Gegensatz zum klassischen Hypothesentesten – erlaubt, Wahrscheinlichkeiten von Hypothesen gegeneinander abzuwägen.⁵¹ Das Entscheiden zwischen Hypothesen mit Bayes wird für geeignet für niedrigere Stufen gehalten, da es „mit der Pfadregel auskommt, insbesondere keine Testgrößen benötigt“ (ebenda, S.9). Eine Umsetzung für die Sekundarstufe I erfolgte z.B. im Schulbuch von Schmid & Weidig (1996, S.180f.). Auch in den neueren Schulbüchern von Cukrowicz & Zimmermann (2001) und Lergenmüller & Schmidt (2003) wird der Bayesianische Zugang als „*Lernen aus Erfahrung*“ zentral entwickelt.

Die Argumentation in Riemers' Zugang wäre auf das Diagnosebeispiel angewendet die folgende. Eine Untersuchung soll helfen zwischen zwei alternativen Hypothesen abzuwägen, „Brustkrebs“ oder „kein Brustkrebs“. Als Indiz zählt ein positives Mammografieergebnis. Zur grafischen Darstellung wird ein sogenanntes „Bayes-Diagramm“⁵² verwendet, in dem nur die Änderung der A-priori-Wahrscheinlichkeiten durch das tatsächlich beobachtete Indiz betrachtet wird (vgl. Abb.II.7).

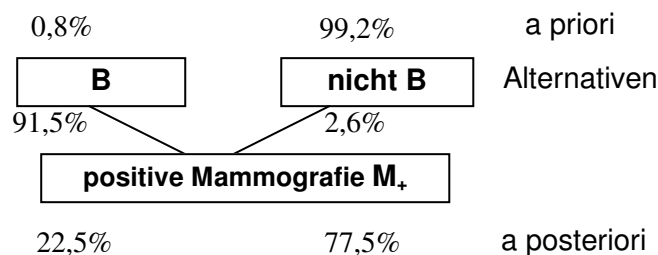


Abb.II.7: Darstellung des Diagnosebeispiels im „Bayes-Diagramm“ (vgl. Riemer, 1986, S.5)

Aus dem Bayes-Diagramm wird wie folgt die Regel von Bayes abgeleitet:

„Man erhält die *a posteriori* Wahrscheinlichkeit einer Alternative, indem man die Wahrscheinlichkeit des zugehörigen Pfades durch die totale Wahrscheinlichkeit des beobachteten Indizes teilt“ (Schmid & Weidig, 1996, S.184).

Die (totale) Wahrscheinlichkeit, das Indiz zu beobachten ist $0,8\% \cdot 91,5\% + 99,2\% \cdot 2,6\%$, denn die Pfade zum Indiz „positive Mammografie“ haben die Wahrscheinlichkeiten:

$$\text{Pfad über „Brustkrebs“: } 0,008 \cdot 0,915 = 0,0075$$

$$\text{Pfad über „nicht Brustkrebs“: } 0,992 \cdot 0,026 = 0,0258$$

Die nach Beobachtung des Indizes „M₊“ revidierte Wahrscheinlichkeit für Brustkrebs ist also:

⁵¹ Das hielt Lindley (z.B. 1965) für das schlagendste Argument für den Bayesianischen Zugang.

⁵² Eine ähnliche Darstellung findet sich auch bei Wickmann (1990), die dort als Pfaddiagramm bezeichnet wird.

$$P(B | M_+) = \frac{0,0075}{0,0075 + 0,0258} = 22,5\% \text{ }^{53}$$

Vorteilhaft kann das „Bayes-Diagramm“ der Anschauung zugute kommen, da es nur die Daten beinhaltet, die für eine konkrete Fragestellung gebraucht werden und damit deutlich weniger komplex ist als „komplette“ Baumdiagramme. Deswegen eignet es sich auch gut zur Darstellung iterierter Anwendung der Bayesschen Regel, indem man einfach mehrere Bayes-Diagramme „hintereinanderschaltet“ (vgl. ebenda, S.185). Das Diagramm unterstützt die Herleitung der Bayesregel und schafft die Basis, Aufgaben schematisch mit einfachen Mitteln lösen zu können, was eine wesentliche Intention dieses Lehrgangs zu sein scheint. Ob durch diese Darstellung aber die Herleitung der Regeln tiefer begründet oder leichter begreifbar wird, ist fraglich. Im nächsten Abschnitt werden hierfür meines Erachtens geeignete Zugänge gezeigt.

2.2.3 Häufigkeitsbasierte Regelherleitung

Die Bayesianische Denkweise des „Lernen aus Erfahrung“ ist auch im Schulbuch von Kuypers, Lauter & Wuttke (1999) vorherrschend. Im Unterschied zur traditionellen Herleitung der Bayesregel wird hier häufigkeitsbasierter vorgegangen. Ein medizinisches Diagnoseproblem mit fiktiven Daten dient als Einführungsbeispiel. Es erfolgt eine *Häufigkeitsdarstellung im Baumdiagramm* ausgehend von 1000000 Personen, ähnlich der „Häufigkeitsbäume“ bei Gigerenzer (Abb.II.8; vgl. auch Kap.I, Abb.I.2).

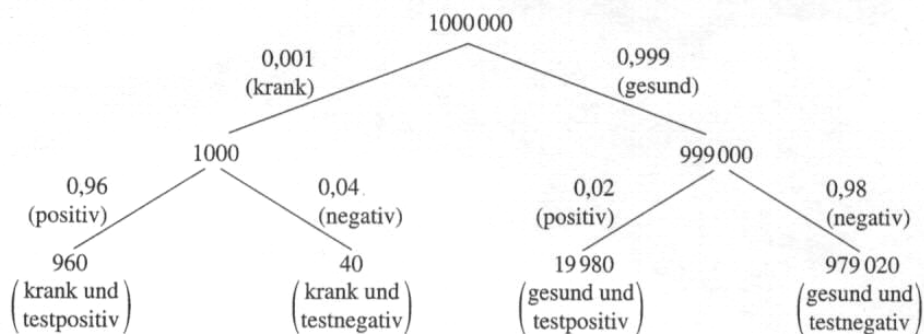


Abb.II.8: Diagnosebeispiel in einer Baumdarstellung mit Häufigkeiten (Kuypers et al., 1999, S. 175)

Die bedingte Wahrscheinlichkeit für die Hypothese „Krankheit“ unter der Bedingung des Datums „positives Testergebnis“ wird als Verhältnis von Häufigkeiten berechnet. Diese Überlegung wird nun genutzt, um die Bayesregel mit Wahrscheinlichkeiten formal zu begründen (vgl. Abb.II.9). Die Multiplikationsregel wurde vorher ebenfalls durch Häufigkeitsüberlegungen mit Hilfe des Baumdiagramms abgeleitet. Es ist anzunehmen, dass die Begründung der Regel von Bayes über Häufigkeitsüberlegungen bei Lernenden eine anschaulichere Vorstellung der Regel bewirken kann. In folgenden Übungen werden weitere Beispiele nur noch mit hergeleiteten Wahrscheinlichkeitsregeln gelöst. Auf absolute Häufigkeiten wird kaum mehr Bezug genommen. Aus den in Kapitel I dargestellten kognitiven Überlegungen und Befunden macht das Einführungskonzept bei Kuypers et al. (1999) einen Schritt in Richtung der Betonung häufigkeitsbasierter Intuitionen. Kernaspekte dieser Arbeit befassen sich mit einer noch wesentlich zentraleren Verwendung von Häufigkeitsmodellierung (vgl. Kap.II.3 und IV).

⁵³ 22,5% (anstatt 22,3% direkt aus den Häufigkeiten 118 / 530) ergibt sich aufgrund von Rundungen aller relativen Häufigkeiten (Rundungen wie in obigen Baumdiagrammen).

Wir haben diese Wahrscheinlichkeit oben dadurch berechnet, dass wir die Zahl der Fälle mit (H₁ und D) durch die Zahl **aller** Fälle mit D, also mit (H₁ und D) und mit (H₂ und D), dividiert haben:

$$P(H_1|D) = \frac{H(H_1 \text{ und } D)}{H(H_1 \text{ und } D) + H(H_2 \text{ und } D)} = \frac{960}{960 + 19980}$$

Stattdessen können wir natürlich auch die zugehörigen Wahrscheinlichkeiten durcheinander dividieren:

$$P(H_1|D) = \frac{P(H_1 \text{ und } D)}{P(H_1 \text{ und } D) + P(H_2 \text{ und } D)}$$

Mit den Gleichungen von Seite 177 ergibt sich daraus:

$$P(H_1|D) = \frac{P(H_1) \cdot P(D|H_1)}{P(H_1) \cdot P(D|H_1) + P(H_2) \cdot P(D|H_2)}$$

Abb.II.9: Häufigkeitsbasierte Herleitung der Regel von Bayes (Kuypers et al., 1999, S. 178)

2.2.4 Chancenkonzept und Konzept des „Begünstigens“

Die Formel von Bayes wird von Borovcnik als zu „schwerfällig“ für das Erkennen des Zusammenhanges der A-priori- und A-posteriori-Wahrscheinlichkeiten charakterisiert (1992, S.177). Zur Überwindung dieser „Schwerfälligkeit“ wird vorgeschlagen, von der intuitiven Interpretation für Wahrscheinlichkeit auszugehen, nämlich der Idee des Wertes einer Wette bzw. von *Chancenverhältnissen* (im Sinne des „Wert eines Spiels“ bei Huygens, 1657)⁵⁴. Konsistent wird auch die Bayesformel mit Chancenverhältnissen⁵⁵ so dargestellt, dass sie an die Tätigkeit des Neubewertens von Wahrscheinlichkeit direkter anbindet (Borovcnik 1992, S.230f.; 1990, S.32f.). Als direkte Vorteile eines solchen Ansatzes werden die Vereinfachung der Berechnung und der Strukturen und durch die Betonung des Denkens in Informationen eine Unterstützung von Intuition und Einsicht in die Verwendungsbereiche des Kalküls hervorgehoben.

Eine neue Information I verändert das Chancenverhältnis $W(E) : W(nE)$ zu $W(E \wedge I) : W(nE \wedge I)$. Mit der Produktregel ergibt sich:

$$W(E) \cdot W(I | E) : W(nE) \cdot W(I | nE) \Leftrightarrow W(E) : W(nE) \cdot W(I | E) : W(I | nE)$$

Man kann diese Beziehung nun so interpretieren:

$W(E \wedge I) : W(nE \wedge I)$	= $W(E) : W(nE)$	· $W(I E) : W(I nE)$
neues Chancenverhältnis (posteriori odds)	= altes Chancenverhältnis (priori odds)	· Trennkraft (Likelihood-Quotient)

Im Chancenkonzept von Borovcnik wird ebenfalls eine „verkürzte“ Version eines Baumdiagramms verwendet (vgl. Abb.II.10). Die Darstellung ist strukturell analog zu Riemer’s Bayes-Diagramm (vgl. Abb.II.7). Der Hauptunterschied liegt in der Angabe der A-

⁵⁴ Die Frage „Was ist der faire Preis für ein bestimmtes Spiel?“ wird auch als die Geburtsstunde der mathematischen Betrachtung von Wahrscheinlichkeit gesehen, als ein gewisser Chevalier de Méré sich mit dieser Frage über bestimmte Würfelspiele 1654 an die Mathematiker Pascal und Fermat wandte.

⁵⁵ *Definition:* Für ein Ereignis E, dessen Wahrscheinlichkeit W(E) als bekannt vorausgesetzt wird, wird der Quotient $W(E) : W(nE)$ als Chancenverhältnis von E gegen nE bezeichnet („odds-ratio“).

priori-Wahrscheinlichkeit als Chancenverhältnis. Borovcnik (1984, S.218) spricht auch von „inversen“ Baumdiagrammen, obwohl die Sichtweise der eines „normalen“ Baumdiagramms entspricht.

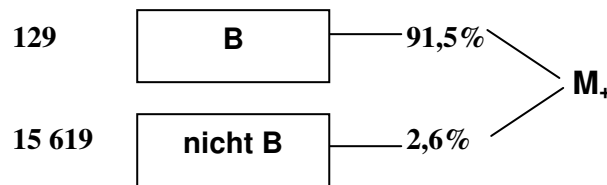


Abb.II.10: Darstellung des Diagnosebeispiels im „inversen“ Baumdiagramm (vgl. Borovcnik 1992, S.229)

Die Rechenregel für das a posteriori-Chancenverhältnis lautet:

$$A\text{-posteriori-Chancenverhältnis} = A\text{-priori-Chancenverhältnis} \cdot \text{Likelihood-Quotient}$$

Im Beispiel:

$$\begin{aligned} \text{Chancen von } (B \mid M_+) \text{ gegen } (\text{nicht } B \mid M_+) &= \\ &= (129 \cdot 0,915) : (15619 \cdot 0,026) \approx 118 : 406 \end{aligned}$$

Daraus folgt per Def. die Wahrscheinlichkeit:

$$P(B \mid M_+) = 118 / (118 + 406) = 22,5\%^{56}$$

Das sogenannte „*Begünstigen*“-Konzept als didaktische Idee wurde u.a. von Borovcnik (1992, S.198f.; vgl. auch Borovcnik, 1985, 1990; Scholz, 1981; Falk & Bar-Hillel, 1983) verwendet, um die oft vorhandene Lücke zwischen Theorie und Praxis zu schließen. Die Neubewertung von Wissen als Anwendung der Bayesregel wird durch ein Werkzeug ergänzt, um die *qualitative* Veränderung einer bedingten Wahrscheinlichkeit formal zu beschreiben und Verwechslungen dieser „Bayesianischen Inferenz“ mit kausalem oder logischem Schlussfolgern zu vermeiden. Im Paradebeispiel der medizinischen Diagnose entspricht der Intuition: Wer krank ist, „verursacht“ einen positiven Befund. Stochastisch haben jedoch diese tief verwurzelten, kausal-logischen Intuitionen keine Bedeutung. Es macht ebenso Sinn die Wahrscheinlichkeit von Befunden (z.B. positive Testergebnisse) neu zu bewerten mit Informationen über das Vorliegen der entsprechenden Krankheit. Das Hauptproblem für Lernende ist demnach also die Wirkungsweise der Formel von Bayes zu verstehen, die auch oft mit dem Schlagwort „Lernen aus Erfahrung“ charakterisiert wird. Die „Begünstigen“-Idee soll diesen Mechanismus besonders einleuchtend machen. Der Kern des Konzeptes ist die systematische Untersuchung des an sich trivialen Zusammenhangs, dass sich die Wahrscheinlichkeit durch das Einbeziehen einer Bedingung bzw. neuer Information im allgemeinen ändert (wenn nicht, ist die Information irrelevant). Im Zusammenhang mit Wetten bedeutet die Beziehung „Begünstigen“ für zwei Ereignisse E und F etwa eine Erhöhung des Chancenverhältnisses von F durch Kenntnis von E (formal: $E \nearrow F$).

Didaktische Möglichkeiten dieser Konzepte bestehen im Wesentlichen daraus, dass der Zusammenhang zwischen Wahrscheinlichkeit, Chancenverhältnissen und Gewinnquoten einerseits auf einer sehr intuitiven Ebene ausgebaut oder andererseits mathematisch gerechtfertigt werden kann (Borovcnik, 1992, S.182). Es gibt aber keine empirischen Befunde, ob die Chancensichtweise für Schüler leichter oder intuitiver als die Vorstellung über relative Häufigkeiten ist.

⁵⁶ 22,5% (anstatt 22,3% direkt aus der relativen Häufigkeit) ergibt sich aufgrund der Rundung der Likelihoods (entsprechend zu den Baumdiagrammen oben).

Während in der üblichen Theorie z.B. Unabhängigkeit als statistische Eigenschaft des Objektes betrachtet wird, wird der Begriff durch das „Begünstigen“-Konzept mit einer Tätigkeit in Verbindung gebracht, eine Wahrscheinlichkeitsbewertung unter Einbeziehung neuer Information zu verändern oder nicht (Borovcnik, 1992, S.218). Tief verwurzelte fehlerhafte Intuitionen in diesem Bereich, die unzweifelhaft vorhanden sind und oft durch Vermischung mit kausalem oder logischem Denken erzeugt werden, können durch das Bild der Neubewertung von Wahrscheinlichkeiten besser überwunden werden, im Sinne einer „sinnstiftenden Handlung“. Die didaktische Chance des Konzeptes besteht somit in der Möglichkeit, das „Spannungsfeld von Begriff und Tätigkeit“ besser zu überwinden. Spezifische konzeptuelle Eigenschaften des stochastischen Denkens können dadurch leichter ausgebildet werden (ebenda, S.220).

2.2.5 Betonung des Realitätsbezugs

Eine intensive Diskussion in der Mathematikdidaktik wurde seit Jahrzehnten über Fragen des Realitäts- und Lebensbezugs der im Unterricht zu behandelnden Themen geführt (für eine Übersicht Blum, 1996). Argumente für die Verstärkung von Realitätsbezügen im Mathematikunterricht sind v.a. pragmatisch die Erziehung der Schüler zu *mündigen Bürgern*, formal die Förderung der Problemlösekompetenz durch das *Übersetzen zwischen Realität und Mathematik*, kulturbezogen die Schaffung eines *reflektierten und ausgewogenen Bildes von Mathematik* als „gesellschaftliches Gesamtphänomen“ und lernpsychologisch als Motivation und Unterstützung des Verstehens und Behaltens mathematischer Inhalte (vgl. Blum, 1991; 1996).

Die Relevanz realitätsbezogenen Modellierens wurde als Ziel insbesondere eines *allgemeinbildenden Stochastikunterrichts* von Schupp (z.B. 1984; im Druck) immer wieder betont und genauer dargelegt: „Die Phasen des [...] Interpretierens und Rückkoppelns, also die beidseitigen Übergänge zwischen Mathematik und sonstiger Welt, stellen erhebliche Anforderungen und stehen immer in der Gefahr des Verkürztwerdens. So sind wir heute mehr denn je im Unterricht gehalten, unsere stochastischen Vorgehensweisen und Resultate sowie diejenigen anderer kritisch zu betrachten, sie eher als Entscheidungshilfen denn als Entscheidungen anzusehen“ (Schupp, im Druck). Schupp hebt hervor, dass der Stochastikunterricht *über* den Gesamtbeitrag zum Mathematikunterricht hinaus „spezifische, nicht auswechselbare Gelegenheiten bietet, auf welche die allgemeinbildende Schule nicht verzichten kann“.

Ein Hauptversäumnis des Mathematikunterrichts, insbesondere in Stochastik, wird oft darin gesehen, die Lernenden von der Relevanz der Inhalte für ihr weiteres Leben zu überzeugen.⁵⁷ Heymann (1995, S.20) führte diesen Aspekt der Lebensvorbereitung genauer aus: Einerseits braucht ein Schüler, der keinen mathematischen oder mathematiknahen Beruf ergreift, für sein späteres Leben nur sehr wenige Inhalte des derzeitigen Mathematikunterrichts. Damit steigt die Schwierigkeit, Schüler für das zu motivieren, was sie im Mathematikunterricht lernen sollen. Andererseits wird der Aspekt der Lebensvorbereitung auch unterschätzt, da gerade solche mathematische Aktivitäten, die im Alltag wichtig sind, systematisch vernachlässigt werden. Zu einseitig wird auf das korrekte Lösen eindeutig lösbarer Aufgaben Wert gelegt, was meist auf ein Abarbeiten von Algorithmen hinausläuft.

Der Stochastikunterricht und insbesondere das Problemfeld Bayesianwendungen eröffnet Möglichkeiten für genau solche vernachlässigten Aktivitäten, die im Alltag jedes

⁵⁷ Z.B. für Stochastik trat dies bei Befragungen von Schülern von Rathgeber (1999) deutlich zu Tage.

Bürgers eine mehr oder weniger große Rolle spielen (vgl. Heymann, 1995, S.21 bzw. Schupp, im Druck) wie z.B.

- Schätzen, Überschlagen, intuitives Erfassen von Größenordnungen,
- Übersetzen von Sachproblemen in einfache mathematische Modelle,
- Interpretation und Erstellung von grafischen Darstellungen,
- Mathematik als Kommunikations- und Argumentationsmittel,
- Umgang mit statistischen Daten und Interpretation von Wahrscheinlichkeitsaussagen in realen Zusammenhängen,
- Verständiger Einsatz technischer Hilfsmittel (z.B. Computer)

Ein unübersehbarer Trend hin zu höherem Realitätsbezug ist sicher in der heutigen Alltagspraxis des Mathematikunterrichts zu beobachten. Dennoch kann man weiterhin feststellen, dass zwischen mathematikdidaktischen Vorschlägen und unterrichtlicher Praxis Lücken klaffen und oft die Kalkülorientierung immer noch die dominierende Rolle zu spielen scheint. Was führt zum Widerstand von Mathematiklehrenden, möglichst realitätsbezogenen Unterricht zu machen? Der didaktische Weg ist anspruchsvoller in organisatorischer, schüler- und lehrerbezogener Hinsicht als der übliche und kann auf besondere Hindernisse stoßen (vgl. Blum, 1996, S.28). Die Hauptangst besteht wohl darin, dass durch eine „Öffnung“ die Unbestimmtheit des Modellierungsprozesses ansteigt und eventuelle Ersatzstrategien der Lernenden, die mit dem eigentlichen Problem nichts mehr zu tun haben, eine vernünftige Kommunikation erschweren.

Ich werde aus Platzgründen in dieser Arbeit die umfassende Diskussion in der Mathematikdidaktik über verschiedene Formen und Methoden realitätsbezogenen Mathematikunterrichts, Aspekte von Lebensrelevanz insbesondere Schülerrelevanz bzw. Authentizität von Beispielen nicht weiter ausführen, sondern einen konkreten Vorschlag zu Anwendungen des Satz von Bayes aufzeigen. In diesem Unterrichtsvorschlag (vgl. Kap.VI) sind realitätsbezogene und schülerrelevante Elemente zentral.

Die hohe praktische Relevanz des Bayesianischen Denkens wurde bereits einführend herausgestellt. Warum sonst wären derartige Problemstellungen unter den bekanntesten und meist untersuchtesten in der psychologischen Urteils- und Entscheidungsforschung, deren Untersuchungsgebiet vor allem Alltagsprobleme der Menschen mit Unsicherheit und Zufall sind?

Es gibt z.B. einen ausgearbeiteten Vorschlag für einen realitätsbezogenen Zugang von Böer (1997) aus der Material-Reihe „ISTRON“ (Blum et al., 1997). Anhand der auch oft im üblichen Stochastikunterricht thematisierten Anwendung des Satzes von Bayes, der Diagnose seltener Ereignisse, wird ein problemorientierter und realitätsbezogener Unterrichtsvorschlag gemacht. Die Fragestellung „Was ist von einem positiven AIDS-Test zu halten?“ dient als Einstieg in den Themenbereich. Stochastische Inhalte werden dabei nicht anhand von pseudorealen Textaufgaben behandelt. Man nimmt sich bewusst Zeit, diese anhand von realistischem Textmaterial, z.B. Medientexten, mit den Lernenden zu erarbeiten. Es dauert eine geraume Zeit, bis es überhaupt zu Wahrscheinlichkeitsberechnungen kommt. Diese Berechnungen werden aus „normalen“ Baumdiagrammen mit Pfadwahrscheinlichkeiten abgeleitet, wobei keine allgemeine Herleitung der Formel von Bayes erfolgt oder von bedingter Wahrscheinlichkeit gesprochen wird. Der Kontext „AIDS“ eignet sich auch deshalb so gut in diesem Zusammenhang, weil er für Schüler eine ernsthafte Bedeutung in ihrem Alltag hat (Böer, 1997, S.56). Der Vorschlag von Böer zeigt eine Möglichkeit auf, dass und wie eine schülerrelevante Situation (AIDS-Test) im Stochastikunterricht behandelt werden könnte.

3 Das didaktische Konzept der „natürlichen Häufigkeiten“

Es besteht eine gewisse Grundüberzeugung, dass der Weg einer formalen Behandlung der bedingten Wahrscheinlichkeit und des Satzes von Bayes mit Schülern schwer ist und daher als Thema in früheren Stufen vermieden wird. Dieser Überzeugung möchte ich entgegenhalten: Es scheint tatsächlich so zu sein, dass eine wahrscheinlichkeitstheoretische Herleitung bedingter Wahrscheinlichkeit und der Bayesregel eine zu große Herausforderung zumindest in der Sekundarstufe I darstellt. Dennoch müssen zentrale Konzepte so früh wie möglich eingeführt werden und vorläufige, tragfähige Abschlüsse geschaffen werden (z.B. Schupp, 1984, S.238). Der Zugang über „natürliche Häufigkeiten“ kann einen anschaulichen, vereinfachten und intuitionsbezogenen Weg dazu bieten. Das Konzept beruht auf den in diesem Zusammenhang in Kapitel I dargestellten Erkenntnissen der kognitionspsychologischen Urteilsforschung. Wie bereits ausführlich erläutert, besteht die grundsätzliche Überzeugung, dass stochastische Intuitionen und Präkonzepte auf Häufigkeitsanschauung basieren. An diese wird angeknüpft, indem (externe) Repräsentation als wichtiger Einflussfaktor auf mentale Strategien stochastischen Denkens (vgl. Kap.I.3.) als maßgebliches Element des didaktischen Konzeptes verstanden wird.

Zunächst möchte ich als Kern des Konzeptes Modellierungsprozesse näher betrachten und anschließend auf Realisierungen in Lehr-Lern-Mitteln eingehen.

3.1 Modellierungsprozess

Der didaktische „Trick“ des Konzeptes der natürlichen Häufigkeiten besteht in der veränderten Mathematisierung der Ausgangssituation (vgl. Abb.II.11). Gegeben ist eine realistische Situation, die Informationen als absolute Häufigkeiten oder bereits „aufbereiteter“ als relative Häufigkeiten bzw. Wahrscheinlichkeiten enthält. Die übliche Mathematisierung einer stochastischen Problemstellung würde im ersteren Fall eine Übersetzung der absoluten in relative Häufigkeiten bzw. Wahrscheinlichkeiten bedeuten.

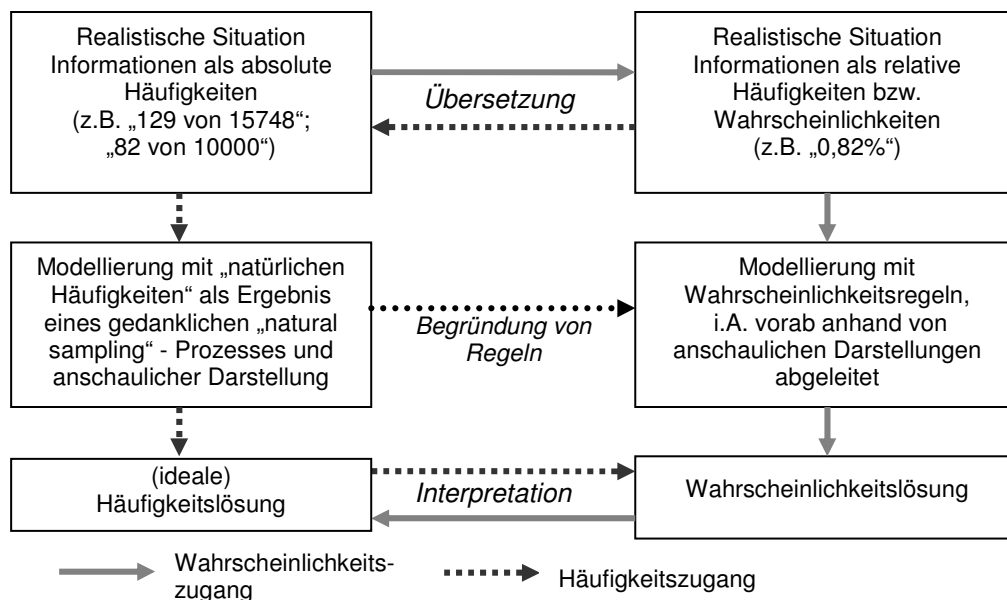


Abb.II.11: Modellierungsprozess im üblichen wahrscheinlichkeitstheoretischen Zugang und im „Konzept der natürlichen Häufigkeiten“ im Vergleich

Anschließend würden zur Verarbeitung bereits hergeleitete formale Regeln zur Anwendung kommen. Das Wahrscheinlichkeitsergebnis wäre zu interpretieren, z.B. basierend auf einer frequentistischen Anschauung.

Das Häufigkeitskonzept geht gewissermaßen „umgekehrt“ vor (vgl. Abb.II.11): Wenn Informationen als relative Häufigkeiten bzw. Wahrscheinlichkeiten vorliegen, werden sie in „natürliche“ Häufigkeiten „zurückübersetzt“. Dies geschieht in der in Kapitel I.3.1 erläuterten Weise eines gedanklichen, idealisierten Experiments. Die entstehenden absoluten Häufigkeitsverhältnisse sind das Ergebnis eines „natural-sampling“ - Prozesses. Diese (ideale) Häufigkeitslösung ist als Wahrscheinlichkeit zu interpretieren. „Umkehrung“ meint also genauer, dass eine Wahrscheinlichkeitsinterpretation erst *nach* der innermathematischen Verarbeitung erfolgt.

Der Hauptzweck ist die Erhöhung der Anschaulichkeit und Vereinfachung des Modellierungsprozesses. Große Schwierigkeiten im Umgang mit Verhältnisbegriffen (z.B. Prozentrechnung) sind bekannte didaktische Hürden in der Sekundarstufe I. Das Rechnen mit „Prozenten“ führt nicht nur bei Schülern, sondern auch bei vielen Erwachsenen immer wieder zu erschreckenden Fehlleistungen. Die natürlichen Häufigkeiten vermeiden diese zusätzliche Hürde.

Nach meiner Einschätzung wird aufbauend auf konkrete und intuitive Anschauungen mit Häufigkeiten eine wahrscheinlichkeitstheoretische Formalisierung, wenn dieses Ziel in der Sek.II erreicht werden soll, leichter fallen und begründeter erfolgen können als ohne Häufigkeitszugang. Die Regelherleitung wird ja keineswegs behindert, sondern im Gegenteil vorbereitet. Tab.II.2 zeigt einige Beispiele, welche „intuitiven Regeln“ bei der Modellierung mit natürlichen Häufigkeiten formalen Wahrscheinlichkeitsregeln entsprechen.

	Häufigkeitsmodellierung	Wahrscheinlichkeitsregeln
Bedingte Wahrscheinlichkeit	$P(A B) = \frac{ A \cap B }{ B }$	$P(A B) = \frac{ A \cap B : \Omega }{ B : \Omega } = \frac{P(A \cap B)}{P(B)}$
Produktsatz (1. Pfadregel)	$ A \cap B = B \cdot P(A B)$	$P(A \cap B) = P(B) \cdot P(A B)$
Satz von der totalen Wahrscheinlichkeit (2. Pfadregel)	$ A = A \cap B + A \cap \bar{B} $	$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(B) \cdot P(A B) + P(\bar{B}) \cdot P(A \bar{B})$
Satz von Bayes	$P(B A) = \frac{ A \cap B }{ A }$ $= \frac{ A \cap B }{ A \cap B + A \cap \bar{B} }$	$P(B A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B) \cdot P(A B)}{P(B) \cdot P(A B) + P(\bar{B}) \cdot P(A \bar{B})}$

Tab.II.2: „Intuitive Regeln“ bei der Häufigkeitsmodellierung und entsprechende Wahrscheinlichkeitsregeln

Arthur Engel (1975; 1976) hatte bereits mit dem „Wahrscheinlichkeitsabakus“ einen häufigkeitsbasierten Algorithmus entworfen, der die Anschauung vielfältiger diskreter⁵⁸ Wahrscheinlichkeitsprozesse erleichtern und diese für jüngere Schüler begreifbar machen kann. Es wird davon ausgegangen, dass jeder diskrete, stochastische Prozess in einen zufälligen „Durchlauf von Spielchips“ durch einen Graphen übersetzt werden kann. Dafür werden Regeln angegeben, die auch den frequentistisch motivierten Wahrscheinlichkeitspfadregeln entsprechen. Wenn der Algorithmus beendet ist, erhält man

⁵⁸ Engel (1975, S.1) vertrat dort die Meinung, dass auf Schulniveau die Behandlung diskreter Prozesse ausreichend ist.

immer absolute Häufigkeiten in den „Endzuständen“, die die Chancenverhältnisse dieser Zustände angeben. Das Ergebnis einer Modellierung einer „typischen Bayessituation“ mit dem Wahrscheinlichkeitsabakus wären natürliche Häufigkeiten; der Algorithmus beschreibt gewissermaßen eine Art „natural sampling“. Als Vorteile dieser häufigkeitsbasierten Algorithmen nennt Engel (1975, S.7) u.a.: „The method [...] leads in a most direct way into the main stream of modern probability, by skipping the dull and useless initial chapters in the usual college text books on probability. [...] The method enables children to compute probabilities and expectations for highly sophisticated stochastic processes usually studied at the college level. And they do it in the most efficient way by a method which is in many respects superior to the method used by experts“. Die Arbeiten zum Wahrscheinlichkeitsabakus von Engel (1975) sind nach meiner Einschätzung als eine algorithmische Grundlegung von Häufigkeitsmodellierungen diskreter, stochastischer Prozesse leider nicht genügend beachtet worden und sollen hier ausdrücklich zur Gewinnung neuer didaktischer Ideen (wieder)empfohlen werden.

3.2 Realisierung in Lernmitteln

In einem Schulbuch von Sedlmeier & Köhlers (2001) wurde das Konzept der natürlichen Häufigkeiten konsequent umgesetzt. Es wird bewusst versucht, möglichst ohne wahrscheinlichkeitstheoretische Formalisierung die wichtigen Grundinhalte der Stochastik einzuführen.⁵⁹ Dazu dienen viele alltagsrelevante Probleme, die mit möglichst einfachen und anschaulichen Mitteln gelöst werden. Eine begleitende Software ermöglicht die Modellierung vieler, auch selbstständig konstruierter, Problemsituationen. Darüber hinaus bestehen auch Möglichkeiten zur Simulation von Zufallsprozessen (z.B. einfache und mehrstufige Zufallsexperimente, zum empirischen Gesetz der großen Zahlen, zum Einfluss der Stichprobengröße auf Verteilungen) und zu speziellem Training (z.B. zur Bayesschen Revision von Wahrscheinlichkeiten). Die Gestaltung der speziellen Trainings basierte auch auf empirischen Erkenntnissen. In Kapitel III wird noch ausführlicher darauf eingegangen. Im Folgenden sollen im Vergleich zu obigen Zugängen die Inhalte zu bedingter Wahrscheinlichkeit, Konjunktion und Revision von Wahrscheinlichkeiten (Satz von Bayes) diskutiert werden (Sedlmeier & Köhlers, 2001, Kapitel 3 und 4).

Die Modellierung von Alltagsproblemen als mehrstufige Zufallsversuche wird durch das Simulationsinstrument der Software sehr gefördert. Es handelt sich dabei um eine „virtuelle Urne“, die man mit beliebigen Populationsverteilungen füllen und Ziehungen durchführen kann. Die Darstellungsform zur Speicherung der Ergebnisse der Zufallsversuche ist zunächst ein sogenanntes „Häufigkeitsraster“ (nach Sortierung: vgl. Abb.II.12, S.63). Jedes Kästchen im Raster repräsentiert in der Regel ein konkretes „Ziehungsergebnis“. Alle Wahrscheinlichkeiten lassen sich direkt aus der Darstellung als Verhältnisse natürlicher Zahlen bilden. Die Summe gleichartiger Kästchen in einem sortierten Häufigkeitsraster ergibt auch eine Art Fläche. Die verschiedenen Flächen stellen die Elemente der Ergebnismenge dar (z.B. helle Kästchen mit Kreuz: Krebs und Befund positiv). Es handelt sich gewissermaßen um eine Häufigkeitsform eines Flächendiagramms (vgl.II.2.2.1.b). Die Flächen, die die Häufigkeit eines Ergebnisses wiedergeben, repräsentieren in dem Verhältnis zur Gesamtfläche die jeweilige Wahrscheinlichkeit des Auftretens.

⁵⁹ Die traditionelle Formalisierung und Herleitung der Regeln ist aber auch im Buch als Anhang enthalten. So kann das Buch als Grundlage für Stochastikunterricht in der Sekundarstufe I und II dienen.

Die Abb.II.12 zeigt eine Häufigkeitsrasterdarstellung des obigen Diagnosebeispiels.⁶⁰ Die Legende zeigt an, dass helle Kästchen „Krebs“, dunkle „kein Krebs“, Kreuz „positiver Befund“ und kein Kreuz „negativer Befund“ bedeuten. Die Häufigkeitslösung ist der Quotient der Krebsfälle mit positiver Mammografie geteilt durch alle Fälle mit positiver Mammografie. Die Häufigkeiten werden in der Legende auch numerisch angezeigt.

$$P(\text{Krebs} \mid \text{Befund positiv}) = 12 \text{ helle Kästchen mit Kreuz} / 53 \text{ Kästchen insgesamt mit Kreuz} = 22,6\% \text{ }^{61}$$

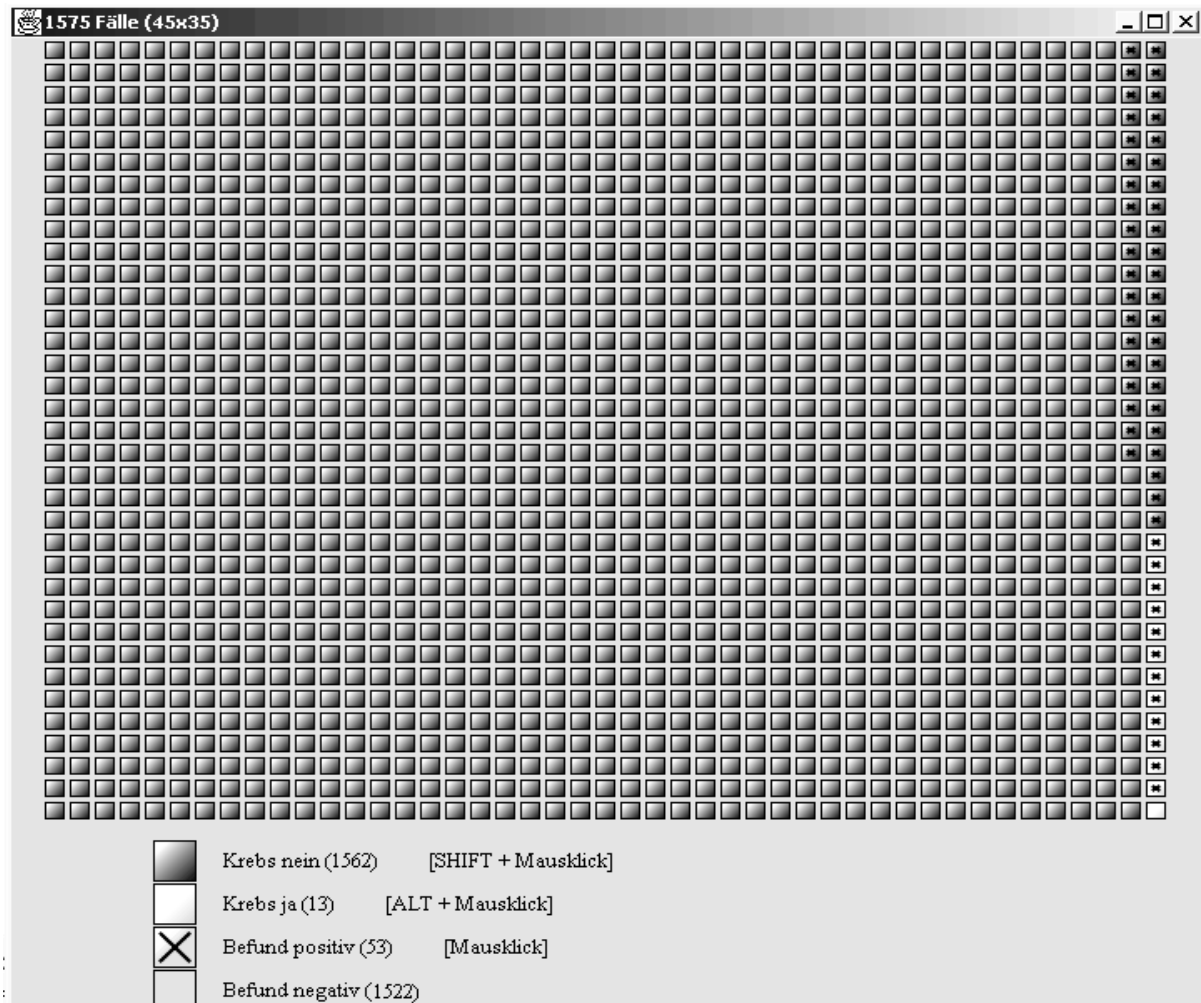


Abb.II.12: Darstellung des Diagnosebeispiels im Häufigkeitsraster (erstellt mit der Software in Sedlmeier & Köhlers, 2001)

Welchen Vorteil bietet diese Darstellung? Im Häufigkeitsraster lassen sich alle Wahrscheinlichkeiten als Verhältnisse von Häufigkeiten ablesen. Im Buch von Sedlmeier und Köhlers (2001) wird deutlich, dass die Rasterdarstellung sehr hilfreich ist, um qualitative Eigenschaften und Wahrscheinlichkeitsregeln bei mehrstufigen Zufallsexperimenten herauszuarbeiten. Die Stärke der flächenmäßigen Visualisierung liegt – wie bereits in 2.2.1.b

⁶⁰ Zwecks Übersichtlichkeit soll ein Kästchen 10 untersuchte Frauen („Fälle“) symbolisieren. Die Häufigkeitsverhältnisse und damit die Wahrscheinlichkeiten bleiben natürlich gleich.

⁶¹ entspricht 120 Krebsfälle mit positivem Befund / 530 Fälle insgesamt mit positivem Befund. 22,6% (anstatt 22,3%) ergibt sich aufgrund der vergrößerten Darstellung von eigentlich 118 Krebsfällen, die positiv testeten, mit 12 Kästchen (entspricht 120 Fälle). Derartig geringe Abweichungen sind für eine Interpretation ohne Belang.

erwähnt – in einer visualisierten Begründung von Regeln im Zusammenhang mit Wahrscheinlichkeiten. Z.B. kann die Größe zueinander „inverser“ bedingter Wahrscheinlichkeiten gegeneinander leicht abgeschätzt werden oder die „Konjunktionsregel“ $P(A \cap B) \leq P(A)$ anschaulich begründet werden, zu deren Anwendung in der kognitionspsychologischen Forschung kontrovers diskutiert wurde (vgl. für einen Überblick: Hertwig & Gigerenzer, 1999). Auch die Produktregel und totale Wahrscheinlichkeit können sehr anschaulich hergeleitet werden. Für die selbstständige Modellierung realistischer Probleme durch die Schüler sind allerdings Häufigkeitsraster womöglich weniger gut geeignet, wenn nicht wie in diesem Schulbuch Software als Hilfsmittel zur Verfügung steht.

Für den Themenbereich Bayesianwendungen („Revision von Wahrscheinlichkeit“) wird zusätzlich der „Häufigkeitsbaum“ (Abb.II.13) zur Modellierung verwendet. Die Erstellung wird ebenfalls durch die Software unterstützt. Dieser Häufigkeitsbaum entspricht der Darstellungsform, die auch in kognitionspsychologischen Arbeiten von Gigerenzer u.a. zur Veranschaulichung vorgeschlagen wurde (vgl. auch Kap.I.3.2).

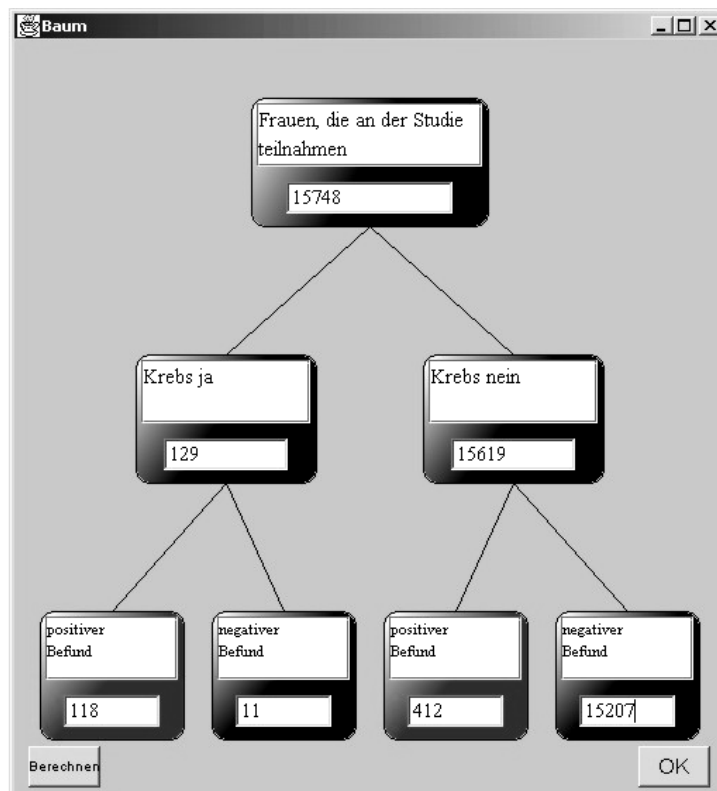


Abb.II.13: Darstellung des Diagnosebeispiels im Häufigkeitsbaum (erstellt mit der Software in Sedlmeier & Köhlers, 2001)

Auf dem didaktischen Konzept der natürlichen Häufigkeiten und der Darstellungsform Häufigkeitsbaum basiert auch der in dieser Arbeit untersuchte Unterrichtsentwurf. Der dort verwendete Häufigkeitsbaum ist die Erweiterung des Häufigkeitsbaums mit einem „umgekehrten Baumdiagramm mit Häufigkeiten“ (vgl. Abb.II.14; vgl. auch I.4.2.2). Neben dem Häufigkeitskonzept ist in dieser Darstellungsform, die ich kurz als „Doppelbaum“ bezeichnen will, „normale“ und „umgekehrte“ Sichtweise integriert.

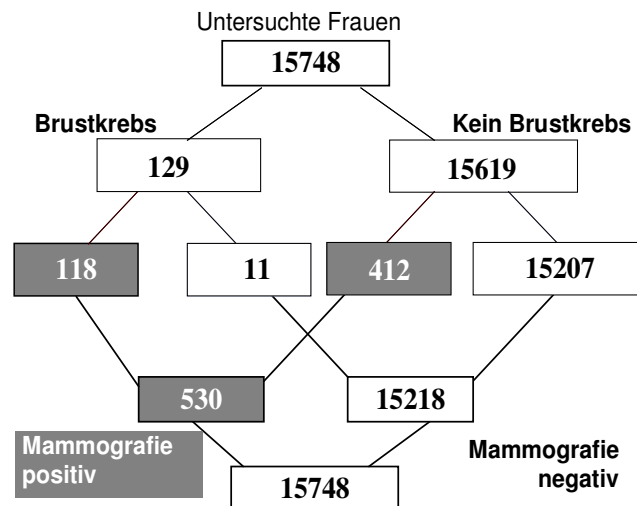


Abb.II.14: Darstellung des Diagnosebeispiels im „Doppelbaum“ mit Häufigkeiten
(vgl. Wassner, Krauss & Martignon, 2002, S.14)

Die gesuchte Wahrscheinlichkeit von z.B. $P(\text{Brustkrebs} \mid \text{Mammografie positiv})$ ergibt sich als Verhältnis von absoluten Häufigkeiten, die direkt aus der Darstellung abzulesen sind. Die Antwort, die die Darstellung „liefert“ basiert auf „natürlichen Häufigkeiten“: In 118 von 530 Fällen⁶², in denen die Mammografie positiv war, lag tatsächlich eine Krebserkrankung vor. Diese natürliche Häufigkeit kann wie folgt als gesuchte Wahrscheinlichkeit interpretiert werden:

$$P(B \mid M_+) = \frac{\text{Fälle mit Brustkrebs und positiver Mammografie}}{\text{Alle Fälle mit positiver Mammografie}} = 118 / 530 = 22,3\%$$

Es sei darauf hingewiesen, dass der Doppelbaum bezüglich der Informationsmenge identisch mit der Vierfeldertafel ist, die bei Strick (2000; vgl. Abb.II.1a) verwendet wurde.⁶³ Alle bedingten und konjunktiven Wahrscheinlichkeiten sind als Quotienten ablesbar. Meines Erachtens ist die Integration beider Sichtweisen in einer Darstellung ein weiterer Vorteil für die Veranschaulichung Bayesianischen Denkens. Die zugrundeliegende Gesamtpopulation wird unter einer Bedingung „aufgeteilt“ und unter der anderen Bedingung wieder „zusammengefügt“. Je nach betrachteter Bedingung (B, nicht B, M_+ , M_-) wird der Wahrscheinlichkeitsraum anders reduziert; die bedingten Wahrscheinlichkeiten unterscheiden sich entsprechend. Insbesondere wird die Einsicht erleichtert, dass die stochastische und die kausale Schlussrichtung nicht identisch sein muss: Sowohl in „normaler“ als auch „umgekehrter“ Sichtweise haben Bayesianische Folgerungen ihre Berechtigung (vgl. Falk, 1988).

4 Vergleichende Bewertung

Ich möchte abschließend eine zusammenfassende didaktische Bewertung aus meiner Sicht der vorgestellten Modellierungsformen von Bayesianischen Problemsituationen – dem wesentlichen Fokus meiner Arbeit – unter verschiedenen Aspekten geben, wobei sich der

⁶² Aus dem „normalen“ Häufigkeitsbaum wird genauso die Summe der positiven Befunde $118 + 412 = 530$ berechnet, die aber nicht in der Darstellung integriert ist.

⁶³ Anmerkung: Ich benutzte in diesem Kapitel absichtlich das Diagnosebeispiel mit Daten in absoluten Häufigkeiten von Strick (2000), um die verschiedenen didaktischen Zugänge und Vorschläge v.a. in ihrem Wert der Modellierung realitätsnaher Problemstellungen zu vergleichen. Derartige Aufgaben mit Daten in absoluten Häufigkeiten sind allerdings leider nicht unbedingt in Schulbüchern üblich.

Begriff „Modellierungsform“ hier vornehmlich auf Darstellungsform und Verarbeitungsformat beziehen soll. Wie deutlich wurde, bestimmt die Modellierungsform meist wesentlich den didaktischen Zugang. Tab.II.3 gibt noch einmal einen Überblick über mögliche bzw. vorgestellte Modellierungsformen. Man beachte, dass Modellierungsformen mit Chancen, die gewissermaßen Häufigkeitsverhältnisse sind, als „spezielle Häufigkeitsmodellierungen“ angesehen werden können.

Darstellungsform	Verarbeitungsformat		
	Wahrscheinlichkeit/ relative Häufigkeit	„natürliche“ Häufigkeit	Chancen- verhältnisse
1. Ohne grafische Darstellung	Bayesformel	Kap.I.3.1	Kap.II.2.2.4
2. Baumdiagramm	Abb.II.2 (1)	Abb.II.13 u. 14	Abb.II.10 (2)
3. Umgekehrte Baumdiagramm	Abb.II.4	Abb.II.14 (3)	
4. Mehrfeldertafel	Tab.II.1b	Tab.II.1a	
5. Flächendiagramm („Unit Square“)	Abb.II.5	Abb.II.12 (4)	
6. Mengendiagramm	Abb.II.3; nur qualitativ	nur qualitativ	

Anmerkungen:

(1) Das Bayes-Diagramm (vgl. Abb.II.7) ist eine Teildarstellung des Baumdiagramms

(2) Bei Borovcnik als „inverses“ (*nicht gleich* umgekehrtes) Baumdiagramm bezeichnet. Nur die A-priori-Wahrscheinlichkeiten werden als Chancenverhältnisse angegeben.

(3) Z.B. integriert im „Doppelbaum“ (Wassner et al., 2002)

(4) Z.B. als „Häufigkeitsraster“ (Sedlmeier & Köhlers, 2001)

Tab.II.3: Überblick über Modellierungsformen bei Bayesianischen Problemen

Die folgende Tabelle II.4 stellt einen Versuch dar, die gezeigten Modellierungsformen unter folgenden Aspekten zu bewerten:

1. Ist eine selbstständige Modellierung eines realistischen Problem es durch den Lernenden ohne Hilfsmittel möglich?
2. Können aus dem Modell allgemeine Wahrscheinlichkeitsregeln abgeleitet werden?
3. Können diese Regeln durch das Modell begründet werden?
4. Wird das sequenzielle, mehrstufige Vorgehen des Zufallsprozesses unterstützt?

Aspekte	Baumdiagramm (bzw. umgek. Baum)		Mehrfelder- tafel		Flächen- diagramm		Mengen- diagramm
	W	H	W	H	W	H	
1. Modellierung realer Probleme	o	+	o	+	-	-	-
2. Ableitung von Wahrscheinlichkeitsregeln	o	o	o	o	o	o	o
3. Begründung von Regeln	o	+	o	+	+	+	+
4. Sequenzielle Darstellung	+	+	o	o	-	-	-

Legende:

W: „im Verarbeitungsformat Wahrscheinlichkeit / rel. Häufigkeit“, H: „im Verarbeitungsformat natürl. Häufigkeit“
 +: „besonders gut geeignet“, o: „geeignet“, -: „weniger gut geeignet“.

Tab.II.4: Bewertung von Modellierungsformen bei Bayesianischen Problemen

Diese vergleichenden und diskussionsoffenen Einschätzungen möchte ich im Folgenden näher erläutern und begründen:

zu 1.

Generell liegt der schlagende Vorteil der Häufigkeitsdarstellungen im Vergleich zu den Wahrscheinlichkeitsdarstellungen in der Einfachheit und Anschaulichkeit. Das obige Diagnosebeispiel stellt eine realistische Bayesianische Problemsituation dar, die mit

Baumdiagrammen und Mehrfeldertafeln gut zu modellieren ist. Hinsichtlich der selbstständigen Anwendung durch die Lernenden erscheinen mir aber der Häufigkeitsbaum (am besten als Doppelbaum) und die Mehrfeldertafel mit Häufigkeiten besser geeignet als die entsprechenden Wahrscheinlichkeitsversionen. Flächendarstellungen mit Wahrscheinlichkeiten und auch das Häufigkeitsraster erweisen sich für das selbstständige Modellieren und Anwenden durch den Lernenden (ohne Softwareunterstützung) eher als zu kompliziert. Das Mengendiagramm ist zu ungenau und damit eigentlich hierfür ungeeignet.

zu 2 und 3.

Die Herleitung von Wahrscheinlichkeitsregeln, die zu Zielen des gymnasialen Abschlussniveaus gehört, geschieht meist mithilfe von Wahrscheinlichkeitsbäumen und Mehrfeldertafeln mit Wahrscheinlichkeiten in der Schule. Allerdings lassen sich z.B. die Pfadregeln zwar ablesen, aber nur mit wahrscheinlichkeitstheoretischem Grundwissen (z.B. Additionssatz bei dichotomen Ereignissen) exakt begründen. Diese Regeln können aus Häufigkeitsbäumen direkt abgeleitet und auch durch Häufigkeitsanteile sofort begründet werden. Deshalb sehe ich unter dem Begründungsaspekt klare Vorteile von Häufigkeitsmodellierungen.

Im Diagnosebeispiel (vgl.II.2.1):

Pfadproduktregel: $P(B) \cdot P(M_+ | B) = (129 : 15748) \cdot (118 : 129) = 118 : 15748 = P(B \text{ und } M_+)$

Pfadsummenregel (totale Wahrscheinlichkeit):

$P(B \text{ und } M_+) + P(\text{nicht } B \text{ und } M_+) = 118 : 15784 + 412 : 15784 = 530 : 15784 = P(M_+)$

Flächen- und Mengendiagramme ermöglichen über Flächeninhaltsvergleiche bzw. mengentheoretische Anschauung auch direkte Veranschaulichungs- und Begründungsmöglichkeiten der Regeln (vgl. 2.2.1.b).

zu 4.

Baumdiagramme bieten gegenüber anderen Darstellungen weitere Vorteile bei Bayesianischen Problemen. Sie unterstützen die Mehrstufigkeit des Zufallsprozesses und die sequenzielle Denkweise einer Neubewertung von Wahrscheinlichkeiten in besonderer Weise. Mehrfeldertafeln und noch mehr Flächen- bzw. Mengendiagramme sehe ich dabei im Nachteil, da sie weniger eindeutig eine „Leserichtung“ beinhalten, die auch dieser Denkweise entspricht. Darüberhinaus sind Prozesse mit mehr als zwei „Stufen“ nur in Baumdiagrammen übersichtlich darstellbar, also z.B. bei Neubewertung von Wahrscheinlichkeit anhand mehrerer Daten (vgl. auch Kap.I, Abb.I.5).

Gemäß dieser didaktischen Analyse haben unter diesen Aspekten Baumdiagramme mit Häufigkeiten das höchste „didaktische Potential“ als Modellierungsform von Bayesianischen Problemen. Es ist meines Erachtens voll gerechtfertigt, dass die Baumstruktur im Zusammenhang mit mehrstufigen Zufallsversuchen, insbesondere bedingter Wahrscheinlichkeit und Bayes, das vorherrschende grafische Modell ist. Nach dieser didaktisch-theoretischen Betrachtung werden im folgenden Kapitel Trainingsstudien näher erläutert, die einige der dargestellten Modellierungsformen in ihrem Nutzen für das Lehren und Lernen Bayesianischen Denkens vergleichend untersuchen.

KAPITEL III

TRAINING ZUR FÖRDERUNG BAYESIANISCHEN DENKENS

Verglichen mit der hohen Anzahl an Untersuchungen zu intuitivem Bayesianischem Denken (vgl. Kapitel I) gab es erstaunlich wenige systematische Untersuchungen, die eine Förderung durch Instruktion bzw. Training als Gegenstand haben. Aus den frühen Befunden eines „Konservatismus“ bei Bayesianischem Denken (Phillips & Edwards; 1966) wie auch dem „Heuristikprogramm“ (Kahneman, Slovic & Tversky, 1982) gingen kaum erfolgreiche Versuche hervor, Bayesianisches Denken systematisch zu fördern (vgl. genauer Kap.I.1 und I.2.1). Shaughnessy stellte 1992 fest (S.475): „To my knowledge, there has not been any research reported that has attempted to change students’ misconceptions or beliefs specifically on conditional tasks. The literature is long on excellent didactical suggestions, but short on hard research in the area“. Diese Situation, die sich nur wenig verändert hat, folgte sicher nicht zuletzt aus den eher pessimistischen Folgerungen der „Heuristics and biases“-Forschung.

Die Studien zu intuitivem Bayesianischem Denken (ohne Instruktion) lieferten wichtige Erkenntnisse, dass intuitive Strategien, die auf Häufigkeitsanschauungen beruhen, in hohem Maße auch unter formal-theoretischen Kriterien zu korrekten Ergebnissen führen können. Es ergaben sich direkte Ansatzpunkte, wie die Weiterentwicklung dieser häufigkeitsbasierten „Grundintuitionen“ gefördert werden kann, die in Kapitel II bereits didaktisch analysiert wurden.

In diesem Kapitel werden Trainingsstudien im Labor zum Nutzen des „Häufigkeitskonzepts“ mit Universitätsstudenten von Sedlmeier und Gigerenzer (2001; Sedlmeier, 1999, Kap.7 und 8) dargestellt. Anschließend werden hierzu eigene Trainingsstudien mit Schülern der Sekundarstufe I und II besprochen. Vorher gehe ich kurz auf eine Trainingsstudie von Bea (1995) ein, in der auf unterschiedlicher Regeldarstellung basierende Trainingsformen zu „bedingter Wahrscheinlichkeit“ verglichen wurden.

1 Wahrscheinlichkeitsbasiertes Training

Frühe Trainingskonzepte basierten meist darauf, dass ein Teilaspekt der Modellierung für Schwierigkeiten der Lernenden verantwortlich gemacht wurde und deshalb dieser fokussiert wurde. Die teilweise konträren Befunde über Präkonzepte beim Bayesianischen Denken (z.B. „Base rate neglect“ und „conservatism“) lassen den Schluss zu, dass die Schwierigkeiten bei „Wahrscheinlichkeitsmodellierungen“ nicht durch einen einzigen Aspekt verursacht werden. In einer Studie von Bea (1995) wurden basierend auf der Identifizierung verschiedener Schwierigkeiten (z.B. Verwechslung des bedingten und bedingenden Ereignisses) der Lernenden Trainingsvarianten u.a. zur Anwendung des Satzes von Bayes verglichen.

1.1 Trainingsbeschreibung

Der Trainingsinhalt war folgendermaßen zweigeteilt:

- a) Einer allgemeinen Einführung der bedingten Wahrscheinlichkeit (mit Mengendarstellungen) folgte die wahrscheinlichkeitstheoretische Ableitung des Satzes von Bayes (über den Satz von der totalen Wahrscheinlichkeit) mit einer jeweiligen

trainingspezifischen Darstellungsform. Es wurden 4 verschiedene Trainingsvarianten unterschieden, die auf folgenden Darstellungsformen⁶⁴ basierten:

- numerisch-formale Darstellung
- Baumdiagramm mit Wahrscheinlichkeiten (vgl. II.2.1)
- Umgekehrtes Baumdiagramm mit Wahrscheinlichkeiten (vgl. II.2.2.1.a)
- Einheitsquadrat (vgl. II.2.2.1.b)

b) Anhand dreier Übungsaufgaben wurde des theoretisch Gelernte angewandt. Die Aufgaben fokussierten dabei auf typische Schwierigkeiten der Lernenden in diesem Bereich (vgl. Bea, 1995, S.142f.).

1.2 Methodisches Vorgehen

Untersucht wurden 289 Studenten der Wirtschaftswissenschaften im 3. Fachsemester mit einem Durchschnittsalter von 22 Jahren. Die Versuchspersonen waren (zum größten Teil) durch den Stochastikunterricht in der schulischen Sekundarstufe, durch eine Vorlesung in beschreibender und das erste Drittel einer Vorlesung in beurteilender Statistik vorgebildet.

Zur Erfassung von Trainingserfolgen wurde von den Versuchspersonen ein Vortest und ein unmittelbarer Nachtest absolviert, sowie ein späterer Nachtest nach etwa 6 Monaten (jedoch nur etwa von einem Drittel der Teilnehmer). Es wurden 12 Testitems entworfen, die jeweils das Auftreten spezieller Schwierigkeiten der Lernenden im Zusammenhang mit bedingten Wahrscheinlichkeiten provozieren sollten (vgl. Bea, 1995, S.142f.). Die einzelnen Aufgabenitems sind somit rein nach inhaltlichen Kriterien klassifiziert.

Zwei Aufgabenitems erforderten im engeren Sinne Bayesianisches Denken, die bei Bea fehleranalytisch als „Aufgaben, die zur Verwechslung von bedingtem und bedingendem Ereignis verleiten“ (ebenda, S. 191, Komplex 3) bezeichnet wurden.⁶⁵ Es wird nicht rechnerisches Modellieren abverlangt, sondern die Beurteilung mathematischer Zusammenhänge.⁶⁶ Z.B. lautet eine Aufgabe im Kontext einer medizinischen Diagnosesituation (HIV-Test), wie sich die A-posteriori-Wahrscheinlichkeit $P(\text{krank} \mid \text{Test positiv})$ ändert, wenn die Basisrate $P(\text{krank})$ in der Population sinkt [Antwortmöglichkeiten: sie steigt, sie sinkt, sie ändert sich nicht, die Frage läßt sich mit den Daten nicht beantworten].

1.3 Ergebnisse

Die Evaluation der Trainingseffekte erfolgte in einer Ermittlung der Veränderungen der (mittleren) Leistungen zwischen Vor- und Nachtest(s). Diese Leistung wurde durch Orientierung an der normativen Lösung mit „richtig“ bzw. „falsch“ operationalisiert. Die Summe aller richtig gelösten Testitems diente als Indikator für die „globale“ Problemlöseleistung, wobei eine „analytische“ und eine „intuitive“ Problemlöseleistung unterschieden wurde. Als „analytisch korrekt“ wurden normativ korrekte Lösungen bezeichnet, bei denen auch Strategien, Begründungen oder Argumentationen als richtig bewertet werden konnten. Wenn keine als richtig bewertbare Strategie oder Begründung erkennbar war, aber trotzdem ein normativ korrektes Ergebnis vorlag, wurde die Aufgabe als „intuitiv korrekt“ bezeichnet. Als Maß für die „globale Verbesserung“ dient die Differenz zwischen der Anzahl korrekter Lösungen im Nach- und Vortest. Eine weitere

⁶⁴ Genaue Erläuterungen zu den grafischen Modellen in Bea, 1995, S.149f. bzw. Kap.II dieser Arbeit.

⁶⁵ Diese Verwechslung wird als hauptverantwortlich für das Auftreten von Fehlurteilen bei Bayesianischem Denken gesehen.

⁶⁶ im Sinne der Unterscheidung von Hiebert (1986): Überwiegend „konzeptuelles“ Wissen, weniger „prozedurales“ (vgl. auch Neubrand, M. et al., 2002).

Evaluationskomponente war die „kognitive Zugänglichkeit“ einer grafischen Darstellungsform. Als Maß dafür wurde die Differenz zwischen den Anzahlen der Items im Nach- und Vortest gewählt, bei denen die trainierte grafische Darstellung explizit zur Lösung herangezogen wurde.

Das Ergebnis der Auswertungen des *unmittelbaren Nachtests* ergab zunächst eine deutlich höhere Leistungssteigerung in allen Trainingsgruppen im Vergleich zu einer Kontrollgruppe, die kein Training erhalten hatte. Das galt sowohl für intuitive als auch analytische Problemlöseleistungen. Allerdings konnte kein Unterschied zwischen den verschiedenen Trainingsvarianten festgestellt werden. Insbesondere war also Training mit grafischen Darstellungen dem rein numerischen Training nicht überlegen. Auch in der kognitiven Zugänglichkeit wurde kein Unterschied zwischen den grafischen Darstellungsformen festgestellt.

Der nach 6 Monaten durchgeführte *zweite Nachtest* ergab ein anderes Bild. Während die Leistungen der Gruppe mit numerischem Formeltraining deutlich fielen, erhöhten sie sich mit grafischen Darstellungen sogar. Hochsignifikante Unterschiede ergaben sich zwischen der numerischen und den grafischen Trainingsbedingungen. Darüber hinaus wurde innerhalb der Trainings mit grafischen Darstellungen das Einheitsquadrat als überlegenes Darstellungsmittel identifiziert. Bei der Frage der kognitiven Zugänglichkeit schnitt das umgekehrte Baumdiagramm deutlich schlechter ab als das normale Baumdiagramm und das Einheitsquadrat.

Wenn man differenziert nach Aufgabengruppen nur die beiden Aufgaben zu Bayesianischem Denken (Komplex 3) betrachtet, war auch der Unterschied zwischen numerischen und grafischen Trainingsbedingungen sehr deutlich, jedoch das Einheitsquadrat nicht mehr statistisch abgesichert überlegen (Bea, 1995, S.192).

1.4 Diskussion

Als deutliches Ergebnis dieser Studie zeigte sich, dass *statistisches Training* im Bereich bedingte Wahrscheinlichkeit *durchaus effektiv* sein kann. Es sei darauf hingewiesen, dass positive Befund sich von bisherigen Trainingsversuchen unterschied. Insofern sind Ergebnisse der Trainingsstudie von Bea ein Gegenbeweis für rigide Folgerungen aus dem „Heuristikprogramm“, dass der Mensch nicht zu reparierende kognitive Defizite im Umgang mit Wahrscheinlichkeit hat (z.B. Gould, 1992; vgl. Kap.I.2.1).

Generell weist die Studie auf eine Erreichbarkeit längerfristiger Lernerfolge durch Training im Bereich „bedingte Wahrscheinlichkeit“ hin. Ergebnisse stützen die Interpretation, dass grafische Darstellungen besser rekonstruierbar und damit besser erinnerbar als formal-numerische Darstellungen sind. Innerhalb der Gruppen mit grafischer Darstellung erwies sich das Einheitsquadrat als überlegen. Daraus wird gefolgert, dass im Bereich bedingte Wahrscheinlichkeit das Einheitsquadrat didaktisch am besten geeignet ist und als einzige Modellierungsform verwendet werden sollte (Bea, 1995, S.195). In der Tat können Flächendiagramme die Herleitung von Wahrscheinlichkeitsregeln, wie in Kap.II.2.2.1.b erläutert, besonders gut veranschaulichen und Begründungen liefern. Allerdings zeigen sich erhebliche Schwächen, wenn sie zur Modellierung realistischer Probleme angewendet werden. Meines Erachtens reichen Flächendiagramme somit nicht als Modellierungshilfen für Lernende bei Bayesianischem Denken aus, können aber sicher wertvolle Hilfe auf dem Weg zu einer formalen Behandlung sein.

2 Training mit dem Konzept der „natürlichen Häufigkeiten“

Das Konzept der „natürlichen Häufigkeiten“ wurde bereits aus kognitionspsychologischer (Kapitel I) und didaktischer Sicht (Kapitel II) genau erläutert und diskutiert. Aufgrund der theoretisch begründeten und empirisch erreichten Befunde (z.B. Gigerenzer & Hoffrage, 1995) wurde dieses Konzept als zentrales Element in Trainingsprogrammen von Sedlmeier & Gigerenzer (2001; Sedlmeier, 1999) realisiert und gegen andere Trainingskonzepte getestet.

Der Kern des Trainingsprogramms ist die Modellierung von Bayesianischen Problemstellungen, die Wahrscheinlichkeiten bzw. relative Häufigkeiten enthalten, mit natürlichen Häufigkeiten. Durch diese „Übersetzung“ (vgl. II.3.1) soll ein deutlicher Trainingseffekt auftreten, da in dieser Form offensichtlich intuitives Denken unterstützt wird (genaue Erläuterungen in Kap. I.3).

Die Realisierung erfolgte in *computerbasierten Tutorien*⁶⁷, die auf zwei speziellen Häufigkeitsdarstellungen, dem Häufigkeitsbaum und dem Häufigkeitsraster (vgl. Kap. II.3.2, Abb. II.12 und II.13) basierten. Getestet wurde der unmittelbare Lernerfolg und die zeitliche Stabilität des Lernerfolges verglichen mit nicht-häufigkeitsbasierten Trainingsversionen. Im Folgenden werde ich diese Trainingsstudien genauer beschreiben und Ergebnisse diskutieren. Hingewiesen sei auf weitere empirische Befunde zum „Häufigkeitskonzept“ als Trainingsmethode von Kurzenhäuser & Hoffrage (2002). Eine Umsetzung erfolgte in einem einstündigen Seminar für Medizinstudenten, denen die Verwendung des Häufigkeitskonzeptes zur Beurteilung diagnostischer Informationen gelehrt wurde. Im Vergleich mit einem regelbasierten Unterricht ergaben sich deutliche Leistungsunterschiede (gemessen nach etwa 2 Monaten) bei der Lösung eines typischen Bayes-Diagnoseproblems (z.B. „Mammografie“). 47% der Studenten, die mit dem Häufigkeitskonzept unterrichtet wurden, gaben das korrekte Ergebnis an im Vergleich zu 16% nach regelbasiertem Unterricht (vgl. ausführlich Kurzenhäuser & Hoffrage, 2002).

2.1 Trainingsbeschreibung

In den beiden Versionen der „Häufigkeitstrainings“ (vgl. Sedlmeier, 1999, S.93) war der Haupttrainingsinhalt die bereits gezeigte Modellierungsmethode von „dichotomen“ Bayesproblemen⁶⁸ durch Übersetzung der numerischen Informationen in Aufgabentexten (meist im Prozentformat) in natürliche Häufigkeiten und Darstellung mit einem entsprechenden grafischen Modell (Baum bzw. Raster).

Als Kontrollversion wurde ein inhaltlich analoges Training entworfen, das die Regelherleitung und Anwendung zum Inhalt hatte. Die theoretische Herleitung der Bayesregel (in einer Terminologie mit Hypothese und Daten, vgl. Einführung, Gleichung 0) erfolgte mit mengentheoretischen Begründungen. Anschließend wurde vorgeführt und trainiert, wie Aufgaben mit der hergeleiteten Regel formal zu modellieren und zu lösen sind (vgl. Kap. II.2.1).

Alle Trainingsvarianten beinhalteten zwei Modi: In einem „Instruktionsmodus“ wurde den Lernenden anhand zweier anwendungsbezogener Bayesprobleme schrittweise erklärt, wie sie mit der jeweiligen Modellierungsmethode zur Lösung gelangen. Danach konnten die Lernenden in einem „Praxismodus“ *selbstständig* an acht analogen Aufgaben die jeweilige

⁶⁷ Versionen der Tutorien sind Inhalt des computergestützten Tutorsystems „BasicBayes“, das mit theoretischen Hintergründen in Sedlmeier (1997) detailliert beschrieben wird.

⁶⁸ Zur Verdeutlichung vgl. Kap. I.4.1.3: Der „einfachste“ dichotome Fall kann z.B. zu „komplexeren“ Bayessituationen (Krauss et al., eingereicht) erweitert werden. In Kap. III.3.2.2 werden weitere Aufgabentypen zu Bayesianischem Denken unterschieden und klassifiziert.

erlernte Methode üben. Bei Erreichen bzw. Nicht-Erreichen einzelner Lösungsschritte, gab das Programm „Feedback“ und Hilfestellungen („step-by-step feedback“). Die gestuften Lösungshilfen wurden im Bedarfsfall detaillierter, so dass gewährleistet war, dass *alle* Versuchspersonen korrekte Lösungen erreichen und das Training vollständig durchlaufen konnten. Das Training dauerte je nach persönlichem Arbeitstempo und Lernfortschritt 1 bis 2 Stunden.

Abb.III.1 (S.72) zeigt beispielhaft eine Kopie des Computerbildschirms, der während des Trainings (in allen Varianten) in drei Fenster unterteilt war:

- Das „*Problemfenster*“ (rechts, oben) enthält die Problemstellung mit den numerischen Informationen (meist in Prozentwerten).
- Das „*Tutorfenster*“ (rechts, unten) dient zur Erklärung, Instruktion und Kommunikation mit dem Lernenden über dort erscheinende Eingabefelder⁶⁹ bzw. „Buttons“.
- Im „*Repräsentationsfenster*“ (links) wird jeweils die in der entsprechenden Trainingsversion verwendete Modellierungsform (Häufigkeitsbaum, Häufigkeitsraster, formale Wahrscheinlichkeitsregel) angezeigt. Die Lernenden interagieren auch mit dem Trainingsprogramm über das Einfügen von numerischen Werten oder Termen in die Darstellungen.

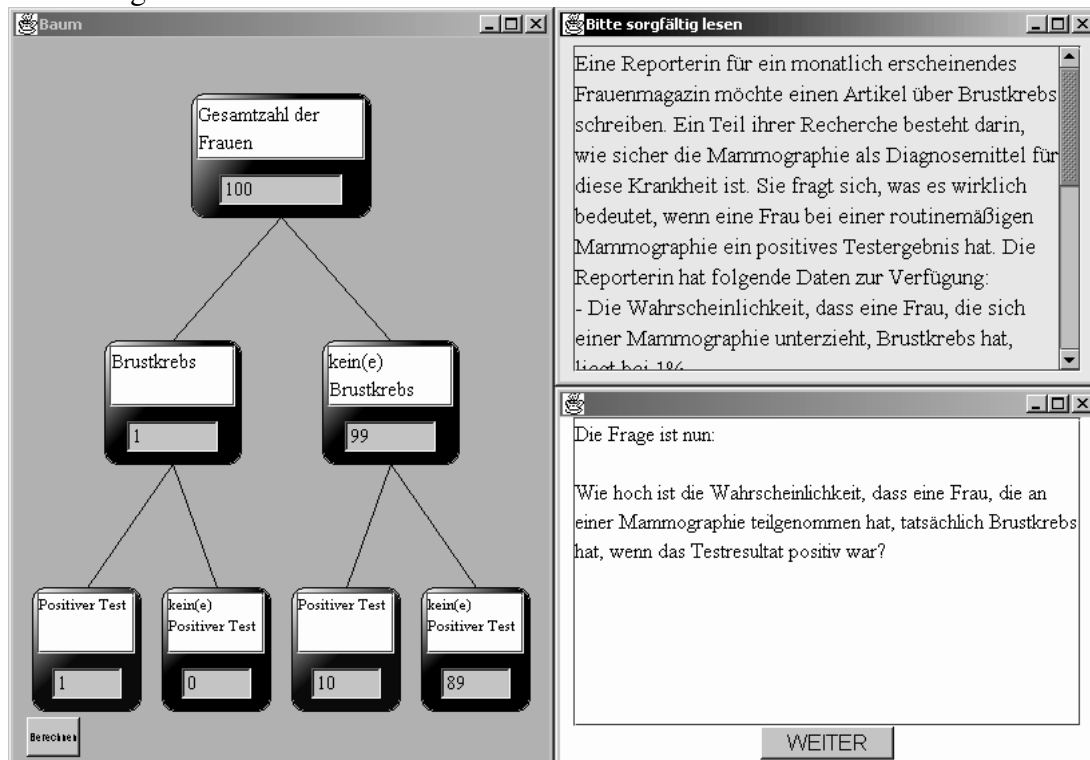


Abb.III.1: Training mit dem Häufigkeitsbaum (vgl. Sedlmeier 1997, 2001⁷⁰)

⁶⁹ Mögliche Eingabeformate waren numerische Werte oder Terme bzw. „multiple choice“.

⁷⁰ „Screenshots“ stammen aus einer späteren, inhaltlich identischen Version der Bayes-Tutorien (Sedlmeier & Köhlers, 2001). Die ursprünglichen Versionen in der Computersprache Macintosh CommonLISP wurden zur Erhöhung der Benutzbarkeit in JAVA übersetzt.

2.2 Methodisches Vorgehen

Untersucht wurden zunächst 62 Universitätsstudenten verschiedener Fachbereiche und Semesterzahl. Die Versuchspersonen verfügten außer Stochastikunterricht in der Schule über keine spezielle Ausbildung in Stochastik an der Universität.

Zur Erfassung von unmittelbaren Trainingserfolgen wurde ein Vortest (Test 1) und ein Nachtest (Test 2) absolviert. Zur Erfassung der Stabilität des Trainingserfolges wurden auch zwei spätere Nachtests nach 1 Woche (Test 3) und nach 5 Wochen (Test 4) durchgeführt. Die Tests waren auch computergestützt und bestanden in allen Trainingsgruppen aus jeweils 10 verschiedenen „dichotomen“ Bayesproblemen. Auf dem Computerbildschirm erschienen jeweils ein Aufgabentext⁷¹, die Fragestellung und ein Lösungsfenster, in das das errechnete Wahrscheinlichkeitsergebnis oder ein Lösungsterm⁷² eingetragen werden sollte.

Unterschieden wurden bei den strukturell analogen Testproblemen zwei Kategorien: Sogenannte „Transferprobleme“ unterschieden sich im Kontext von den Trainingsproblemen und jeweils 2 Probleme der Nachtests waren identisch mit Trainingsproblemen. Diese Zusammensetzung der Tests sollte eine Differenzierung zwischen reiner Reproduktion und Anwendung auf andere Kontexte ermöglichen. Grundsätzlich war ein spezieller Aufgabentyp – der des dichotomen Bayesproblems – Gegenstand von Training und Tests. Die Erfassung von Kompetenzen durch diesen Test fokussierte also, anders als bei Bea (1995; I.1), auf rechnerisches Modellieren (im Sinne von „procedural knowledge“; Hiebert, 1986).

2.3 Ergebnisse

Unmittelbare Lernerfolge wurden durch die Veränderung der (mittleren) Testleistungen zwischen Test 1 und Test 2 operationalisiert. Die Stabilität der Trainingseffekte über die Zeit hinweg wurde durch Vergleich von Test 3 bzw. 4 mit Test 2 gemessen. Die „Transferleistungen“ wurden durch den Vergleich von Leistungen in Transferproblemen mit Leistungen in Trainingsproblemen bewertet. Die Leistung pro Aufgabenitem wurde gemessen am normativen Ergebnis mit „richtig“ bzw. „falsch“ bewertet. Dabei gab es zwei Bewertungskriterien, ein „striktes“ und ein „liberales“. Mit „Strikt richtig“ wurden nur exakt dem normativen Ergebnis entsprechende Lösungen (evtl. korrekt gerundet) gewertet, „liberal richtig“ gewertet wurden auch Lösungen, die in einem Bereich ± 5 Prozentpunkte um das normativ richtige Ergebnis lagen.⁷³

Die Versuchspersonen zeigten kaum bewertbare Leistungen im Vortest (Test 1; vgl. Abb.III.2a). Das Ergebnis des unmittelbaren Nachtests (Test 2) ergab eine deutliche Leistungssteigerung in allen Trainingsgruppen. Im „strikten“ Bewertungskriterium war die Testleistung in den Gruppen mit Häufigkeitstraining (70-80 Prozentpunkte) in etwa *doppelt* so hoch wie nach Regeltraining (ca. 35%).

Als noch wesentliches Kriterium ist die *Stabilität des Lernerfolges* anzusehen. Der meist feststellbare nach gewisser Zeit einsetzende Verfall des Lernerfolges war nach dem Regeltraining zu beobachten. Nach Häufigkeitstraining ergab sich diese Verfallskurve nicht, im Gegenteil lagen die Leistungen nach 5 Wochen (Test 4) sogar noch auf höherem Niveau.

⁷¹ Dabei waren numerische Angaben in Wahrscheinlichkeiten oder relativen Häufigkeiten.

⁷² bestehend aus rationalen Zahlen und additiven oder multiplikativen Verknüpfungen.

⁷³ Ich berichte hier nur Ergebnisse im strikten Kriterium.

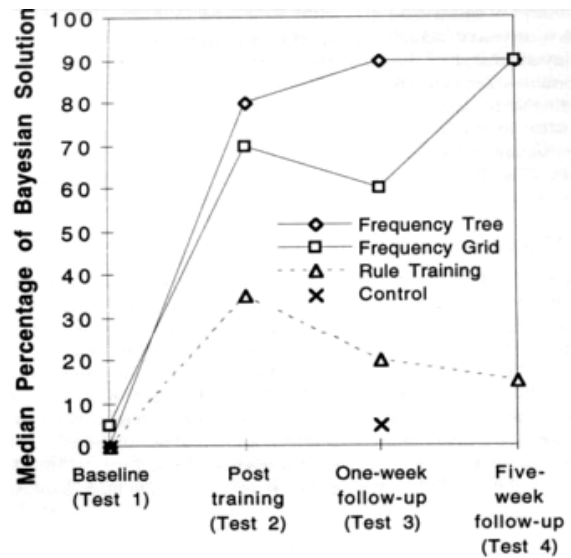


Abb.III.2a: Testleitungen in den verschiedenen Trainingsgruppen in Studie 1 (vgl. Sedlmeier, 1999, S.102)

„Transferaufgaben“ (mit neuen Anwendungskontexten) wurden in allen Trainingsgruppen fast genauso gut wie Trainingsaufgaben (ohne neue Anwendungskontexte) gelöst. Es war also kein Unterschied zwischen Regel- und Häufigkeitstrainings hinsichtlich Anwendung auf andere Kontexte festzustellen, die in allen Trainingsbedingungen nahezu optimal erfolgte.

In einer *zweiten Studie* sollte geklärt werden, ob die höheren Lernerfolge nach Häufigkeitstraining aus der Tatsache resultieren, dass grafische Modelle verwendet wurden und bei Regeltraining nicht. Bea (1995) hatte diesbezüglich einen Unterschied in langfristigen Effekten von „Regeltraining ohne“ und „Regeltraining mit grafischer Darstellung“ festgestellt. Sedlmeier & Gigerenzer (2001) verwendeten zur Untersuchung eine analoge computerbasierte Trainingsversion, die auf der Modellierungsform „Baumdiagramme mit Wahrscheinlichkeiten“ basierte, und wiederholten die Studie entsprechend mit den 3 Trainingsbedingungen „Regeltraining“, „Häufigkeitsbaumtraining“ und „Wahrscheinlichkeitsbaumtraining“.

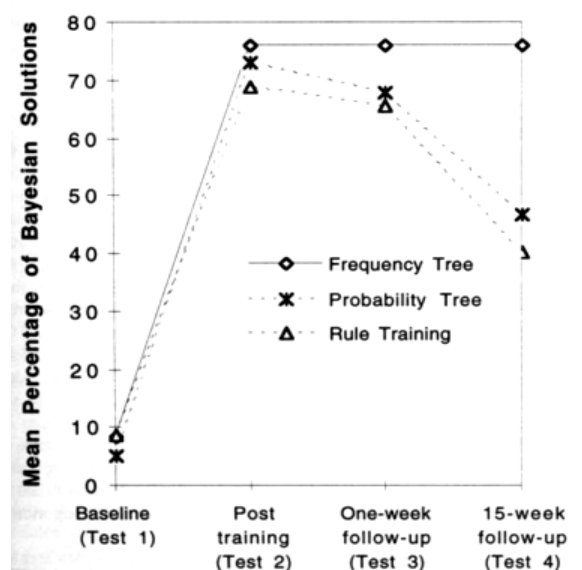


Abb.III.2b: Testleitungen in den verschiedenen Trainingsgruppen in Studie 2 (vgl. Sedlmeier, 1999, S.100)

Sie fanden zur Stabilität des Lernerfolgs folgenden Befund (vgl. Abb.III.2b): Trainingseffekte nach „Häufigkeitsbaumtraining“ blieben stabil (3 Messungen: unmittelbar, 1 Woche und 15 Wochen nach Training), doch sowohl Regeltraining *als auch* Training mit dem Wahrscheinlichkeitsbaum unterlagen dem „üblichen“ Verfall des Lernerfolgs.

2.4 Diskussion

Diese Studie zeigte, wie bereits III.1, dass stabile Lernerfolge durch Training bei Bayesianischem Denken erreichbar sind. Es sei noch einmal betont, dass solche Erfolge bis dato vielfach als nicht erreichbar galten. Im Unterschied zu Trainingsstudien von Bea (1995), die den Nutzen verschiedener grafischer Modelle für Kompetenzen zur Verwendung von Wahrscheinlichkeitsregeln im Bereich bedingte Wahrscheinlichkeit verglichen⁷⁴, verdeutlichten die Studien von Sedlmeier & Gigerenzer basierend auf dem Ansatz der „ökologischen Rationalität“ (Kap.I.3), dass ein derartiges wahrscheinlichkeitsbasiertes Training weit weniger effektiv als Training mit Häufigkeitsmodellierungen ist. Das gilt insbesondere, wenn man die langfristigen Lernerfolge betrachtet. Stabile Lernerfolge stellen das grundsätzliche Ziel jeglichen Trainings dar, das sich von kurzfristigem „Einpauken“ unterscheidet. Ergebnisse der Tests 3 und 4 sind deshalb als die bedeutenderen einzustufen. Die Verwendung von grafischen Darstellungen per se kann gemäß der Ergebnisse von Studie 2 nur wenig an einem längerfristigen Verfall des Lernerfolgs ändern, wenn sie nicht auf Häufigkeitsformaten basieren.

In einem umfassenderen Rahmen betrachtet, wurden Trainingsszenarien miteinander verglichen, die einerseits formal-theoretische andererseits intuitiv-anwendungsbezogene Kompetenzen betonen. Das Ergebnis stützt die Argumentation der zugrundeliegenden Theorie der „ökologischen Rationalität“: Grundlegende Intuitionen beim Urteilen und Entscheiden unter Unsicherheit basieren auf Häufigkeitsformaten. Der Umgang mit Wahrscheinlichkeitsformaten (etwa als Prozentwerte) erfolgt gewissermaßen nicht intuitiv und muss erst mühsam, gegebenenfalls aufbauend auf Häufigkeitsanschauungen entwickelt werden. Deshalb erreichen Trainingsszenarien, die auf grundlegenden häufigkeitsbasierten Intuitionen aufbauen, höhere und stabilere Lernerfolge.

3 Trainingsvergleiche mit Schülern

Wie bereits ausführlich diskutiert, ist das Themenfeld der bedingten Wahrscheinlichkeit, insbesondere der Anwendung des Satzes von Bayes auch ein schuldidaktisch umfangreich reflektierter Bereich. Wie soll man diese oft als schwer (zu vermitteln) empfundenen Probleme mit Schülern⁷⁵ behandeln? Die heutige durchschnittliche Vorgehensweise im Unterricht und in Schulbüchern ist wahrscheinlichkeits- und regelbasiert (vgl. Kap.II). Der Erfolg eines Trainings mit dem Konzept der „natürlichen Häufigkeiten“ von Sedlmeier und Gigerenzer bei Studenten der Universität legte nahe, vor einer Realisierung im Unterricht (vgl. Kap.IV) den Einfluss auf entsprechende Kompetenzen zunächst auch mit Schülern im Labor zu untersuchen.

Im Folgenden behandle ich eigene Trainingsstudien mit *Schülern der gymnasialen Sekundarstufe I und II* auf Grundlage der Trainingsstudien von Sedlmeier & Gigerenzer

⁷⁴ Es sei darauf hingewiesen, dass die Studien von Bea nicht explizit das rechnerische Modellieren und Lösen von Bayesproblemen testeten, so dass ein Vergleich der Studien von Bea und Sedlmeier nur sehr begrenzt möglich ist.

⁷⁵ Im Folgenden verwende ich vereinfachend für Gruppen von Schülerinnen und Schülern die maskuline Form „Schüler“, für Lehrerinnen und Lehrer „Lehrende“.

(2001; Sedlmeier 1999, 1997). Untersuchungsgegenstand sind Trainingswirkungen auf Grundkompetenzen im Bereich „Bedingte Wahrscheinlichkeit und Bayesanwendungen“ („Bayesianisches Denken“).

3.1 Trainingsbeschreibung

Häufigkeits- und Regeltraining wurden mit dem Tutorsystem „BasicBayes“ (Sedlmeier, 1997) durchgeführt, wie in III.2.1. beschrieben. Es wurden 3 Trainingsvarianten verglichen: Training mit dem Häufigkeitsbaum (*FREQ*), Regeltraining im Wahrscheinlichkeitsformat (*PROB*). In einer dritten Variante wurde zunächst eine häufigkeitsbasierte Einführung mit dem Baumdiagramm und anschließend auch die Herleitung und Verwendung der Wahrscheinlichkeitsregel gezeigt. Die Lernenden trainierten mit beiden Modellierungsformen Aufgaben zu lösen (*COMBI*).

3.2 Methodisches Vorgehen

3.2.1 Stichproben

Die untersuchte Stichprobe aus der Sekundarstufe I bestand aus Gymnasialschülern der 10. Klasse ($N = 80$, 35 männlich / 45 weiblich; Verteilung auf Trainingsgruppen: $n_{\text{prob}} = 27$, $n_{\text{freq}} = 27$, $n_{\text{combi}} = 26$), die vorher noch keinen Stochastikunterricht hatten. Die Stichprobe aus der Sekundarstufe II bestand aus Schülern der 13. Klasse, die an Mathematik-Leistungskursen teilnahmen ($N = 47$, 31 männlich / 16 weiblich; Verteilung auf Trainingsgruppen: $n_{\text{prob}} = 23$, $n_{\text{freq}} = 24$), in denen das Themengebiet „Bedingte Wahrscheinlichkeit und Satz von Bayes“ jeweils unmittelbar vorher im Unterricht behandelt worden war.⁷⁶ Die Schüler stammten aus verschiedenen Gymnasien aus den Ländern Berlin und Brandenburg und wurden den Gruppen zufällig zugewiesen. Kontrolliert wurde vorab, dass alle Gruppen hinsichtlich geschlechtsbezogener und allgemeiner leistungsbezogener Merkmale (bisherige Schulleistung in Mathematik) weitgehend parallel waren.

3.2.2 Erhebungsinstrumente

Leistungstests bei Sedlmeier und Gigerenzer fokussierten auf einen Aufgabentyp „dichotomer“ Bayesprobleme, bei dem überwiegend prozedurales Denken erforderlich ist („rechnerisches Modellieren“). Bei der Konstruktion des eigenen *Kompetenztests zu Bayesianischem Denken* wurde eine Differenzierung von Aufgabenklassen und Anforderungen angestrebt. Die Testkonstruktion basierte auf einem Itempool, der verschiedene „Itemtypen“ umfasste, die im Folgenden genauer diskutiert werden sollen (Tabelle III.1 gibt einen Überblick).

Bezüglich der Art der mathematischen Tätigkeit erforderten alle Aufgaben außermathematische Modellierung, d.h. ein Überführen einer mehr oder weniger realistischen Situation in ein Modell. Sie stammten weitgehend aus vorangegangenen Studien bzw. aus bekannten Anwendungsproblemen. Die Testitems waren in allen Trainingsgruppen völlig identisch, außer einem Hinweis zur Größenordnung der Grundgesamtheit bei Häufigkeitsgruppen (z.B. „Betrachten Sie 1000 Fälle“).

Die schriftlichen Tests mit weitgehend offenen Antwortformaten ermöglichten die Bewertung und den Vergleich der Modellierungstätigkeit als Kern mathematischen Denkens der Schüler. Alle Überlegungen, Darstellungen, Begründungen oder Argumentationen sollten

⁷⁶ Es erfolgten durchschnittlich 8 Unterrichtsstunden. Im Unterricht wurde generell nach dem Schulbuch von Bigalke & Köhler (1997) vorgegangen (Berlin und Brandenburg).

auf die Testbögen geschrieben werden. Mehrfache Aufforderung und Überprüfung durch die Testleiter sowie Belohnung der Versuchsschüler konnten eine möglichst umfassende Dokumentation der Modellierungstätigkeiten ermöglichen. Die detaillierte Untersuchung der Modellierungs- und Verarbeitungsprozesse diente als Grundlage für quantitative und qualitative Bewertungen.

Folgende Tabelle III.1 gibt einen Überblick über die im Kompetenztest vorkommenden Itemtypen. Angegeben ist jeweils eine Anforderungsbeschreibung und ein Beispielitem.

ITEMTYP	Beschreibung / Beispiel
Itemtyp A0	<p><i>Beschreibung:</i> Es soll angegeben werden, ob gegebene Informationen aus dem Aufgabentext für eine mathematische Lösung benutzt werden müssen.</p> <p><i>Beispielitem (bezieht sich jeweils auf ein Item vom Typ A2):</i> Brauchen Sie für die Antwort eigentlich die Information A unbedingt? <input type="radio"/> ja <input type="radio"/> nein Begründen Sie ihre Entscheidung: _____ (<i>Antwortformat:</i> multiple choice + offen)</p>
Itemtyp A0+	<p><i>Beschreibung:</i> Es soll angegeben werden, ob und wie die Änderung gegebener Informationen das Ergebnis beeinflussen kann.</p> <p><i>Beispielitem (bezieht sich jeweils auf ein Item vom Typ A2):</i> Ändert sich etwas, wenn die Prozentangabe in Information A kleiner wird (z.B. 5%)? <input type="radio"/> nein <input type="radio"/> ja, das Ergebnis wird größer <input type="radio"/> ja, das Ergebnis wird kleiner Begründen Sie ihre Entscheidung: _____ (<i>Antwortformat:</i> multiple choice + offen)</p>
Itemtyp A1	<p><i>Beschreibung:</i> Es soll angegeben werden, ob mit den gegebenen Informationen ein Problem im Bereich "totale Wahrscheinlichkeit" (außermath. Kontext) gelöst werden kann (A1a). Wenn mit „ja“ geantwortet wird, soll es gelöst werden (A1b).</p> <p><i>Beispielitem (bezieht sich jeweils auf ein Item vom Typ A2):</i> Kann man mit den gegebenen Informationen beantworten, wie hoch die Wahrscheinlichkeit ist, dass ein in die Ambulanz kommender Patient an Fieber, Schüttelfrost und Hautabschürfungen leidet? <input type="radio"/> ja, und zwar _____ <input type="radio"/> das kann man mit den obigen Informationen nicht eindeutig beantworten Begründen Sie ihre Entscheidung und geben Sie Überlegungen an: _____ (<i>Antwortformat:</i> multiple choice + numerisch + offen)</p>
Itemtyp A2	<p><i>Beschreibung:</i> Aus den gegebenen Informationen soll ein kontextbezogenes dichotomes Bayesproblem gelöst werden (nur dieser Itemtyp war expliziter Gegenstand der Trainings).</p> <p><i>Beispielitem (BlutA2):</i> Sie arbeiten in der Ambulanz eines Krankenhauses. Ein Patient mit hohem Fieber und Schüttelfrost kommt herein, bei dem Sie auch Hautabschürfungen bemerken. Folgende Informationen aus der Krankenstatistik haben Sie zur Verfügung: A Durchschnittlich 10% der in die Ambulanz kommenden Patienten haben Blutvergiftung. B Wenn ein Patient Blutvergiftung hat, besteht eine Wahrscheinlichkeit von 80%, dass er an diesen Symptomen (Fieber, Schüttelfrost, Hautabschürfungen) leidet. C Wenn ein Patient keine Blutvergiftung hat, besteht trotzdem eine Wahrscheinlichkeit von 10%, dass er an diesen Symptomen leidet. Die Frage lautet nun: Wie groß ist die Wahrscheinlichkeit, dass dieser Patient Blutvergiftung hat, wenn er diese Symptome aufweist? Antwort: _____ Geben Sie Ihre Überlegungen ausführlich an: _____ (<i>Antwortformat:</i> numerisch + offen)</p>

<p><i>Itemtyp</i> A3</p>	<p><i>Beschreibung:</i> Aus den gegebenen Informationen soll ein kontextbezogenes Bayesproblem gelöst werden, das in der Struktur komplizierter als im Training ist. Anstatt der üblichen dichotomen Struktur, kommen trichotome Ereignisse vor.</p> <p><i>Beispielitem (WetterA3):</i> Das Wetter in Nürnberg (Nordbayern) ist an 20% aller Tage regnerisch, an 50% wechselhaft und an 30% heiter. Es hat sich herausgestellt, dass am Vorabend eines Tages mit heiterem Wetter die Wettervorhersage für Nordbayern zu 80% gut und zu 20% schlecht lautet. Bei wechselhaftem Wetter am nächsten Tag lautete sie zu 50% schlecht und zu 50% gut. Ein Regenwettertag wird zu 60% mit schlecht angekündigt und zu 40% mit gut. Gestern Abend wurde für heute gutes Wetter in Nordbayern vorhergesagt. Mit welcher Wahrscheinlichkeit ist heute heiteres Wetter in Nürnberg? Antwort: _____ Geben Sie Ihre Überlegungen ausführlich an: _____ (<i>Antwortformat:</i> numerisch + offen)</p>
<p><i>Itemtyp</i> A4</p>	<p><i>Beschreibung:</i> Aus den gegebenen Informationen soll ein kontextbezogenes Bayesproblem gelöst werden, das im Unterschied zu Trainingsaufgaben einen komplexeren Kontext mit extrem seltenen bzw. sicheren Ereignissen enthält.</p> <p><i>Beispielitem (DnaA4):</i> Im Rahmen der Untersuchungen zu einem Mord im Raum Berlin wurde eine DNA-Analyse von potentiell Tatverdächtigen durchgeführt. Dabei wurde eine am Tatort gefundene Spur des Täters untersucht und es konnte ein DNA-Profil der Spur erstellt werden. Es wurde schließlich eine Übereinstimmung des DNA-Profiles eines Untersuchten mit dem DNA-Profil der Spur festgestellt. Der Mann wurde angeklagt und es gab Stimmen, dass dieser Mann aufgrund der Übereinstimmung mit "hundertprozentiger" Sicherheit auch der Täter sei. Der Verteidiger wollte das nicht einfach so hinnehmen, schließlich ging es für seinen Mandanten um eine lebenslange Freiheitsstrafe, und er hinterfragte die Genauigkeit der DNA-Analyse. Versetzen wir uns in die Rolle des Verteidigers des Angeklagten und betrachten wir die DNA-Analyse genauer. Ein Sachverständiger erstellt ein Gutachten, das folgende Informationen enthält: A Ein im Rahmen des DNA-Tests untersuchter Mann ist ohne weitere Einschränkung durch Indizien zu 0,0001% der Spurverursacher. B Wenn der Verdächtige wirklich der Spurverursacher ist, dann können wir davon ausgehen, dass es der durchgeführte DNA-Test mit Sicherheit (also zu 100%) anzeigt. C Die Wahrscheinlichkeit, dass das DNA-Profil eines Unschuldigen rein zufällig mit dem DNA-Profil der Spur übereinstimmt, beträgt 0,0001%. Sie wollen als Verteidiger nun wissen, mit welcher Wahrscheinlichkeit der Angeklagte der Verursacher der DNA-Spur ist, wenn der DNA-Test Übereinstimmung zeigte. Antwort: _____ Geben Sie Ihre Überlegungen ausführlich an: _____ (<i>Antwortformat:</i> numerisch + offen)</p>

Tab.III.1: Itemtypen des Kompetenztests zu Bayesianischem Denken

Ich möchte versuchen, eine Abgrenzung der Itemtypen hinsichtlich einiger Klassifikationskriterien für mathematische Aufgaben zu geben.⁷⁷ Die Itemtypen A0, A0+ und A1a erfordern vorwiegend begriffliches Modellieren in Form des Verstehens und Beurteilens von Modellen bzw. (funktionalen) Zusammenhängen mit Argumentieren auf mittlerem Niveau. Die Itemtypen A1b, A2, A3 und A4 verlangen vorwiegend rechnerisches Modellieren. Technische Anforderungen (wie z.B. Termumformungen, Bruchrechnung) sind jeweils mit unterschiedlicher Schwierigkeit in A1b bis A4 enthalten; jedoch ist keine Aufgabe rein technischer Natur (im Sinne von rein innermathematischer Modellierung; Neubrand, Klieme et al., 2002, S.103).

Es ist zu beachten, dass *nur* der Itemtyp A2, wie oben beschrieben, expliziter Inhalt der Trainings war, d.h. das Training betraf vorwiegend das rechnerische Modellieren von dichotomen Bayesproblemen. Diese Grundaufgabe wurde auch nach dem Klassifikations-

⁷⁷ Kriterien und Kategorien sind z.B. in Neubrand M., Klieme et al. (2002); Neubrand J. (2002); Ross, Jordan, Baumert, Blum et al. (im Druck) theoretisch verankert beschrieben bzw. operationalisiert.

schema zur Analyse von Mathematikaufgaben von Ross, Jordan, Baumert, Blum et al. (im Druck) genauer bewertet, das im Rahmen des Projektes COACTIV⁷⁸ entwickelt wurde. Das „Rating“ erfolgte durch eine Expertenlehrerin, die umfassende Schulung hierzu erhalten hatte. Die Ergebnisse der Bewertung sind im *Anhang A* genauer ausgeführt.

Die notwendigen Lösungsschritte bei Itemtyp A1b stellen als „Berechnung einer totalen Wahrscheinlichkeit“ eine Teilaufgabe der A2-Items dar. Insofern ist A1b hinsichtlich des Umfangs der Bearbeitung niedriger als A2 einzustufen.

Der Itemtyp A3 ist gewissermaßen eine strukturell komplexere, rechnerische Bayesianische Modellierungsaufgabe, bei der über Anwendungswissen hinaus auch Problemlösekompetenz erforderlich ist, da eine *strukturell* neue Situation modelliert werden muss (vgl. Stern & Hardy, 2001, S.168f.). A3-Items erfordern insbesondere im Vergleich zu A2-Items einen höheren Bearbeitungsumfang, da mehr Informationen (6 anstatt 3) bei der Modellierung verarbeitet werden müssen⁷⁹. Bei A3-Items sind außerdem gewissermaßen „irrelevante Zahlen“ in der Aufgabenstellung enthalten, etwa im Wetterbeispiel (vgl. Tab.III.1) werden die Angaben zur Vorhersage „schlechtes Wetter“ zur Bearbeitung nicht benötigt.

Bei A4-Items ist der Bearbeitungsumfang gegenüber A2 zwar nicht erhöht. Allerdings müssen bei A4-Aufgaben „extreme“ Wahrscheinlichkeitssituationen modelliert werden (Wahrscheinlichkeit von 100% bzw. 0,0001%). Die A4-Items weisen zudem ein höheres Niveau der sprachlogischen Komplexität auf, bedingt durch eine authentischere Darstellung der Situation (vgl. Cohors-Fresenborg, 1996).

Die genannten Merkmale können als Indikatoren dafür dienen, dass Itemtypen A3 bzw. A4 insgesamt anspruchsvoller als A2-Items einzustufen sind. Die Schwierigkeitsbeurteilung der Itemtypen wird durch eine ex-post-Analyse der Leistungen bei Wahrscheinlichkeitsmodellierung (10./13.Klasse) unterstützt. Eine deutliche Heterogenität der Schwierigkeit ist festzustellen, wobei A3-Items tendenziell am schwersten einzustufen sind (Tab.III.2). Die subjektive ex-post-Bewertung der Schüler der 10. bzw. 13. Klasse deutet ebenfalls auf heterogene Verteilung der Schwierigkeit der Itemtypen mit steigender Tendenz hin.

Schwierigkeit	A0	A0+	A1	A2	A3	A4⁸⁰
in % der erreichten Nachtest-Leistung, 10. / 13.Klasse	52 / 72%	42 / 52%	21 / 43,5%	46 / 64%	6 / 40%	43%
subjektive Bewertung, 10.Klasse	leicht: 72% mittel: 26% schwer: 2%	leicht: 65% mittel: 28% schwer: 7%	leicht: 14% mittel: 70% schwer: 16%	leicht: 8% mittel: 72% schwer: 20%	leicht: 4% mittel: 68% schwer: 28%	-
subjektive Bewertung, Leistungskurs, 13.Klasse	leicht: 90% mittel: 10% schwer: 0%	leicht: 55% mittel: 43% schwer: 2%	leicht: 52% mittel: 32% schwer: 16%	leicht: 28% mittel: 66% schwer: 6%	leicht: 6% mittel: 66% schwer: 28%	leicht: 22% mittel: 39% schwer: 39%

Tab.III.2: Ex-post-Bewertung der Schwierigkeit der Itemtypen (objektiv und subjektiv)

3.2.3 Design

Tab.III.3 gibt einen Überblick über die Maßnahmenfolge in allen Trainingsgruppen. Die fünf Gruppen ergeben sich aus den Trainingsvarianten PROB, FREQ und COMBI (vgl. 3.1.) und der Stichprobe (10.Klasse und 13.Klasse, Leistungskurs). „Kombitraining“ wurde nur bei Zehntklässlern untersucht. Auf eine Kontrollgruppe ohne Training wurde verzichtet, da

⁷⁸ „Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz“, Projektleiter: J.Baumert, W.Blum & M.Neubrand, Projektorganisation: S.Krauss.

⁷⁹ Dadurch steigt die Anzahl von „kleinen Lösungsschritten“ (als Maß des Bearbeitungsumfangs) natürlicherweise an.

⁸⁰ A4-Items wurden bei der Studie mit Zehntklässlern nicht verwendet, da die sprachliche Komplexität dieser Aufgabenstellung im Rahmen des Tests zu hoch war.

aufgrund der Ergebnisse vorangegangener Studien substantielle Effekte durch Training (im Vergleich zu keinem Training) angenommen werden können.

Alle Trainingsgruppen erhielten zunächst einen Vortest, um das Grundniveau festzustellen. Der Vortest setzte sich aus jeweils einem Item des Typs A0, A0+, A1 und A2 zusammen. Itemtypen A3 und A4 waren im Vortest nicht enthalten; mit ihnen wurden Transferleistungen („strukturelle Verallgemeinerung“) nach Training geprüft.

Anschließend erfolgte das Training in der jeweiligen Trainingsvariante. Der Lernerfolg wurde unmittelbar nach dem Training mit dem kompletten Kompetenztest (Nachttest 1) gemessen. Dieser Nachttest enthielt drei Items vom Typ A2 (jeweils mit A0, A0+, A1; vgl. Tab.III.1), ein Item vom Typ A3 und ein Item vom Typ A4 (nur bei LK-Schülern). Die kontextuelle Einkleidung der Aufgabenitems in Vortest, Training und Nachttest war immer unterschiedlich. Alle Testitems erforderten also über Reproduktion hinausgehendes „Anwendungswissen“ in dem Sinne, dass keiner der Aufgabenkontexte, die die Schüler während Vortest- oder Trainingsphase bearbeiteten, identisch zu einem Aufgabenkontext im Nachttest war. Nach ca. 12 Wochen wurde der Leistungstest in derselben Weise wiederholt (Nachttest 2), um die Nachhaltigkeit des Lernerfolges zu prüfen.

Gruppe (Stichprobe)	Vortest	Training	Nachttest 1	Nachttest 2
Prob LK (Sek. II, LK)	unmittelbar vor Training	PROB	unmittelbar nach Training	ca. 12 Wochen nach Training
Freq LK (Sek. II, LK)		FREQ		
Prob 10 (Sek. I, 10.Kl.)		PROB		
Freq 10 (Sek. I, 10.Kl.)		FREQ		
Combi 10 (Sek. I, 10.Kl.)		COMBI		

Tab.III.3: Design der Studie mit Schülern

3.3 Ergebnisse

3.3.1 Analyse der Modellierungsprozesse und Operationalisierung

Ein Bewerten individueller Verarbeitungsprozesse aus den numerischen Ergebnissen ist oft nur unzureichend möglich. Eine angemessene Bewertung von *Kompetenz* (in einem höheren Sinne als „Leistung“) sollte Lösungswege und Argumentationen miteinbeziehen (vgl. Stern & Hardy, 2001, S.164f.). Bei der Auswertung des Kompetenztestes erfolgte deshalb *vor* der Leistungsbewertung eine Analyse und Klassifikation der Modellierungsprozesse. Diese Analyse war Grundlage für die Operationalisierung des *Lernerfolges*. Schritte im Modellierungsprozess sollten identifiziert und klassifiziert werden. Anschließend wurden bei jedem Schüler die schriftlich aufgezeichneten Lösungswege und Argumentationen bei jedem Item überprüft und nach dieser Klassifikation mit Punkten bewertet. Die Testleistung (= Summe der Punkte in allen Items) spiegelt eine differenzierte Bewertung der Kompetenz der Schüler wider. Tab.III.4 (S.80) zeigt die Schritte⁸¹ für einen „idealen“ Modellierungsprozess dichotomer Bayesprobleme (Itemtyp A2) differenziert nach Modellierungsform (links: Wahrscheinlichkeitsregel, rechts: Häufigkeitsbaum; vgl. auch Kap.II.2.1 und 3.1).

⁸¹ Man beachte, dass die Bezeichnung der vier Schritte speziell für diese Arbeit erfolgte. Im gebräuchlichen Prozessmodell des mathematischen Modellierens (z.B. Schupp, 1984; Blum, 1996; Klieme et al., 2001) entsprechen diese Schritte etwa den beiden ersten „Übergängen“ von der Situation (Welt) zum Modell (Mathematik) und der „innermathematischen Weiterverarbeitung“ des Modells bis zur wahrscheinlichkeitstheoretischen Lösung. Die Interpretation der Lösungen war nicht Gegenstand dieses Tests.

Die Modellierungsschritte lassen sich bei Itemtypen A3 und A4 analog zu A2 klassifizieren, wobei die Schritte 1 bis 3 jeweils entsprechend modifiziert sind. Wie bereits angedeutet, ist die Modellierung des Itemtyps A1 ein Teilprozess der A2-Items. Schritt 1 und 2 sind analog. In Schritt 3 müssen folgende Lösungsterme gebildet werden („totale Wahrscheinlichkeit“):

$$A \cdot C + B \cdot (1 - C) \quad \text{bzw.} \quad A + B / \text{„Gesamtpopulation“}$$

Bei den Itemtypen mit überwiegend begrifflicher Modellierung war Argumentieren verlangt. Durch bewertbare Argumentationsschritte der Schüler wurde in ähnlicher Weise „konzeptuelles“ Denken genauer bewertet.

ITEM „PILOT“ (Typ A2, Antwortformat offen):	
<p>Ein Pilot bemerkt während eines Fluges Rauch, der aus dem Motor aufsteigt. Er muss sich entscheiden, ob er den Flug vorzeitig abbricht oder ob er weiterfliegt, je nachdem wie wahrscheinlich es ist, dass der Motor aussetzt. Er verfügt über die folgenden Informationen aus der Flugstatistik: A Wenn der Motor aussetzt, deutet in 90% der Fälle aufsteigender Rauch vorher darauf hin. B Während 10% der Flüge, bei denen der Motor nicht aussetzt, kommt Rauch aus dem Motor. C Der Motor seines Flugzeugtyps setzt in 1% aller Flüge aus. Die für den Piloten (lebens-)entscheidende Frage ist nun: Wie wahrscheinlich ist es, dass der Motor aussetzt, wenn Rauch von ihm aufsteigt?</p>	
Modellierung mit Wahrscheinlichkeitsregeln:	Modellierung mit Häufigkeitsbaum:
<p>Schritt 1 „Mathematisieren (i.e.S.)“ (25% der Punkte pro Item)</p> <ul style="list-style-type: none"> • Herausfiltern der relevanten Informationen aus dem Aufgabentext • Übersetzung der Situation in eine „verarbeitungsfähige“ Form • Welche Informationen sind relevant und was bedeuten sie? In welcher Form können/sollen sie mathematisch bearbeitet werden? 	
<p>A) $P(\text{Rauch} \mid \text{Motor aus}) = 90\% = 0,9$ B) $P(\text{Rauch} \mid \text{Motor nicht aus}) = 10\% = 0,1$ C) $P(\text{Motor aus}) = 1\% = 0,01$</p>	<p>C) Bei 1% von 1000 = 10 Flügen setzt der Motor im Durchschnitt aus A) Von diesen 10 kommt bei 90% = 9 Rauch aus dem Motor. B) Von den restlichen 990 kommt bei 10% = 99 auch Rauch aus dem Motor.</p>
<p>Schritt 2 „Strukturieren“ (25% der Punkte pro Item)</p> <ul style="list-style-type: none"> • Weiterverarbeitung der übersetzten Informationen • Wie lassen sich die Informationen verwerten und in Beziehung setzen? Wie kommt man zum Modell? 	
<p>Mit Produktregel / Additionsregel bei dichotomen Ereignissen: $P(\text{Rauch und Motor aus}) = A \cdot C$ $P(\text{Rauch und Motor nicht aus}) = B \cdot (1 - C)$, wobei $P(\text{Motor nicht aus}) = 1 - C$</p>	<p>Bildung der Baumstruktur und einsetzen der (übersetzten) Häufigkeitswerte Ergänzung der fehlenden Baumhäufigkeiten (Summenbildung)</p>
<p>Schritt 3 „Verknüpfen“ (25% der Punkte pro Item)</p> <ul style="list-style-type: none"> • Aufstellen eines „Lösungsterms“ (Mathematisches Modell i.e.S.) • Wie kann eine stochastische Lösung erreicht werden? 	
<p>Bayesregel (über die totale Wahrscheinlichkeit): $P(\text{Motor aus} \mid \text{Rauch}) = A \cdot C / [A \cdot C + B \cdot (1 - C)] = 0,9 \cdot 0,01 / [0,9 \cdot 0,01 + 0,1 \cdot (1 - 0,01)]$</p>	<p>Bedingte Wahrscheinlichkeit (frequentistische Interpretation): $P(\text{Motor aus} \mid \text{Rauch}) = A / [A + B] = 9 / 9 + 99$</p>
<p>Schritt 4 „Berechnen“ (25% der Punkte pro Item)</p> <ul style="list-style-type: none"> • Berechnung einer Lösungswahrscheinlichkeit (mit evtl. Umformung numerischer Formate) 	
<p>$P(\text{Motor aus} \mid \text{Rauch}) = 0,9 \cdot 0,01 / [0,9 \cdot 0,01 + 0,1 \cdot (1 - 0,01)] = 0,0833 (= 8,33\%)$</p>	<p>$P(\text{Motor aus} \mid \text{Rauch}) = 9 / 9 + 99 = 0,0833 (= 8,33\%)$</p>

Tab.III.4: Modellierungsschritte bei dichotomen Bayesproblemen (Itemtyp A2)

3.3.2 Leistungen im Kompetenztest

Verglichen werden zunächst deskriptiv trainingsbedingte Veränderungen der Testleistungen pro Gruppe (= Punktsummenwerte im Vortest, Nachtest 1, Nachtest 2). Unterschiede zwischen Trainingsgruppen werden als Effektgrößen⁸² auf Basis von varianzanalytischen Ergebnissen jeweils in Fußnoten angegeben (vgl. Cohen, 1988). Ausdrücklich hingewiesen sei an dieser Stelle auf die umfassende Diskussion zu Möglichkeiten und Grenzen der Verwendung üblicher Signifikanztestmethoden (z.B. Gigerenzer, 1993a; Sedlmeier, 1996) und möglichen Alternativen, wie z.B. die Verwendung von Effektgrößen als Grundlage für Metaanalysen (vgl. Cohen, 1988; 1994) oder explorative Datenanalyse (Tukey, 1977; vgl. eine ausführliche Analyse von Biehler, 1982). Ich möchte diese oft sehr kontrovers geführte Diskussion im Rahmen dieser Arbeit nicht vertiefen. Im Sinne einer Vermeidung „methodischer Rituale“ soll (inferenz-)statistische und datenanalytische Methodik ergänzend nebeneinander verwendet werden (vgl. Box, 1983).

Abb.III.3a zeigt für alle fünf Trainingsgruppen die durchschnittlich erreichten Anteile am maximal erreichbaren Punktsummenwert („mittlere Testleistung“). Im Vortest erreichten die *Zehntklässler* in den drei Versuchsgruppen Prob(10), Freq(10) bzw. Combi(10) mittlere Lösungsraten von 15,1%, 18,5% bzw. 12,6%.⁸³ Im unmittelbaren Nachtest 1 ergaben sich Leistungen von 36,1%, 59,1% bzw. 43,3%.⁸⁴ Im späteren Nachtest 2 zeigten sich deutlich zunehmende Unterschiede zwischen Prob (12,9%) und Freq (44,3%) bzw. Combi (31,1%).⁸⁵ Nach Wahrscheinlichkeitstraining ist ein deutlicher „Verfall“ der Trainingseffekte zu beobachten, während dieser Verfall in den Häufigkeits- und Kombigruppen weniger stark auftritt.

Die *LK-Schüler* erreichten im Vortest in den beiden Versuchsgruppen Prob(LK) 47,8% und Freq(LK) 51,1% der erreichbaren Punktsumme.⁸⁶ Im unmittelbaren Nachtest 1 ergaben sich Leistungen von 51% bzw. 78,9%.⁸⁷ Im späteren Nachtest 2 verringert sich der Lernerfolg bei Prob (21,1%) deutlich stärker als bei Freq (61,8%)⁸⁸.

⁸² Der verwendete Effektgrößenindex η^2 (Populationsvarianz) ist definiert als das Verhältnis der Varianz der Mittelwerte zur Varianz in der Gesamtpopulation („superpopulation“; vgl. Cohen, 1988, S.281). Anschaulich: „Welcher Anteil der Varianz in der Gesamtpopulation kann durch die Gruppenzugehörigkeit erklärt werden? Wertebereich ist [0; 1] (folgt aus Def.). η^2 ist eine Verallgemeinerung des punkt-biserialen r^2 für mehr als 2 Populationen. Bei Cohen (1988, S.285f.) werden Angaben zur Bewertung der Effektgröße für „Index f“ gegeben, wobei $f = \sqrt{\eta^2 / (1 - \eta^2)}$. Demnach ist $f = 0,1$ ($\eta^2 = 0,01$) als ein *kleiner*, $f = 0,25$ ($\eta^2 = 0,059$) als ein *mittlerer* und $f = 0,4$ ($\eta^2 = 0,138$) als ein *großer Effekt* zu bewerten.

⁸³ Varianzanalytisch ergibt sich kein signifikanter Unterschied ($F = 1,864$; $df = 2$; $p = 0,162$; Effektgrößen-schätzer $\eta^2 = 0,039$), obwohl Freq tendenziell am besten im Vortest abschneidet. Unterschiede im Vortest werden in folgenden Analysen jeweils als Kovariate berücksichtigt.

⁸⁴ Eine Kovarianzanalyse (Vortestleistungen als Kovariate) ergibt einen hochsignifikanten Unterschiedseffekt ($F = 6,165$; $df = 2$; $p = 0,003$; $\eta^2 = 0,144$). Bei paarweisen Vergleichen erweist sich nur der Unterschied zwischen Freq und Prob als hochsignifikant ($p = 0,003$), die Unterschiede zwischen Freq und Combi bzw. Prob und Combi als nicht signifikant.

⁸⁵ Eine Kovarianzanalyse ergibt einen hochsignifikanten Unterschiedseffekt ($F = 19,867$; $df = 2$; $p < 0,001$; $\eta^2 = 0,357$). Bei paarweisen Vergleichen erweist sich der Unterschied zwischen Freq und Prob als hochsignifikant ($p < 0,001$) sowie der Unterschied zwischen Combi und Prob ($p < 0,001$). Der Unterschied zwischen Combi und Freq ist nicht signifikant ($p = 0,183$).

⁸⁶ Eine Varianzanalyse ergibt keinen signifikanten Unterschiedseffekt ($F = 0,140$; $df = 1$; $p = 0,71$; $\eta^2 = 0,0031$), obwohl Freq tendenziell besser abschneidet.

⁸⁷ Eine Kovarianzanalyse (Vortestleistungen als Kovariate) ergibt einen hochsignifikanten Unterschiedseffekt ($F = 26,459$; $df = 1$; $p < 0,001$; $\eta^2 = 0,376$).

⁸⁸ Eine Kovarianzanalyse (Vortestleistungen als Kovariate) ergibt einen hochsignifikanten Unterschiedseffekt ($F = 141,698$; $df = 1$; $p < 0,001$; $\eta^2 = 0,762$).

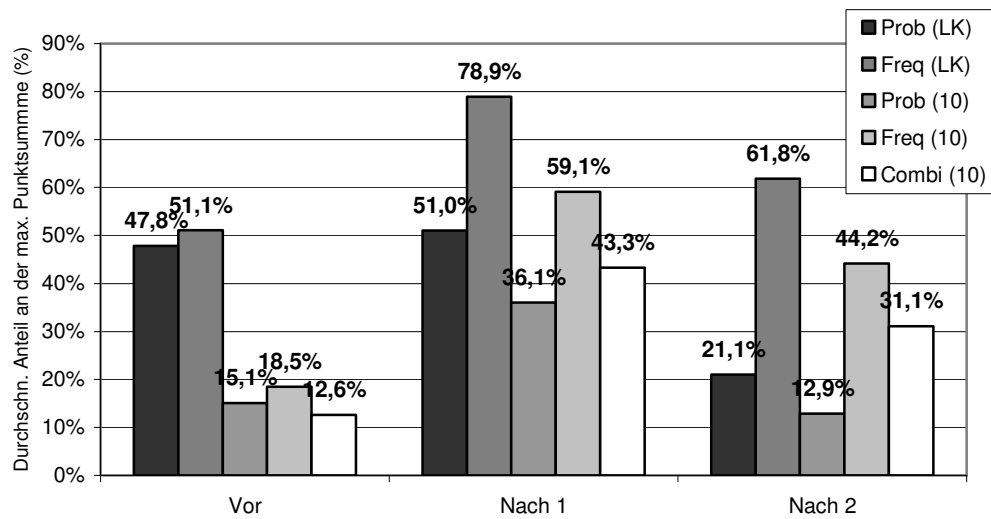
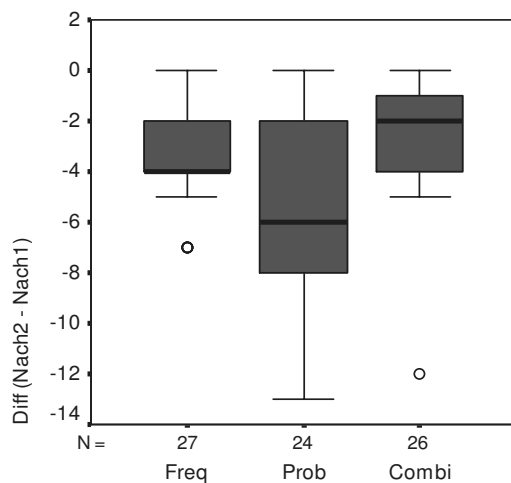
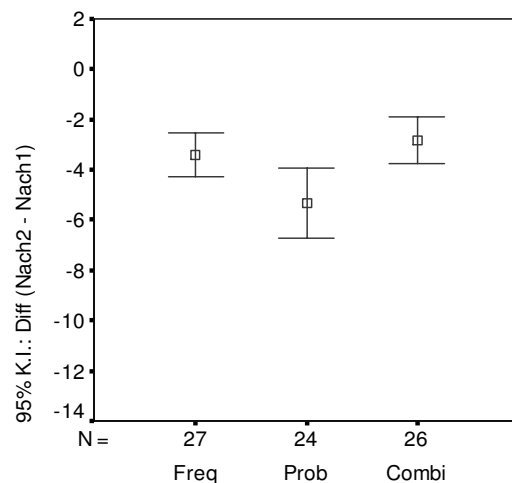


Abb.III.3a: Testleistungen im Gruppendurchschnitt, gesamter Test

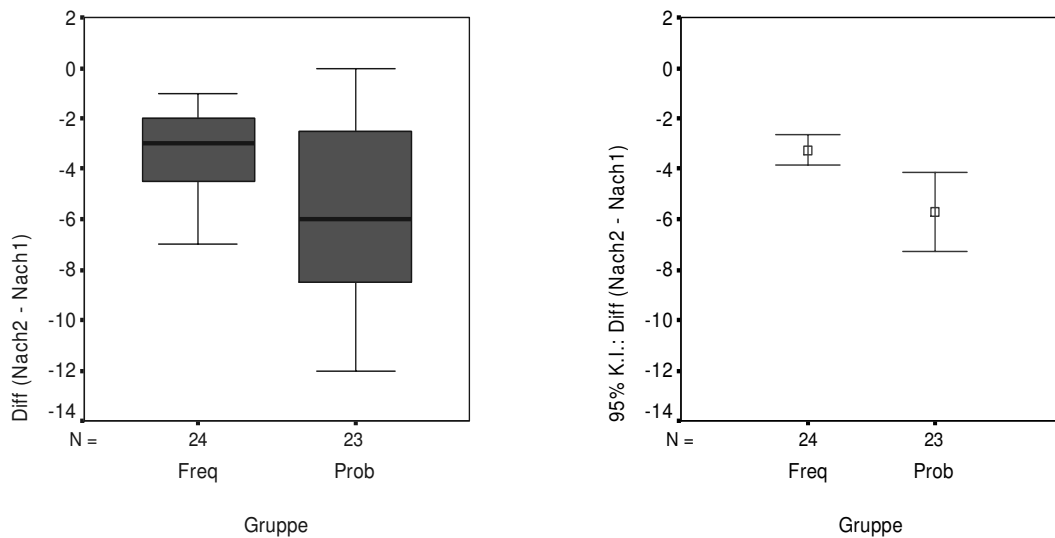
Der Zusammenhang der unmittelbaren zur längerfristigen Nachtestleistung (Nach1 – Nach2), d.h. die Frage welche Nachhaltigkeit der Lernerfolg mit der jeweiligen Trainingsmethode aufweist, soll genauer analysiert werden. Der in den verschiedenen Trainingsgruppen unterschiedliche Rückgang der Leistungen wird in den Verteilungen der (absoluten) Abnahme der Punktsomme zwischen Nachtest 1 und 2 deutlich (vgl. Abb.III.3b, links). In den Prob-Gruppen (10. bzw. 13.Klasse, LK) sind der Median, das Maximum und Spannweite dieser Rückgänge deutlich höher als in den anderen Gruppen, wie die Boxplots anschaulich machen. Etwa 50% der Schüler in den Prob-Gruppen zeigen höhere Leistungsrückgänge als alle Schüler in den anderen Gruppen (ohne „Ausreißer“). Diese Unterschiede in der Stabilität der Lernerfolge werden auch jeweils im Vergleich der Konfidenzintervalle (95%) um die mittlere Abnahme der Punktsommen deutlich (Abb.III.3b, rechts).



Differenzverteilung Punktsomme Nach 2 – Nach 1 (10.Klasse)



95%-Konfidenzintervalle um mittlere Differenzen Punktsomme Nach 2 – Nach 1 (10.Klasse)



Differenzverteilung Punktsomme
Nach 2 – Nach 1 (Leistungskurs, 13.Klasse)

95%-Konfidenzintervalle um mittlere Differenzen
Punktsomme Nach 2 – Nach 1 (13.Klasse)

Abb.III.3b: Abnahme der Testleistungen zwischen Nachtest 1 und Nachtest 2

3.3.3 Leistungen differenziert nach Itemtypen bzw. Art der Modellierung

Im Folgenden werden die Testleistungen nach Itemtypen bzw. der vorherrschenden Art des mathematischen Arbeitens (überwiegend rechnerische oder begriffliche Modellierung) differenziert betrachtet.

Abb.III.4 zeigt den Leistungsvergleich bei dichotomen Bayesproblemen (Itemtyp A2), die als einzige expliziter Trainingsinhalt waren. Die Zehntklässler zeigten hier kaum bewertbares Vorwissen, während die LK-Schüler mit Vorbildung bei etwa 40% lagen. Die Unterschiede in den Trainingseffekten in Nachtest 1 und in der Stabilität der Trainingseffekte (Nach 2) sind qualitativ mit dem Gesamttestergebnis vergleichbar. Jedoch fällt der Unterschied zwischen Prob und Freq im Nachtest 1 bei den Zehntklässlern (17,4 Prozentpkt.) kleiner aus als im Gesamttest (23 Prozentpkt.).⁸⁹

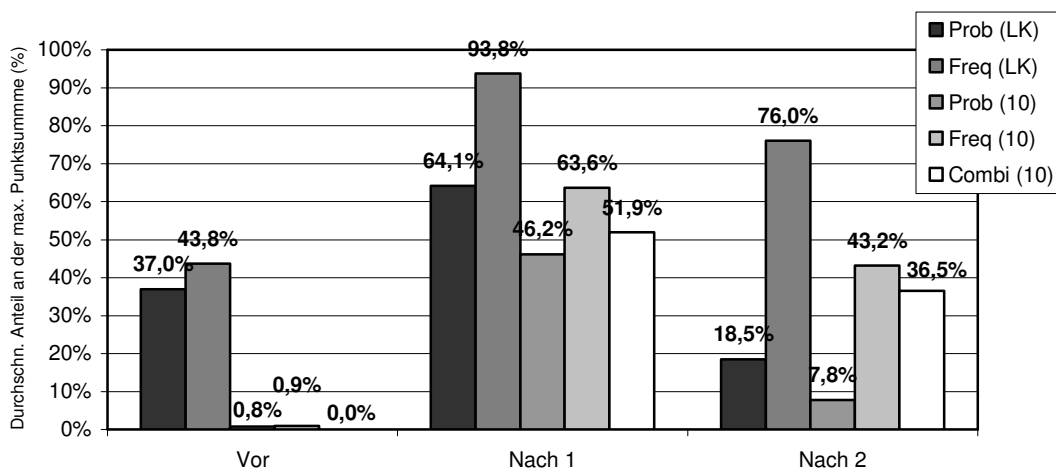


Abb.III.4: Testleistungen im Gruppendurchschnitt, nur Itemtyp A2

⁸⁹ $\eta^2 = 0,063$ (Kovarianzanalyse: $F = 3,214$; $df = 1$; $p = 0,079$) im Vgl. zu $\eta^2 = 0,144$ gesamt.

Die Vortestleistungen beim A1-Itemtyp (Modellierung eines Problems zur totalen Wahrscheinlichkeit, vgl. Abb.III.5) waren bei den *Zehntklässlern* in der Freq-Gruppe höher als in den beiden anderen Gruppen⁹⁰. Es ist anzumerken, dass bei diesem Itemtyp ein Teil der Beantwortung (Teilitems A1a) im Multiple-choice-Format erfolgte. In wenigen Fällen wurden Antwortalternativen ohne Argumentation angekreuzt, so dass nicht möglich war, zufälliges von bewusstem Ankreuzen zu unterscheiden. Es wurden dennoch jeweils Punkte für richtiges Ankreuzen vergeben. Es ist aber davon auszugehen, dass in allen Gruppen in etwa gleich häufig „Zufallstreffer“ bepunktet wurden. Insgesamt wird somit die Qualität eines Leistungsvergleiches nicht beeinträchtigt. Im Vergleich ist erkennbar, dass Kombi- und Prob-Gruppe (Zehntklässler) sich kaum unterscheiden und nach nur schwacher Leistungssteigerung unmittelbar nach Training wieder auf Grundniveau nach 12 Wochen zurückkehren. Der Leistungsunterschied der Freq-Trainingsgruppe zu den übrigen vergrößerte sich unmittelbar nach Training und nach 12 Wochen deutlich.⁹¹ Bei den *LK-Schülern* waren die Unterschiede zwischen Freq und Prob in den Nachtests noch ausgeprägter als im Gesamttest (bis über 50 Prozentpunkte in Nach 2).

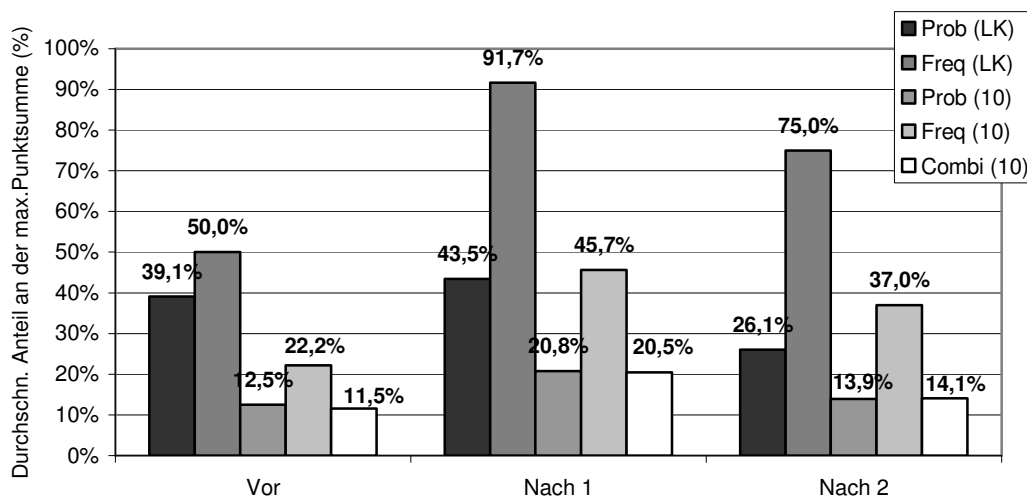


Abb.III.5: Testleistungen im Gruppendurchschnitt, nur Itemtyp A1

Bei „Transferproblemen“ des Itemtyps A3, die abweichend von den trainierten Problemen komplexere Struktur aufwiesen, zeigten sich in den Nachtests besonders unterschiedliche Leistungen. Es ergab sich das qualitative Bild des Gesamttests hinsichtlich der Unterschiede zwischen den Gruppen (vgl. Abb.III.6), die Prob-Gruppen schnitten vergleichsweise jedoch noch deutlich schlechter als bei Items des Typs A2 ab: Zehntklässler kamen kaum mit den A3-Problemen zurecht. Auch LK-Schüler fielen nach 12 Wochen unter 10% Lösungsrate zurück. Häufigkeitsgruppen und auch die Kombigruppe zeigten deutlich höhere und auch stabilere Leistungen bei diesem Itemtyp.

⁹⁰ Der Unterschied ist varianzanalytisch nicht signifikant: $F = 2,08$; $p = 0,132$; $\eta^2 = 0,053$. Die Unterschiede wurden in folgenden Analysen als Kovariate berücksichtigt.

⁹¹ Kovarianzanalyse bei Nach 1 ergibt folgendes: $F = 2,913$; $p = 0,061$; $\eta^2 = 0,074$; bei Nach 2 wird der Unterschied deutlich signifikant: $F = 5,314$; $p = 0,007$; $\eta^2 = 0,127$.

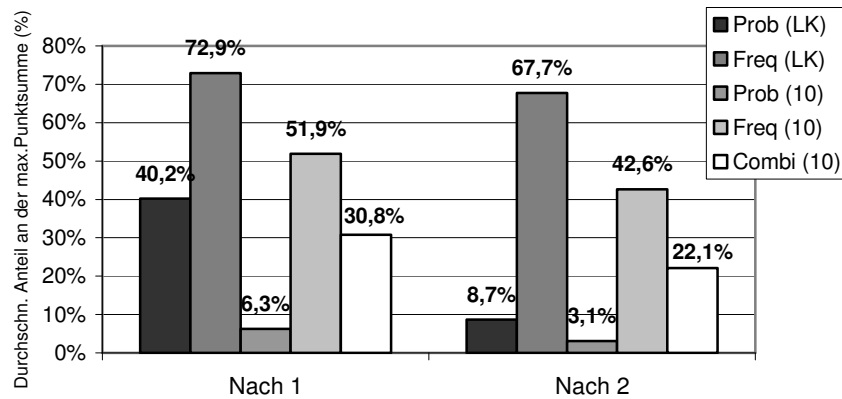
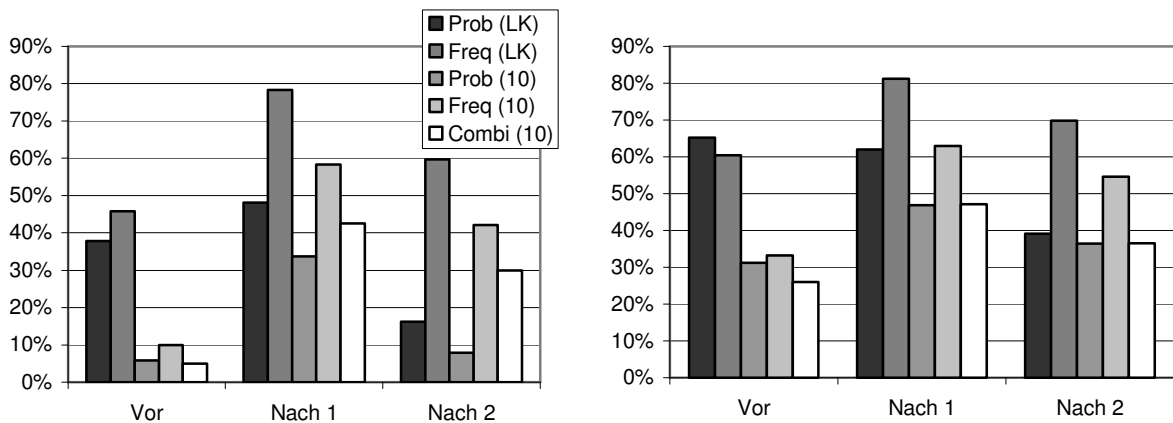


Abb.III.6: Testleistungen im Gruppendurchschnitt, nur Itemtyp A3

Eine weitere Differenzierung der Testleistungen nach Itemgruppen mit überwiegend rechnerischer (A0, A0+, A1a) bzw. begrifflicher Modellierung (A1b, A2, A3, A4) ergibt folgendes Ergebnis (vgl. Abb.III.7).



rechnerisch	Vor	Nach 1	Nach 2	begrifflich	Vor	Nach 1	Nach 2
Prob (LK)	37,9%	48,1%	16,2%	Prob (LK)	65,2%	62,0%	39,1%
Freq (LK)	45,8%	78,3%	59,7%	Freq (LK)	60,4%	81,3%	69,8%
Prob (10)	5,8%	33,8%	7,9%	Prob (10)	31,3%	46,9%	36,5%
Freq (10)	10,0%	58,3%	42,1%	Freq (10)	33,3%	63,0%	54,6%
Combi (10)	5,0%	42,5%	30,0%	Combi (10)	26,0%	47,1%	36,5%

Abb.III.7: Vergleich zwischen Testleistungen bei überwiegend rechnerischer bzw. begrifflicher Modellierung.

Insgesamt sind die Leistungen bei begrifflichen Modellierungsaufgaben höher. Es ist wieder zu beachten, dass aufgrund des Antwortformates „Multiple choice“ in den Fällen ohne die Angabe von Argumentationen ein Teil der Leistung als „Zufallstreffer“ begründbar wären.⁹² Unterschiede in den Gruppenleistungen sind bei beiden Itemgruppen qualitativ ähnlich, d.h. auch bei überwiegend begrifflicher Modellierung zeigten Häufigkeitsgruppen (Freq) nach Training die deutlich höheren Leistungen. Die Leistungsunterschiede zwischen Freq- und Prob-Gruppen fielen jedoch in den Nachtests geringer aus.⁹³ Auffällig ist, dass sich bei überwiegend begrifflicher Modellierung die Leistungen von Kombi-Gruppe und Prob-Gruppe

⁹² Der Erwartungswert bei völlig zufälligem Ankreuzverhalten und ohne erkennbare Argumentation liegt für diese Itemgruppe bei 21% der erreichbaren Punkte (gestrichelte Linie).

⁹³ Ausgedrückt im Effektgrößenindex η^2 : 10.Klasse: Nach 1: 0,075, Nach 2: 0,099; LK: Nach 1: 0,272, Nach 2: 0,416.

kaum unterscheiden, während bei vorwiegend rechnerischer Modellierung Kombi-Training v.a. längerfristig eindeutig höheren Lernerfolg hervorrief.

3.3.4 Verwendung von Modellierungsformen⁹⁴

Welche Darstellungsformen und Verarbeitungsformate werden tatsächlich von den Schülern im Modellierungsprozess verwendet? Keine explizite Form, also z.B. rein formal, verbal oder mit grafischer Darstellung, mit Häufigkeiten oder Wahrscheinlichkeiten, war in den Aufgabenstellungen zwingend gefordert. Einer Interpretation bei Bea (1995, S.161) folgend, können Verwendungsraten als Maß für die *kognitive Zugänglichkeit* einer Modellierungsform dienen, d.h. die Bereitschaft eine trainierte Modellierung als adäquat für die Problemlösung zu akzeptieren und deshalb tatsächlich einzusetzen.⁹⁵

Es ist auch anzunehmen, dass individuell bevorzugte Formen der Modellierung sehr stark von einem vorherrschenden *Denkstil* des Lernenden („preferred way of thinking“, vgl. Sternberg, 1997, S.8) abhängen. Der mathematische Denkstil wird dabei als die vom Individuum bevorzugte Art und Weise charakterisiert, mathematische Sachverhalte durch gewisse interne Vorstellungen und/oder externalisierte Darstellungen zu repräsentieren und durch gewisse Vorgehensweisen zu verarbeiten. So werden visuelle, analytische und konzeptuelle Denkweisen unterschieden, wobei diese oft nicht in reiner Form auftreten. Auch interne Vorstellung und externalisierte Darstellung können sich in einem Individuum unterscheiden. Ich möchte auf die umfassende Diskussion zu mathematischen Denkstilen jedoch nur hinweisen, sie aber in dieser Arbeit nicht vertiefen (vgl. für eine Übersicht Borromeo Ferri, 2003).

Im Folgenden wird der Einfluss des jeweiligen Trainings auf bevorzugte Formen der externalisierten Darstellung bei Schülern gezeigt. Dazu wurde überprüft, *welche* Modellierungsformen Schüler im Nachtest 1 im Vergleich zum Vortest verwendeten und *wie oft*. Tab.III.10 zeigt das Modellierungsverhalten⁹⁶ der *LK-Schüler*, die bereits in einem Leistungskurs Stochastik übliche Vorgehensweisen kennengelernt hatten.

Gruppe	FREQ LK (n=24)				PROB LK (n=23)			
	Vortest A2	Nach A2*	Nach A3	Nach A4	Vortest A2	Nach A2*	Nach A3	Nach A4
1. Keine Angabe	6	-	-	-	6	1	6	6
2. nur verbal, W	1	-	-	-	2	-	-	-
3. nur verbal, H	-	-	-	-	-	-	1	-
4. nur verbal, W und H	1	-	-	-	-	-	-	-
5. nur formal, W	1	-	-	-	1	17	9	12
6. „Häufigkeitsbaum“	-	22	22	22	-	-	-	-
7. nur Baumdiagramm, W	8	1	2	2	9	4	5	4
8. Baumdiagramm u. formal, W	5	-	-	-	5	1	2	1
9. nur Baumdiagramm, Mischform W und H	2	1	-	-	-	-	-	-

Legende: W: „im Verarbeitungsformat Wahrscheinlichkeit/rel. Häufigkeit“; H: „in Verarbeitungsformat abs.

Häufigkeit“; **fett**: im Training geübte Form; *: Da mehrere A2-Items im Nachtest vorkamen, sind mittlere Häufigkeiten (gerundet) angegeben.

Tab.III.10: Verwendungshäufigkeiten (absolut) von Modellierungsformen vor bzw. unmittelbar nach Training (bei Leistungskurs-Schülern; Itemtypen A2, A3, A4)

⁹⁴ Gemäß Kap.II.4. bezieht sich der Begriff „Modellierungsform“ hier insbesondere auf die Art der Darstellung und des numerischen Formates.

⁹⁵ Inwieweit die Verwendung tatsächlich als genereller Indikator für die kognitive Zugänglichkeit einer Modellierungsform gelten kann, ist zu diskutieren.

⁹⁶ Man beachte, dass die Häufigkeiten sich auf erkennbare Modellierungsversuche beziehen, aber keine Aussage über jeweilige Leistungen enthalten.

Es ist zu erkennen, dass die LK-Schüler *vor Training* zumeist eine Modellierung mit Wahrscheinlichkeitsbäumen versuchten, teilweise auch über eine Ableitung formaler Regeln mit Baumdiagrammen arbeiteten. Immerhin etwa ein Viertel der Schüler in beiden Gruppen hatten aber gar keine erkennbaren Vorstellungen zur Modellierung der Probleme bzw. schätzten lediglich aus den Angaben (vgl. auch Abb.III.10).

Nach *Häufigkeitstraining* benutzten 92% der Schüler bei allen Itemtypen⁹⁷ die gelernte Modellierungsform mit dem Häufigkeitsbaum (vgl. Abb.III.8). Es gab insbesondere keine „Ausfälle“, d.h. jeder Schüler war in der Lage eine Modellierung anzugeben.

Nach *Wahrscheinlichkeitstraining* wurde bei A2-Items nicht so häufig mit der im Training hergeleiteten formalen Regel modelliert (ca. 74%). Wenn man jedoch davon ausgeht, dass auch die Schüler, die Probleme (zusätzlich) mit Wahrscheinlichkeitsbäumen darstellten (Modellierungsform 7 und 8), „formal“ modellierten, lag die Verwendungsrate jedoch deutlich über 90%. Bei Items des Typs A3 und A4 fiel die Verwendungsrate auf 39% bzw. 52% (69,5% bzw. 74% inklusive Modellierungsformen 7 und 8). 26% gaben bei den Itemtypen A3 und A4 keine Modellierung an.

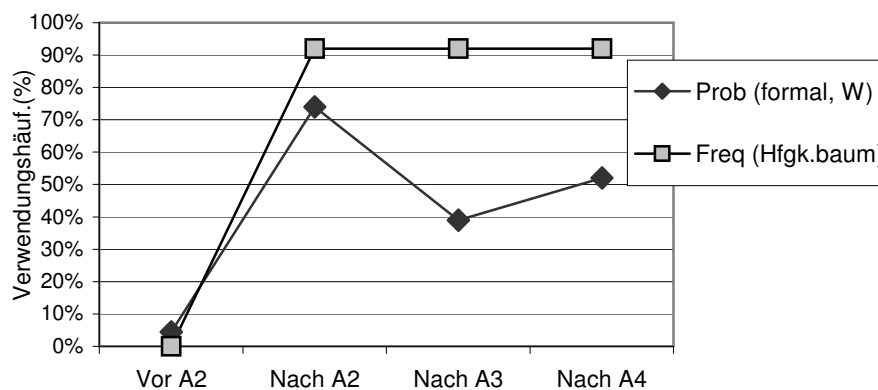


Abb.III.8: Verwendungshäufigkeiten der jeweils trainierten Modellierungsform vor bzw. unmittelbar nach Training (bei Leistungskurs-Schülern; Itemtypen A2, A3, A4)

Die Schüler der *10. Klasse* waren zu Untersuchungsbeginn ohne stochastisches Vorwissen aus dem Unterricht. Im *Vortest* wurde deshalb fast ausschließlich in verbaler Form modelliert (über alle Gruppen ca. 60%, vgl. Tab.III.11) oder gar nicht (ca. 35%, vgl. Abb.III.10).

Nach *Häufigkeitstraining* erfolgte bei A2-Items zu 84% bzw. bei A3-Items zu 66% Modellierung mit der trainierten Häufigkeitsmethode. 14% bzw. 30% konnten keine Modellierungsangaben machen.

Nach *Wahrscheinlichkeitstraining* wurde zu fast 70% formale Modellierung bei A2-Items und 46% bei A3-Items angewandt. 18% bzw. 54% konnten keine Modellierungsangaben machen.

Die favorisierte Modellierungsform nach *Kombi-Training* war deutlich der Häufigkeitsbaum (73% bei A2, 65% bei A3-Items). Die ebenfalls trainierte formale Modellierung wurde nur zu 11,5% bei A2-Items angewandt, bei A3-Items gar nicht. 8% bzw. 31% konnten keine Angaben zur Modellierung machen.

⁹⁷ Unabhängige Modellierung erfolgte nur bei den Itemtypen A2, A3, A4. Die Lösung der Items vom Typ A0 und A1 erfolgte immer im Rahmen der jeweiligen Modellierungsform des zugehörigen A2-Items (vgl. 3.1.2. B).

Gruppe	FREQ 10 (n=27)			PROB 10 (n=24)			COMBI 10 (n=26)		
	Vor A2	Nach A2	Nach A3	Vor A2	Nach A2	Nach A3	Vor A2	Nach A2	Nach A3
1. Keine Angaben	9 33%	11 13,6%	8 30%	8 33%	13 18%	13 54,2%	10 38,4%	6 7,7%	8 31%
2. nur verbal, W	14 52%	1 1,2%	-	12 50%	8 11,1%	-	14 54%	1 1,3%	-
3. nur verbal, H	-	-	-	-	-	-	1 3,8%	1 1,3%	-
4. nur verbal, W und H	3 11%	-	-	3 13%	1 1,4%	-	1 3,8%	-	-
5. nur formal, W	-	-	-	-	50 69,5%	11 45,8%	-	9 11,5%	-
6. „Häufigkeitsbaum“	-	68 84%	18 66%	-	-	-	-	57 73,1%	17 65%
7. nur Baumdiagramm, W	1 4%	1 1,2%	1 4%	1 4%	-	-	-	4 5,1%	1 4%
8. Baumdiagramm und formal, W	-	-	-	-	-	-	-	-	-
9. nur Baumdiagr., Mischform W und H	-	-	-	-	-	-	-	-	-

Legende: W: „im Verarbeitungsformat Wahrscheinlichkeit/rel. Häufigkeit“; H: „in Verarbeitungsformat abs. Häufigkeit“; **fett**: im Training geübte Form.

Tab.III.11: Verwendungshäufigkeiten (absolut und relativ) von Modellierungsformen vor bzw. unmittelbar nach Training (bei Schülern der Klasse 10; Itemtypen A2, A3)

Abb.III.9 fasst vergleichend zusammen, in welchem Umfang jeweils trainierte Modellierungsformen vor und nach Training in den verschiedenen Gruppen verwendet wurden.

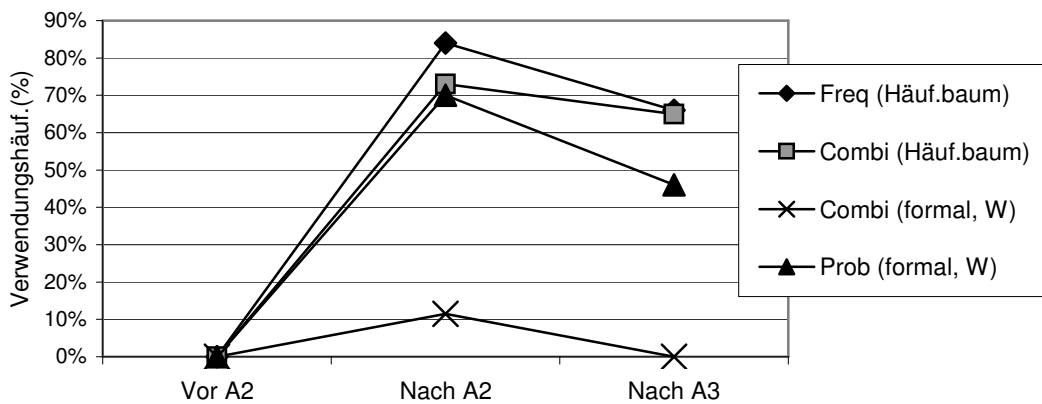


Abb.III.9: Verwendungshäufigkeiten trainierter Modellierungsformen vor bzw. unmittelbar nach Training (bei Schülern der Klasse 10; Itemtypen A2, A3)

Abb.III.10 gibt nochmal einen Überblick über alle Trainingsgruppen, wie oft vor und nach jeweiligem Training gar keine Modellierung angegeben werden konnte (bzw. ein Ergebnis ohne Modellierung geschätzt wurde).

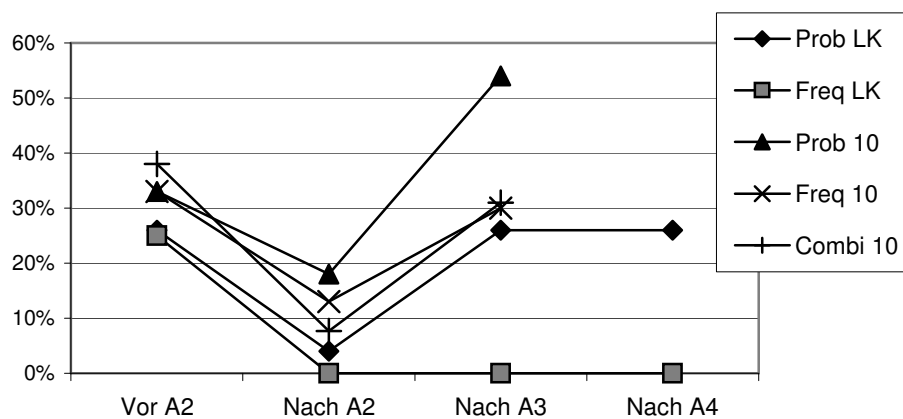


Abb.III.10: Fälle ohne Modellierung vor bzw. unmittelbar nach Training (Leistungskurs und Klasse 10; Itemtypen A2, A3, A4)

3.3.5 Leistungsunterschiede differenziert nach Modellierungsschritten

Bei der Analyse der schriftlichen Überlegungen und Argumentationen der Schüler wurde ein vierschrittiges Schema zur Bewertung des Modellierungsprozesses angewandt (vgl. 3.3.1). Eine naheliegende Frage war: Treten Leistungsdefizite bei Häufigkeits- bzw. Wahrscheinlichkeitsmodellierung in allen Schritten proportional auf oder gibt es in einzelnen Schritten über(unter-) proportionale Unterschiede?

Zur Analyse wurden bei den LK-Schülern die Leistungen in den einzelnen Modellierungsschritten bei Häufigkeitsmodellierern (Modellierungsform 6, vgl. Tab.III.10) mit denen der Wahrscheinlichkeitsmodellierer (Modellierungsformen 5, 7, 8) verglichen. Abb.III.11 zeigt höhere Leistungsdefizite der Wahrscheinlichkeitsmodellierer in Schritt 1 „Mathematisieren (i.e.S.)“ und Schritt 4 „Berechnen“. Abb.III.13 zeigt die Prozentanteile der Leistung der Wahrscheinlichkeitsmodellierer an den Leistungen der Häufigkeitsmodellierer in den jeweiligen Modellierungsschritten. Das Herausfiltern der Informationen aus Texten und die Übersetzung in mathematisch „verarbeitungsfähige“ Informationen („Mathematisieren i.e.S.“) führte zu erhöhten Schwierigkeiten im Vergleich zu folgenden Schritten. Auch bei der Berechnung, also rein technischen Anforderungen, zeigten die Wahrscheinlichkeitsmodellierer deutlich unterproportionale Leistungen.

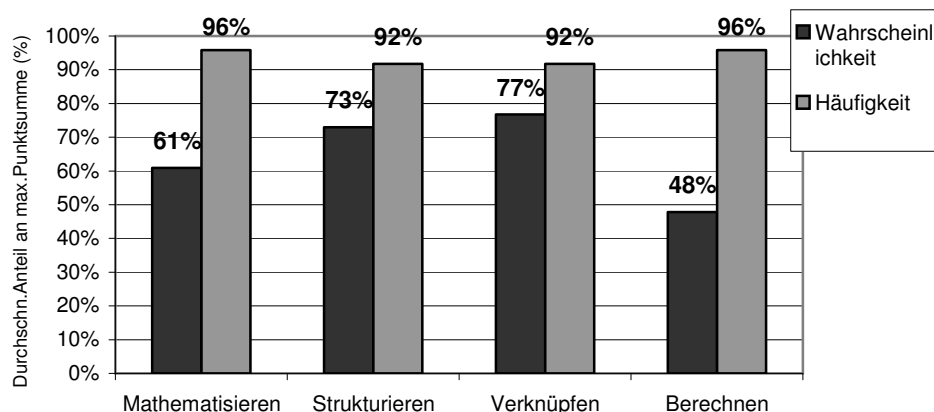


Abb.III.11: Testleistungen bei Wahrscheinlichkeits- bzw. Häufigkeitsmodellierung, differenziert nach Modellierungsschritten (Itemtyp A2)

Beim Itemtyp A3 ergibt sich ein qualitativ ähnlicher Zusammenhang, jedoch sind Unterschiede weniger ausgeprägt (vgl. Abb.III.12 und 13). Auffällig ist, dass „Mathematisieren (i.e.S.)“ sowohl mit Häufigkeiten als auch mit Wahrscheinlichkeit bei A3-Items überproportionale Leistungsdefizite aufwies im Vergleich zu den weiteren Modellierungsschritten.

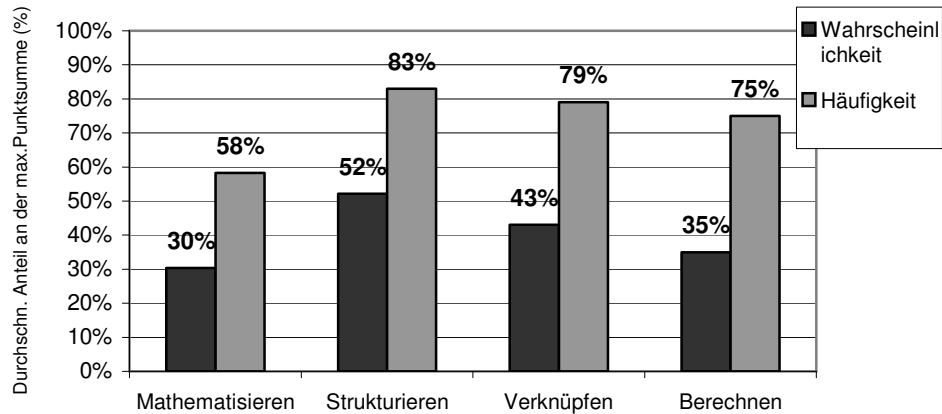


Abb.III.12: Durchschnittliche Testleistungen bei Wahrscheinlichkeits- bzw. Häufigkeitsmodellierung, differenziert nach Modellierungsschritten (Itemtyp A3)

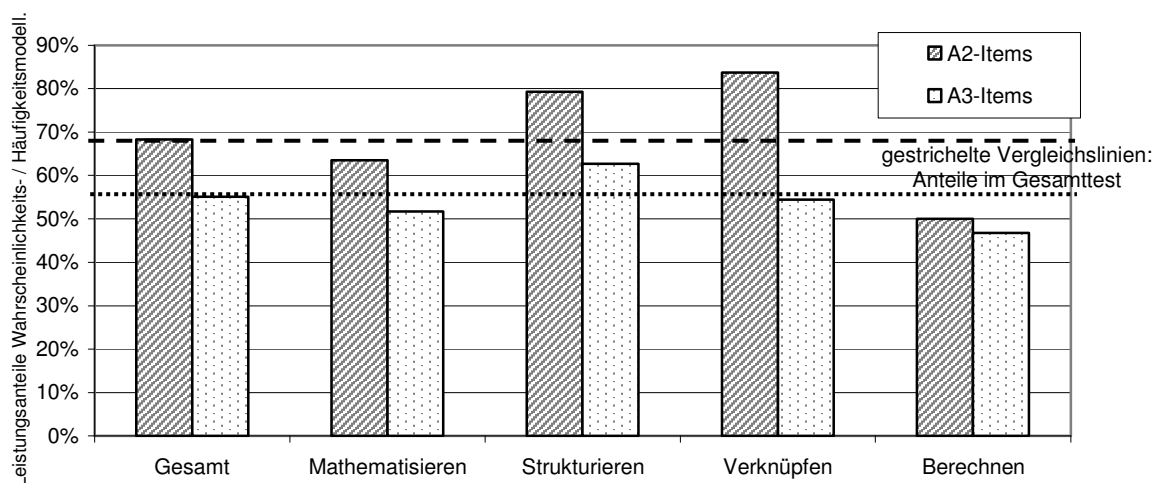


Abb.III.13: Anteile der durchschnittlichen Testleistungen bei Wahrscheinlichkeitsmodellierung von Leistungen bei Häufigkeitsmodellierung, Gesamttest und differenziert nach Modellierungsschritten (bei Itemtypen A2 und A3)

3.4 Zusammenfassung und Diskussion

Die Ergebnisse dieser Trainingsstudie stützen die Hypothese, dass Training mit dem Konzept der „natürlichen Häufigkeiten“ gegenüber einem formalen Wahrscheinlichkeitszugang zu höherem *Lernerfolg bei Bayesianischem Denken* führt. Sowohl für Schüler der Sekundarstufe I (10. Klasse, ohne unterrichtlichem Vorwissen) als auch der Sekundarstufe II (Leistungskurs Mathematik, mit unterrichtlichem Vorwissen) wurden deutliche Unterschiede im Lernerfolg festgestellt: Die Schüler zeigten bereits nach einem maximal 2-stündigen computerbasierten Training mit Häufigkeitsbäumen nicht nur bei den trainierten dichotomen Bayesproblemen deutlich höhere Leistungen als nach entsprechendem Wahrscheinlichkeitstraining, sondern auch bei nicht trainierten komplexeren Problemen, die strukturelle Transferleistungen bei der Modellierung abverlangten. Hinsichtlich der vorherrschenden Art des mathematischen Arbeitens zeigten sich diese Unterschiede gleichermaßen, jedoch stärker bei vorwiegend

rechnerischem Modellieren als bei vorwiegend begrifflichem Modellieren. Die Befunde beziehen sich somit nicht nur auf prozedurales, sondern auch auf konzeptuelles Bayesianisches Denken. Insgesamt lag das in dem entwickelten *Kompetenztest zu Bayesianischem Denken* erreichte Niveau nach Häufigkeitstraining im Gruppendurchschnitt bei Zehntklässlern bei ca. 60% (der maximalen Testleistung) bzw. bei Leistungskurs-Schülern bei 80%, nach Wahrscheinlichkeitstraining bei ca. 35% bzw. 50%.

Ein für den Lernerfolg entscheidendes Kriterium ist seine *Nachhaltigkeit*. Die Leistungen bei Schülern, die mit „natürlichen Häufigkeiten“ trainiert wurden, blieben längerfristig deutlich stabiler, während nach (reinem) Wahrscheinlichkeitsregeltraining meist geradezu ein „Verfall“ des unmittelbaren Trainingseffektes zu beobachten war. Die Leistungsunterschiede zwischen Wahrscheinlichkeits- und Häufigkeitsgruppen vergrößerten sich aus diesem Grunde im Allgemeinen nach etwa 3 Monaten nochmals deutlich. Insbesondere weisen Resultate der Gruppe mit kombiniertem Training (Combi) darauf hin, dass ein Zugang über Häufigkeitsmodellierung zu formalen Wahrscheinlichkeitsregeln im Vergleich zu einem „direkten“ wahrscheinlichkeitstheoretischen Zugang die Stabilität des Lernerfolges erheblich unterstützt. Gewissermaßen erscheint der Häufigkeitszugang als Basis für einen formaleren Ausbau sehr nützlich.

Die Analyse der tatsächlich zur Lösung verwendeten *Modellierungsformen* ergab grundsätzlich hohe Verwendungsraten der im Training benutzten Formen bei den trainierten A2-Itemtypen. Bei „neuen“ Itemtypen (A3, A4) konnten in den Wahrscheinlichkeitsgruppen deutlich weniger Schüler die gelernte Form strukturell modifizieren. Häufigkeitstraining unterstützte diese „Transferleistungen“ besser. Als Folge kam es nach reinem Wahrscheinlichkeitstraining deutlich häufiger zum „Versagen“ bei solchen Transferaufgaben. Ein besorgniserregender Befund, der eine gewisse Ueffektivität derzeitigen Stochastikunterrichts zur Verbesserung stochastischen Denkens (vgl. auch Rasfeld, 2004) unterstreicht, ist meines Erachtens, dass trotz der Behandlung des Satzes von Bayes kurz vorher im Unterricht, ca. ein Viertel der Leistungskurs-Schüler im Vortest gar keine Angaben zur Modellierung von anwendungsbezogenen, dichotomen Bayesproblemen (A2-Items) machen konnte.

Aus den Resultaten der Combi-Gruppe (10.Klasse), die sowohl das Modellieren mit Häufigkeitsbäumen als auch mit Wahrscheinlichkeitsregeln gelernt hatten, ist eine sehr deutliche Präferenz zur Häufigkeitsmodellierung festzustellen. Auch dieser Befund deutet aus meiner Sicht darauf hin, dass häufigkeitsbasiertes Modellieren grundlegende Intuitionen anspricht, während formales Modellieren ihnen womöglich sogar widerspricht. Ich möchte zwei weitere Interpretationsansätze anbieten, die bereits in 3.3.4 genannt wurden: Erstens kann Häufigkeitsmodellierung tatsächlich kognitiv zugänglicher sein, d.h. dass Schüler eher bereit sind, Häufigkeitsdarstellungen als adäquat für die Problemlösung zu akzeptieren als Wahrscheinlichkeitsmodelle. Zweitens wird womöglich ein visueller Denkstil von vielen Schülern bevorzugt, d.h. z.B. dass intern vorherrschende bildliche Vorstellungen sich in einer Präferenz für bildliche, externalisierte Darstellungen äußern. Entscheidendes Argument ist dabei nicht die Verwendung von grafischen Modellen, sondern die Verbindung mit einer konkreten Häufigkeitsvorstellung, die den visuellen Denkstil besser unterstützt als das mathematische Modell „Wahrscheinlichkeit“.

Zur genaueren Analyse individueller *Modellierungsschritte* bei außermathematischen Bayes-Situationen wurde der Modellierungsprozess von der Situation (Welt) zu mathematischen Resultaten in vier „konkretere“ Modellierungsschritte zerlegt, die mit „Mathematisieren i.e.S.“, „Strukturieren“, „Verknüpfen“ und „Berechnen“ bezeichnet

wurden. Ein Vergleich der Leistungen bei den einzelnen Modellierungsschritten wies (bei A2-Itemtypen) auf überdurchschnittliche Leistungsdefizite der Wahrscheinlichkeitsgruppe in den Schritten „Mathematisieren i.e.S.“ und „Berechnen“ hin. Letzteres ist einsehbar, da rein technische Anforderungen (z.B. Bruchrechnen, Prozentrechnen) bei der Häufigkeitsmodellierung geringer sind als bei der Wahrscheinlichkeitsmodellierung. „Mathematisieren i.e.S.“ bezieht sich auf Herausfiltern der relevanten Informationen aus dem Aufgabentext und Übersetzen in eine „mathematisch verarbeitungsfähige“ Form („Welche Informationen sind relevant und was bedeuten sie? In welcher Form können/sollen sie mathematisch bearbeitet werden?“), ohne dass bereits das komplett strukturierte mathematische Modell gebildet wird. Gerade bei diesem grundlegenden Schritt von der Realität zur Mathematik scheint die besondere Stärke von Häufigkeiten zu liegen. Die weiteren Schritte sind womöglich vergleichsweise routinetafeler und leichter zu trainieren („Verwendung von Formeln“ und „umformen“ bzw. „Bilden eines Häufigkeitsbaumes“ und „Lösungsbruch bilden“). Die Leistungsdefizite bei Wahrscheinlichkeitsmodellierung fallen hier weniger stark aus. Diese Befunde deuten darauf hin, dass – neben technischen Grundfertigkeiten – gerade die grundlegenden Schritte des Mathematisierens (Übergang Situation – Modell) bei außermathematischer Modellierung durch eine Häufigkeitsübersetzung besonders unterstützt werden.

Die Befunde stützen die kognitionspsychologische Theorie, dass "natürliche Häufigkeiten" als evolutiv ursprünglichste Form für die Verarbeitung von unsicheren Situationen besonders adaptiv sind („adaptive algorithms“) und somit ein häufigkeitsbasierter Zugang zu stochastischen Themen an grundlegende Intuitionen der Schüler anknüpft. Der Aufbau auf grundlegende Intuitionen und das Wechselspiel mit der Theorie werden als zentrales Element in der zeitgemäßen didaktischen Sichtweise von mathematischer Bildung (wieder) sehr betont, sei es etwa als „mentale Modelle“ gemäß Freudenthal (1983) oder als „primäre Intuitionen“ bei Fischbein (1975, 1987).

Im Sinne der Unterstützung intuitiven stochastischen Denkens empfiehlt sich die zentrale Verwendung des vorgestellten didaktischen Häufigkeitskonzeptes für einen anwendungsbezogenen Einstieg in den Problembereich bedingte Wahrscheinlichkeit und Bayesianwendungen (vgl. auch Wassner, Martignon & Sedlmeier, 2002). Weitergehend ermutigen die Ergebnisse der Trainingsstudie bereits zu diesem Einstieg in der Sekundarstufe I, der bei spiraligem Aufbau des Curriculums in der Sekundarstufe II je nach gefordertem Niveau wahrscheinlichkeitstheoretisch vertieft und präzisiert werden kann. Wie in Kapitel II genauer ausgeführt wurde, werden in üblichen Zugängen bedingte Wahrscheinlichkeiten auch zunächst über absolute Häufigkeiten anschaulich berechnet, und die Überlegungen benutzt, um formale Wahrscheinlichkeitsregeln herzuleiten (z.B. Bayesregel). Bei der späteren mathematischen Modellierung von (anwendungsbezogenen) Aufgaben werden diese Regeln in Verbindung mit grafischen Darstellungen benutzt, aber auf die anschaulichen Häufigkeitsüberlegungen wird meist nicht mehr zurückgegriffen. Genau in diesem Punkt soll sich die im folgenden Kapitel dargestellte Realisierung eines Zugangs zu Bayesianischem Denken grundlegend von bisherigen unterscheiden, indem als grundsätzliches „Werkzeug“ Häufigkeitsmodellierungen benutzt werden.

KAPITEL IV

REALISIERUNG DES KONZEPTE DER „NATÜRLICHEN HÄUFIGKEITEN“ IM UNTERRICHT DER GYMNASIALEN SEKUNDARSTUFE I

Eine oft geäußerte Meinung von Lehrenden bei einer Befragung⁹⁸, die ich vor Beginn meiner Promotion durchführte, lautete: „Die Schüler haben mit dem Satz von Bayes Schwierigkeiten, egal wie er eingeführt wird“. Tatsächlich führt der im derzeitigen Stochastikunterricht verfolgte Zugang zu Themen um bedingte Wahrscheinlichkeit und Bayesanwendungen oftmals offensichtlich nicht zu den gewünschten Erfolgen. Dieses Defizit, das viele Lehrende ohne zu zögern einräumen, kann auch an den Vortestleistungen der untersuchten Leistungskurs-Schüler abgelesen werden, die unmittelbar vorher zu diesen Themen unterrichtet worden waren. 41% der an den Trainingsstudien teilnehmenden Leistungskurs-Schüler gaben in einer Vorbefragung an, sie könnten sich nicht mehr genau an die Bayes-Formel erinnern bzw. sie herleiten, welche sie kurz vorher im Unterricht gelernt und ausgiebig benutzt hatten. Die Vortestleistungen bestätigten, dass der vorherige Unterricht große Mängel hinterließ: 25% konnten gar keine Modellierung angeben; nur ca. 15% konnten ein typisches, dichotomes Bayes-Problem vollständig richtig modellieren und lösen (vgl. Kap.III, 3.3.4). In einer Untersuchung von Rasfeld (2004) bei Zehntklässlern (11 Klassen, NRW) wurde auch eine gewisse Ineffektivität des derzeitigen Stochastikunterrichts festgestellt: Intuitives stochastisches Denken fokussierend zeigten Zehntklässler, die Stochastikunterricht hatten, nur marginal bessere Leistungen als nicht in Stochastik unterrichtete 10.Klassen. Rasfeld (2004, S.59) zieht als Fazit: „Eine Verbesserung des intuitiven Verständnisses stochastischer Problemstellungen [durch üblichen Stochastikunterricht in der Sekundarstufe I, Anm. d. Verf.] findet statt, aber nicht in dem gewünschten Ausmaß. Auch die Schüler mit stochastischen Vorkenntnissen zeigen noch im erheblichen, teilweise unverändert großen Umfang Fehlvorstellungen [...] z.B. Schwierigkeiten beim Modellieren stochastischer Problemstellungen, gehen unreflektiert mit Laplace-Wahrscheinlichkeiten um [etc.]“.

Didaktische Vorschläge zur Verbesserung stochastischen (v.a. Bayesianischen) Denkens wurden an vielen Stellen dieser Arbeit bereits umfassend vorgestellt und diskutiert. Das Ziel des nun dargestellten Zugangs ist, innerhalb eines Stochastiklehrgangs *für die Sekundarstufe I* einen vorläufigen und tragfähigen Abschluss zu *Themen um „Bedingte Wahrscheinlichkeit und Bayesanwendungen“* zu schaffen⁹⁹. Aus der dargestellten Forschungsdiskussion und den Trainingsuntersuchungen wurden einige *Leitideen* für die Realisierung des eigenen Zugangs gewonnen:

- Das Konzept der „natürlichen Häufigkeiten“ unterstützt intuitives stochastisches Denken und ermöglicht den Schülern nachhaltig darauf aufbauend sinnstiftendes und transparentes Modellieren stochastischer Problemstellungen (belegt durch Trainingsstudien in Kap. III).

⁹⁸ 40 Mathematiklehrende wurden aus verschiedenen deutschen Bundesländern zu Einstellungen und Meinungen zum Stochastikunterricht allgemein und speziell zum Themengebiet „Bedingte Wahrscheinlichkeit und Satz von Bayes“ befragt. Ich gehe im Rahmen dieser Arbeit nicht detailliert auf Befunde ein (genauer: Wassner, C., Ergebnisse einer Fragebogenstudie zum Stochastikunterricht, unveröff. Manuskript).

⁹⁹ Dies ist einerseits wichtig für Schüler, die am Ende der Sekundarstufe I ausscheiden, aber auch als Voraussetzung für leistungsstarken Stochastikunterricht in der Sekundarstufe II (vgl. Schupp, 1984, S.238).

- Die Entwicklung des formalen Kalküls wirkt zunächst kontraproduktiv v.a. auf außermathematisches Modellieren und erschwert für viele Schüler die kognitive Zugänglichkeit. Demotivation kann die Folge sein.
- Da die Häufigkeitsanschauung grundlegend für die Entwicklung des formalen Kalküls ist, kann dieser unproblematisch später (wenn erforderlich) daraus entwickelt werden. Diese Vorgehensweise unterstützt auch die Forderung nach einem spiraligen Aufbau des Stochastikcurriculums (z.B. Heitele, 1976).¹⁰⁰

Das primäre Unterrichtsziel ist also gerade *nicht* die Entwicklung und Anwendung des formalen Kalküls, das einem gewissen Pragmatismus zu folgen scheint: Leichtere Testbarkeit, Zwänge des Lehrplans, zeitliche Restriktionen oder schlichtweg fehlendes Wissen der Lehrenden um andere Zugänge. Eine befragte Lehrerin formulierte es so: „Ich hätte gern mehr Zeit und Möglichkeiten für realitätsbezogeneren Aufgaben und schülernah zu unterrichten, also z.B. Eingehen auf außerschulische Interessengebiete der Schüler, Ideen von Schülerseite näher betrachten, Team- und Projektarbeit“. Diese Aussage macht meines Erachtens deutlich, dass oft wichtige andere „Bildungsaspekte“ wie Lebensvorbereitung, Problemlösefähigkeit, kritischer Vernunftgebrauch bei Modellierung und Interpretation, Sinnstiftung etc. (vgl. z.B. Blum, 1996, S.20f.) vernachlässigt werden (müssen). Als Ergebnis wird zwar inhaltlich Konsistenz mit dem mathematischen Theoriegebäude erreicht, welches aber kaum zur realen Welt in Bezug gebracht werden kann. Dieses „Formelwissen“ wird deshalb schnell wieder als „bedeutungslos“ vergessen (vgl. dazu ausführlicher Kap.II.2.2.5).¹⁰¹ Deshalb wurde im folgenden Zugang ebenfalls als wesentlich erachtet:

- Auch für Schüler relevante, interessante, realitätsbezogene Problemstellungen sollen zur motivierenden Heranführung an die entsprechenden Inhalte dienen, da gerade dieser Themenbereich hervorragende Möglichkeiten dazu bietet. Auf keinen Fall darf die Chance zur Motivation, Sinngebung etc. durch die oft zur Einführung übliche Verwendung von „Münzen-Würfel-Urnen“-Aufgaben („Würfelbudenmathematik“, Schupp 1984) verschenkt werden.

Im Folgenden stelle ich zunächst den Entwurf einer einführenden Unterrichtsreihe zum Themenkreis „Bedingte Wahrscheinlichkeit und Bayesianwendungen“ vor, die auf den genannten Leitideen basiert. Anschließend werde ich eine Unterrichtsstudie in der Sekundarstufe I besprechen, deren Untersuchungsgegenstand die Realisierung, Exploration und Optimierung eines auf der entworfenen Unterrichtsreihe basierenden Lehr-Lern-Szenarios war. Das methodische Vorgehen bei der Exploration wird dabei ausführlich beschrieben und die Ergebnisse werden dargestellt und diskutiert.

¹⁰⁰ Schupp (1984, S.238) zum spiraligen Aufbau des Curriculums: „Statt jeweils die Theorie ein Stück voranzutreiben, um sie allenfalls dann anzuwenden, sollte man mehr nach der Komplexität der stochastischen Situationen gliedern, mit denen die Schüler konfrontiert sein werden. [...] Hilfsmittel sollten nur und erst dann gebracht werden, wenn sie beim Schüler wirklich als Hilfsmittel ankommen, weil sie komplexe Geflechte transparent zu machen vermögen“.

¹⁰¹ Bei der Befragung von 40 Mathematiklehrenden hielten 35 die Verbindung von Unterrichtsstoff und täglichem Leben im Stochastikunterricht für besonders wichtig im Vergleich zu sonstigen Mathematikgebieten.

1 Entwurf einer Unterrichtsreihe „Authentisches Bewerten und Urteilen unter Unsicherheit“

Nach Vorliegen der Trainingsbefunde wurde begonnen, auf Grundlage des Lehrplanes von Nordrhein-Westfalen für die gymnasiale Sekundarstufe I eine Unterrichtsreihe zum Thema „Bayessche Regel“ zu entwickeln. Aus obigen Leitideen ergibt sich der treffendere Titel „Authentisches Bewerten und Urteilen unter Unsicherheit“. In der von Rolf Biehler geleiteten Arbeitsgruppe an der Universität Kassel entstand unter Mitarbeit des Mathematiklehrers Stefan Schweynoch ein umfassender Unterrichtsentwurf mit didaktischen Anmerkungen, der nun vorgestellt werden soll.¹⁰²

1.1 Ziele und Einordnung

Der Hauptinhalt der Reihe ist eine realitätsbezogene Erarbeitung des Begriffs der bedingten Wahrscheinlichkeit und das Bewerten und Beurteilen von Anwendungsproblemen zum Themenkreis Bayesregel durch Verwendung von Häufigkeitsdarstellungen.

Die Reihe wurde bewusst auf Basis des Lehrplans von Nordrhein-Westfalen entworfen. Der NRW-Lehrplan sieht (optional) die Behandlung des Themenkreises „Bayessche Regel“ und Schaffung inhaltlichen Verständnisses für Zusammenhänge (ohne formale Entwicklung) im Unterricht der 9. oder 10. Jahrgangsstufe vor. Die Relevanz der Regel für statistisches Denken soll auch deutlich werden (vgl. Kap.II.1.1; Kultusministerium NRW, 1993, S.21). Dies ist u.a. auf den Einfluss von Wolfgang Riemer zurückzuführen, der auch Mitglied der Lehrplankommission war und dessen Publikationen (z.B. Riemer 1985; 1986) das Thema als unterrichtbar in der Sekundarstufe I erscheinen ließen. Der Themenkreis baut auf dem Vorwissen aus früheren Stufen zum Thema mehrstufige Zufallsexperimente (insbesondere Baumdiagramme und Pfadregeln) sowie auf grundlegenden Interpretationen von Wahrscheinlichkeit auf (klassisch, frequentistisch). Die Lehrplanvorgaben haben sich auch in Schulbüchern niedergeschlagen (z.B. Schmid & Weidig, 1996, Mitautor Riemer; Kuypers et al., 1999; vgl. ausführlicher Kap.II.2.2.2). Da insofern in den nordrhein-westfälischen Gymnasien optimale Lehrplan- und Schulbuchvoraussetzungen gegeben waren, wurde die eigene Unterrichtsreihe auf die gymnasiale Sekundarstufe I in NRW abgestimmt und kam dort erstmalig zum Einsatz.

Als übergeordnetes Lernziel sollen die Schüler Situationen modellieren, beurteilen und durchdringen können, bei denen formal die Bayesregel Verwendung findet. Ohne dass diese benannt oder formal abgeleitet wird, lernen sie (mathematische) Mittel zur Modellierung dieser Situationen kennen. Im Einzelnen werden etwa folgende Kompetenzen von den Schülern erwartet:

Die Schüler können (bzw. kennen)...

- realistische außermathematische Problemstellungen, bei denen die Bayesregel eine praktische Bedeutung hat.
- statistische Informationen aus realistischen Texten zur Beurteilung der Problemstellung geeignet mit Häufigkeitsdarstellungen modellieren.
- mithilfe der Modellierungen die Problemstellungen bewerten, beurteilen und im Sachkontext interpretieren.

¹⁰² Es existiert das bisher unveröff. Manuskript: Wassner, C., Biehler, R., Schweynoch, S., Martignon, L. Authentisches Bewerten und Urteilen unter Unsicherheit – Arbeitsmaterialien und didaktische Kommentare für den Themenbereich „Bayessche Regel“ in der Sekundarstufe I (vgl. Überblick in Wassner & Biehler, im Druck).

- ungewöhnlich erscheinende Wahrscheinlichkeitsergebnisse begründen und Zusammenhänge erklären.
- Gegebenenfalls Grundannahmen vernünftig verändern und Auswirkungen erkennen (z.B. Basisratenproblem).
- zentrale Wahrscheinlichkeitsbegriffe wie Einzelwahrscheinlichkeit, UND-Wahrscheinlichkeit und bedingte Wahrscheinlichkeit, a-priori- und a-posteriori-Wahrscheinlichkeit verstehen, abgrenzen und bei der Modellierung verwenden.
- die Problemsituationen auch als „statistische Denkweise“ der Neubewertung der Wahrscheinlichkeit einer Hypothese auf Grund von neuen Daten oder Beobachtungen („Lernen aus Erfahrung“) begreifen.

Die Bildungsziele der Unterrichtsreihe gehen deutlich über mathematisches Modellieren realitätsbezogener Situationen, das Inhalt der Trainings in Kapitel III war, hinaus. Um sich einer praktischen, individuellen oder gesellschaftlichen Bedeutung von Mathematik in diesem Zusammenhang bewusst werden zu können, wurde Zeit für die Erarbeitung authentischer, kontextbeleuchtender Texte eingeplant („Sinnstiftung“). Solche dienen idealerweise als Anstoß für Diskussion, subjektive Meinungen und Deutungen, eigene Ideen, kritischen Vernunftgebrauch etc. (vgl. Biehler, Heymann & Winkelmann, 1995). Vermeintlich eindeutige Phänomene und Vorurteile des Alltags (z.B. bei medizinischen Methoden und Ergebnissen) sollen kritisch beleuchtet und durch begründete Urteile ersetzt werden (vgl. z.B. Krämer, 1995; Beck-Bornholdt & Dubben, 2001; Gigerenzer, 2002). Die gelernte Mathematik soll gleichsam zum „Verstärker des Alltagsdenkens“ werden (vgl. These 4, Heymann, 1995). Die Folge ist, dass sich der größere Teil des Unterrichts um die vielschichtige Erarbeitung und Durchleuchtung konkreter Situationen dreht, anstatt mathematisch möglichst „tief“ vorzudringen und möglichst viele Übungsaufgaben zu bearbeiten (vgl. auch Wagenschein, 1975).

1.2 Inhalte

Die Frage nach der Entwicklung mathematischer Inhalte war durch das Konzept der „natürlichen Häufigkeiten“ bereits weitgehend geklärt. Neben Modellierung mit Häufigkeitsbäumen („Doppelbäumen“; vgl. Abb.II.14) wurden auch analoge Mehrfeldertafeln mit Häufigkeiten („Vierfeldertafel mit absoluten Häufigkeiten“; vgl. Tab.II.1a) verwendet, um den Schülern Wahlmöglichkeiten bei der Modellierung zu überlassen. Verallgemeinerung, Formalisierung und Begriffsgefüge wurden bewusst „vorsichtig“ und nur soweit wie nötig entwickelt. Die Frage nach geeigneten realistischen Situationen und ihrer (stochastischen) Bearbeitung und Erschließung im Unterricht war eine Kernfrage bei der inhaltlichen Gestaltung. Nach Sichtung bereits bestehender bzw. Sammeln und Aufarbeiten geeigneter, realitätsnaher Problembereiche konnte umfangreiches Material für den Unterricht erstellt werden.

Die Inhalte der Unterrichtsreihe waren durch die Vorgabe der Realisierung am nordrhein-westfälischen Gymnasium so zu gestalten, dass der zeitliche Umfang etwa 15 Schulstunden nicht überschreitet. Zur Erleichterung des Einsatzes wurden „Arbeitsblätter“ mit Zusatzmaterialien für Schüler mit didaktischen Kommentaren und Lösungsvorschlägen für die Lehrenden entworfen (komplett in Anhang B). Ich diskutiere im Folgenden überblicksartig den Aufbau der Unterrichtsreihe und möchte dabei vor allem den Bezug zu didaktischen Grundpositionen verdeutlichen.

1.2.1 Übersicht

Tab.IV.1 gibt eine Übersicht über alle im Rahmen der Unterrichtsreihe erstellten Arbeitsblätter und Materialien, die jeweils mit didaktischen Kommentaren und Lösungen für die Lehrenden versehen sind. Zur Chronologie ist zu bemerken, dass nach der Einführung in AB 1 bis zum Überblick 1 prinzipiell die Reihenfolge der Bearbeitung vertauschbar ist. AB 7 gibt die Hinführung zu Überblick 2. Die folgenden Anwendungen und Übungen in AB 9 bis AB 14 sind in der Reihenfolge der Bearbeitung ebenfalls frei wählbar.

ARBEITSBLATT	Titel	Hauptinhalt
AB 1	<i>Wie sicher ist der AIDS-Test?</i>	(Intuitive) Problemlösung
AB 2	<i>Das AIDS-Testverfahren im Detail</i>	Ergebnisvertiefung
AB 3	<i>Mögliche Folgen von positiven Testergebnissen</i>	Kontextvertiefung
AB 4	<i>Einfluss der Basisrate</i>	Vertiefung und Erweiterung der Modellierung
AB 5	<i>Das AIDS-Test Problem aus anderer Sicht</i>	
ÜBERBLICK 1	<i>Wahrscheinlichkeitsbegriffe</i>	Verallgemeinerung Zusammenfassung Begriffsvertiefung Übung
AB 6	<i>Übungen zum ÜBERBLICK 1</i>	
AB 7	<i>Mordfall</i>	Anwendung zu Urteilen mit Indizien (führt zu Überblick 2)
ÜBERBLICK 2	<i>Wahrscheinlichkeiten neu bewerten</i>	Verallgemeinerung & Zusammenfassung Begriffsvertiefung Übung
AB 8	<i>Übungen zum ÜBERBLICK 2</i>	
AB 9 (+Zusatzmaterial)	<i>Schwangerschafts- und Vaterschaftstest</i>	Anwendung (mit Kontextvertiefung)
AB 10 (+Zusatzmaterial)	<i>Mammografie</i>	Anwendung (mit Kontextvertiefung)
AB 11 (+Zusatzmaterial)	<i>BSE-Krise</i>	Anwendung (mit Kontextvertiefung)
AB 12 (+Zusatzmaterial)	<i>Drogen im Straßenverkehr</i>	Anwendung (mit Kontextvertiefung)
AB 13	<i>Übungsaufgaben 1</i>	Übung
AB 14	<i>Übungsaufgaben 2</i>	Übung (erweiterte Modellierung)

Tab.IV.1: Übersicht über die Unterrichtsreihe (komplett in Anhang B)

1.2.2 Einführung mit einer realen Problemstellung

Am Beginn der Unterrichtseinheit steht keine vordergründig wahrscheinlichkeitstheoretische Fragestellung, sondern eine, die für Schüler individuelle Bedeutung haben könnte: „Wie sicher ist ein AIDS-Test?“. Dass sich die AIDS-Test-Problematik als didaktisch besonders lohnendes, realistisches Beispiel für die Anwendung des Satzes von Bayes eignet, ist altbekannt. Z.B. in Unterrichtsvorschlägen von König (1991), Böer (1997) oder Israel (2001) findet es Verwendung. Heymann (1995, S.24) formulierte treffend: „Verstehen ist auf die geistige Aktivität der Lernenden angewiesen, auf die innere Fragehaltung, die sich ‚von außen‘ nicht erzwingen läßt“. Wie ließe sich aber besser eine solche „innere Fragehaltung“ induzieren als durch eine derartig untypische Fragestellung zu Beginn einer Mathematikstunde? Die Neugier der Schüler wäre wenigstens dahingehend geweckt, was diese Frage wohl mit Mathematik zu tun hat. Diese Neugier kann geschickt genutzt werden, um Schüler zum „Mitdenken“ zu bewegen. Das dürfte ungleich schwieriger sein, wenn sich in die erste Stunde um ein banales Würfelspiel oder Urnenexperiment dreht (wohlgemerkt bei 15-jährigen Jugendlichen).

Arbeitsblatt 1 (vgl. Anhang B)

Ein Informationstext dient als Einstieg:

Wie sicher ist der AIDS-Test?

Der sogenannte AIDS-Test ist einer der zuverlässigsten Tests, die jemals entwickelt wurden. Er wird eingesetzt, um eine Infektion mit HIV festzustellen(*). Wegen der hohen Gefahr der Verbreitung der tödlichen HIV-Infektion war sogar lange Zeit in der Diskussion, ob nicht die gesamte Bevölkerung zum AIDS-Test gezwungen werden soll.

Der AIDS-Test ist aber nicht perfekt. Wenn jemand HIV-infiziert ist, soll der Test positiv sein. Zu 99,9% fällt er dann auch positiv aus. Andererseits wenn jemand nicht HIV-infiziert ist, soll der Test natürlich negativ sein. Zu 99,7% fällt er dann tatsächlich negativ aus.

Nehmen wir mal an, dass für alle Menschen in NRW ein AIDS-Test durchgeführt werden soll. Laut Schätzung des Robert-Koch-Instituts sind bundesweit 0,05% der Bevölkerung HIV-infiziert, die Quote kann auch für NRW angenommen werden. Die Bevölkerungsstatistik gibt an, dass in NRW 18.000.000 Menschen leben.

(*) Im Sprachgebrauch hat sich AIDS-Test eingebürgert. AIDS bezeichnet eigentlich die Krankheit, die man bekommen kann, wenn man mit HIV infiziert ist. HIV kommt vom engl. „human immunodeficiency virus“ = „Immunschwäche-Virus beim Menschen“.

Eine Diskussion über die Textinhalte soll zunächst frei von den Schülern geführt werden, wobei eine Verstrickung in Details vermieden werden soll. In diesem Zugang sollen die Fragestellungen nicht erst mit völlig offenem Datenmaterial von den Schülern selbst erarbeitet werden (wie z.B. bei Israel, 2001). Eine solche Erweiterung des Zugangs wäre gleichwohl sicher von Nutzen, wenn eine stärkere Verzahnung von Inhalten der beschreibenden Statistik mit der Wahrscheinlichkeitsrechnung erreicht werden soll. Sie könnte auch dieser Unterrichtsreihe „vorgeschaltet“ werden. Allerdings müsste dann ein größerer Zeitrahmen zur Verfügung stehen, und es würden erheblich höhere Anforderungen an den Lehrenden gestellt, den „roten Faden“ im Unterricht nicht zu verlieren.

Die bereits durch einen Text strukturiertere, aber deswegen nicht unrealistischere Situation kann natürlich einfacher und schneller zu den wahrscheinlichkeitstheoretischen Problemen führen als der völlig offene Zugang. So soll auch bereits *in dieser Einführungsstunde* deutlich werden, was diese Situation mit Mathematik zu tun hat. Das Ziel ist, den *Regelkreis des mathematischen Modellierens* (vgl. z.B. Schupp, 1984) bereits „komplett“ mit dem „Grundproblem“ *zu durchlaufen*, um später die Situation begründeter vertiefen, präzisieren und unter anderen Gesichtspunkten beleuchten zu können:

„Stell Dir vor, eine beliebige Person aus NRW bekommt mitgeteilt, dass ihr Test positiv ist. Wie sicher kann sie sein, dass sie tatsächlich HIV-infiziert ist?“

Der nächste Schritt ist die Formulierung des zugehörigen wahrscheinlichkeitstheoretischen Problems:

„Wie hoch ist die Wahrscheinlichkeit für eine HIV-Infektion, wenn der Test positiv ist?“

Vor dem Versuch, mit den Schülern eine mathematische Modellierung zu entwickeln, sollen *alle Schüler* das Ergebnis *schätzen* und die Schätzungen werden notiert.¹⁰³ Da die meisten Menschen in dieser Situation ein viel zu hohes Ergebnis schätzen (siehe Kap.I.2.1), bietet sich die einmalige Gelegenheit, zu zeigen, dass mit Mathematik gewissermaßen „Entdeckungen“ möglich sind (Wittenberg, 1963, S.46f.).

¹⁰³ Bei der Befragung von Mathematiklehrenden bewerteten 30 von 40 die Berücksichtigung von Intuitionen (z.B. Schätzungen) der Schüler als besonders wichtig im Stochastikunterricht im Vergleich zu sonstigen Mathematikgebieten.

Die Modellierung erfolgt über den im Labor erfolgreich untersuchten Weg der Übersetzung der Informationen in „natürliche Häufigkeiten“ und die Aufstellung eines Häufigkeitsbaums. Zur Hinführung auf diese Modellierung eignen sich einige Schlüsselfragen, über die am besten von den Schülern selbst (z.B. als Gruppenaufträge) mit Unterstützung des Lehrers nachgedacht wird, z.B.:

- *Was kann passieren, wenn ein HIV-Infizierter getestet wird? Was, wenn ein nicht HIV-Infizierter getestet wird? Schreibe alle Möglichkeiten auf!*
- *Welche Möglichkeiten würdest du als „Fehler des Tests“ bezeichnen und wo lag der Test richtig?*
- *Verteile die Bevölkerung von NRW auf die verschiedenen Möglichkeiten. Wie viele Personen sind es jeweils?*

Eine besondere Schwierigkeit bei der Modellierung von Bayesproblemen kann die Klassifizierung des Ergebnisraums des mehrstufigen Zufallsprozesses sein (vgl. insbesondere beim Drei-Türen-Problem: Wollring, 1992, S.10). Die Schüler sollen in dieser Phase eigenständig erkennen, dass vier mögliche Ergebnisse auftreten können, von denen zwei „Fehler des Tests“ darstellen. Ziel ist die Bestimmung einer Wahrscheinlichkeit für einen „richtig-positiven“ Test, wobei es auch „falsch-positiv“, „richtig-negativ“ und „falsch-negativ“ Testergebnisse gibt.

Im Anschluss erfolgt (am besten fragend-entwickelnd an der Tafel oder OHP) die Modellierung und Visualisierung der Problemstellung mit einem Häufigkeitsbaum (vgl. Abb.IV.1a). Dass alle Menschen aus NRW am Test teilnehmen, ist zugegebenermaßen eine unrealistische Annahme. Die Einschränkung der Realität umgeht hier zunächst die schwierige Frage der Schätzung der Basisrate (Vertiefung erfolgt in Arbeitsblatt 4).

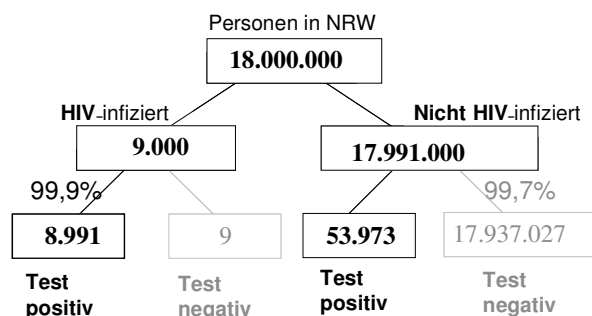
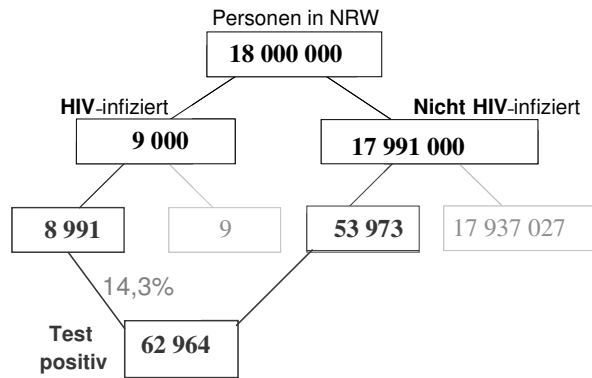


Abb.IV.1a: Modellierung des AIDS-Test Problems im Häufigkeitsbaum

Das Erreichen einer wahrscheinlichkeitstheoretischen Lösung des Problems ist mit dem entwickelten Häufigkeitsbaum, wie ich bereits ausführlich gezeigt habe, bereits jetzt in greifbarer Nähe. Zur Verdeutlichung wird noch die Gesamtzahl aller positiv Getesteten in den Baum integriert (Abb.IV.1b). Die Lösung sollte mit dem klassisch-frequentistischen Vorwissen der Schüler ohne Probleme als Wahrscheinlichkeit interpretierbar sein.



Lösung: „Die Wahrscheinlichkeit für eine HIV-Infektion bei positivem Test“ =
 „Der Anteil der HIV-Fälle unter allen positiv Getesteten“ = $8\,991 / 62\,964 = 14,3\%$

Abb.IV.1b: Wahrscheinlichkeitstheoretische Lösung des AIDS-Test Problems mit dem Häufigkeitsbaum

Nun zeigt sich die besondere Eignung der AIDS-Test-Situation für den Durchlauf des kompletten „Regelkreises der mathematischen Modellierung“. Der Rückbezug der mathematischen Lösung „14,3%“ auf die reale Situation geschieht von selbst, weil sie ganz und gar nicht im Einklang mit der Intuition und den eventuellen Vorerfahrungen der Lernenden steht. Die „Entlarvung“ des eigenen Irrtums – das kann jeder an sich selbst feststellen – bewirkt eine neue Motivation zum aktiven Nachdenken über die mathematische Modellierung, insbesondere den Bezug Welt (Realität) – Mathematik (Modell) (vgl. z.B. Schupp, 1984). Selbst wenn Schüler aufgrund des verblüffenden Ergebnisses noch bezweifeln, dass das benutzte Modell stimmt, schafft das eine wichtige Voraussetzung für vernünftiges Denken, nämlich „[...] dass die Schüler Brücken zwischen ihrem Alltagsdenken und dem von ihnen geforderten mathematischen Denken schlagen können.“ (Heymann, 1995, S.24).

1.2.3 Vertiefung und Erweiterung der Modellierung

Die Modellbildung wie auch die Interpretation und Rückkopplung der mathematischen Lösung mit der Realität (also die Übergänge zwischen Mathematik und realer Welt) stellen erhebliche Anforderungen und stehen immer in Gefahr verkürzt zu werden (vgl. Schupp, im Druck). Deshalb sollen diese beiden Schritte besondere Vertiefung in folgenden Unterrichtsphasen finden (vgl. Arbeitsblätter 2, 4 und 5, Anhang B).

Arbeitsblatt 2: Das AIDS-Testverfahren im Detail

Die Verblüffung über das erhaltene Ergebnis bietet vielfältige Motivation zu kritischem Vernunftgebrauch:

„Warum wird der Test überhaupt durchgeführt, wenn er so „schlecht“ ist?“

Es gilt deutlich zu machen, dass ein positiver Test immerhin einen „Informationsgewinn“ von 0,05% auf 14,3% bewirkt. Wenn aber der Test negativ ausfällt, dann liegt sogar nahezu zu 100% keine HIV-Infektion vor. Die Sichtweise „Informationsgewinn“ unterstützt die vernünftige Interpretation der Ergebnisse

„Gibt es eine Möglichkeit, wie man die „Sicherheit“ für positiv Getestete erhöhen kann?“

In Praxi wird ein ähnlicher Test auf jene nochmal angewandt, die ein positives Testergebnis hatten. So soll die „Sicherheit“ weiter erhöht werden. Diese mehrfache Anwendung verschiedener Tests ergibt weitere Motivation rechnerischer Modellierungen (vgl. auch Kap.I,

Abb.I.5). Wahrscheinlichkeit kann auch hier als „Grad der Sicherheit oder des Vertrauens“ eine völlig neue Interpretation erhalten. Eine subjektivistische Deutung könnte anhand der AIDS-Test-Situation gut entwickelt werden.

Arbeitsblatt 4: Einfluss der Basisrate

*„Woran liegt es eigentlich, dass **ein** positiver Test so wenig aussagt (gegen intuitive Schätzungen) ?“*

Diese naheliegende Frage führt zum zentralen Thema des Einflusses der Basisrate auf das Ergebnis. Aus kognitionspsychologischer Sicht wurde das Erkennen dieses Einflusses als besonders problematisch charakterisiert (z.B. „Base-rate neglect“); jedoch wird diese Schwierigkeit durch den anschaulichen Häufigkeitsansatz entschärft (vgl. dazu ausführlich Kap.I.3.1).

Die Schüler sollen ausgehend von Daten zu verschiedenen „HIV-Risikogruppen“ (z.B. Robert-Koch-Institut, 2002) selbstständig entdecken, dass und wie sich der Vorhersagewert eines Testes ändert (z.B. in Gruppenarbeit, so dass möglichst viele Einzelergebnisse erarbeitet werden). Die Modellierung mit dem Häufigkeitsbaum wird nicht nur vertieft, sondern auch validiert. Eine Vernetzungsmöglichkeit mit der Analysis bietet die funktionale Darstellung des Zusammenhanges.

Eine weitere Darstellungsform, die Häufigkeitstabelle (vgl. Abb.IV.2), erlaubt analoge Modellierungen. Sie kann ebenfalls sehr intuitiv erarbeitet werden. Eine weitere Darstellungsform einzuführen, soll einerseits das Entstehen von „Bearbeitungsroutinen“ vermeiden. Andererseits haben Tabellen eine hohe Benutzungsrelevanz im Alltag bzw. fächerübergreifend in der Schule. Im Vergleich mit der Tabelle wird nun der Häufigkeitsbaum zum „Doppelbaum“ ergänzt (vgl. Abb.IV.3; Kap.II.3.2). Im Folgenden sollen die Schüler beide Modellierungsformen gleichberechtigt verwenden.

	Gesamt	HIV-infiziert	Nicht HIV-infiziert
Gesamt	7 500	848	6 652
TEST positiv	867	847	20
TEST negativ	6633	1	6632

(Bemerkung: Gesamtheit bezieht sich auf „intravenös Drogenabhängige (IVDA) in Berlin“)

Abb.IV.2: Modellierung des AIDS-Test Problems mit der Häufigkeitstabelle

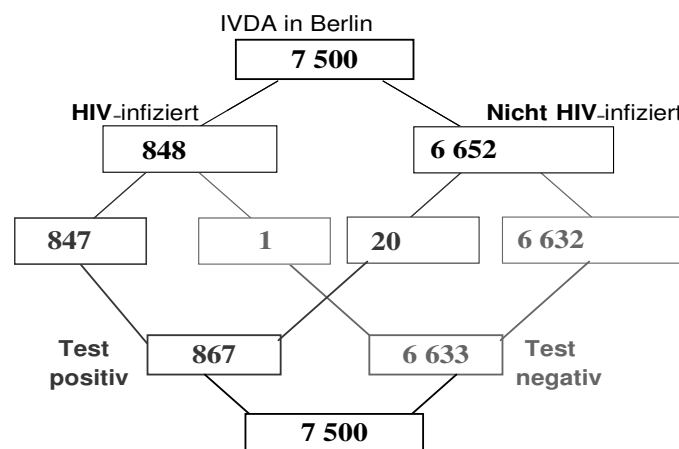


Abb.IV.3: Modellierung des AIDS-Test Problems im vollständigen Häufigkeitsbaum

Arbeitsblatt 5: Das AIDS-Test Problem aus anderer Sicht

Eine weitere Schwierigkeit des stochastischen Denkens tritt erfahrungsgemäß bei Problemen im Zusammenhang mit bedingten Wahrscheinlichkeiten auf: Stochastische Folgerungen sind völlig unabhängig von Kausalzusammenhängen möglich („Falk-Paradoxon“ als Problembeispiel; z.B. Falk, 1979; vgl. auch Kap.II.2.2.4). Beim vorliegenden AIDS-Test Problem wird aus dem Vorliegen des Testergebnisses (Wirkung) diagnostisch auf die Wahrscheinlichkeit einer Erkrankung (Ursache) geschlossen, gleichsam gegen die kausale Richtung.

Umgekehrt ist es ein neues Problem vom Vorliegen der Erkrankung auf die Wahrscheinlichkeit des Testergebnisses zu schließen. Eine realistische Situation ist die Entwicklung eines AIDS-Tests (vgl. Text in Arbeitsblatt 5, Anhang B): Testergebnisse von Versuchspersonen, deren Gesundheitszustand bekannt ist, lassen Schlüsse über die „Güte“ eines neu entwickelten Tests zu. „Doppelbaum“ und Tabellen können dabei als Modelle für beide Schlussrichtungen dienen. Zur Umkehrung kann man den Baum gewissermaßen auch „von unten“ und die Tabelle „in Zeilen“ lesen. Herausgearbeitet werden soll, dass der Formalismus des „Bayesianischen Folgern“ unabhängig von logisch-kausalen Folgerungsrichtungen anwendbar ist.

1.2.4 Vertiefung außermathematischer Kontexte

Arbeitsblatt 3: AIDS-Test positiv und dann?

Am Anfang der Arbeit wurde die Lebensrelevanz des Bayesianischen Denkens betont. Die günstige Gelegenheit, dem mathematischen Tun einen tieferen Sinn zu verleihen, sollte ergriffen werden. Anhand von Medientexten (evtl. begleitend in der Sexualerziehung) soll mit den Schülern die individuelle und gesellschaftliche Bedeutung des Themas AIDS und der Diagnose erörtert werden. Ein positives Testergebnis hat bereits massive Auswirkungen auf das individuelle Leben und auf das Verhalten der gesellschaftlichen Umwelt. In anschließender Diskussion soll die zentrale Bedeutung stochastischen Denkens für solche wichtigen Lebensfragen deutlich werden. Die dafür verwendete Zeit ist meiner Meinung nach keinesfalls „verschwendet“.

Es entstehen Gelegenheiten zu Kommunikation und Abwägen subjektiver Meinungen, die auch Unterstützung durch mathematische Argumente finden können. Z.B. eine Diskussion, ob ein Pflichttest – angesichts einer drohenden Verbreitung – (bevölkerungspolitisch bzw. moralisch) zu vertreten ist; ein Thema, das auch in Deutschland in der öffentlichen Diskussion war (vgl. Böer, 1997, S.38). Vertiefung außermathematischer Kontexte kann zu einer besonderen „Unterrichtskultur“ anregen, die die soziale *und* die inhaltliche Dimension verbindet (vgl. These 5; Heymann, 1995).

Später bietet umfangreiches Material die Möglichkeit, Situationen aus weiteren realistischen Anwendungskontexten auf dieselbe Weise zu behandeln (vgl. Arbeitsblätter 9-12 und Zusatzmaterial; Anhang B).

1.2.5 Verallgemeinerung und Begriffsvertiefung

Trotz der Betonung außermathematischer Anwendungen darf die innermathematische Zieldimension des Mathematikunterrichts nicht außer Acht gelassen werden. Allerdings tritt die hybride Natur der Mathematik (angewandt vs. abstrakt) in ganz besonderer Weise in der Wahrscheinlichkeitstheorie zu Tage (z.B. bereits bei einer Definition von Wahrscheinlichkeit: frequentistisch – axiomatisch; vgl. z.B. Hacking, 1984). Es findet sich bei kaum einer mathematischen Disziplin eine größere Kontroverse um Theorie und Anwendung. Für den

Unterricht mag sich daraus ein Grundsatz ableiten, den mathematischen Kalkül sehr „vorsichtig“ zu entwickeln und kein Wissen auf Vorrat zu produzieren. Dasselbe gilt auch für Begriffe (AK Stochastik, 2002).

Es ist natürlich nicht darüber zu streiten, dass das Entwickeln mathematischer Objekte und Werkzeuge zentraler Bestandteil auch des Stochastikunterrichts sein muss. Eine Frage ist aber, wann und in welchem Umfang? In der Unterrichtsreihe wird in sogenannten „Überblicken“¹⁰⁴ Anwendungswissen verallgemeinert, formalisiert und Begriffe eingeführt (vgl. Anhang, B).

Überblick 1: Wahrscheinlichkeitsbegriffe

In „Überblick 1“ werden ausgehend vom Modell Häufigkeitsbaum die Begriffe „Einzelwahrscheinlichkeit“, „UND-Wahrscheinlichkeit“ und „bedingte Wahrscheinlichkeit“ entwickelt, voneinander abgegrenzt, und jeweils mathematische Sprache, Symbole und Rechenregeln mit Häufigkeiten eingeführt. In zugehörigen Übungen (vgl. Arbeitsblatt 8) sollen die Schüler einerseits mit diesen Objekten vertraut werden, andererseits sie bei neuen Problemstellungen verwenden. Wahrscheinlichkeitstheoretische Formalisierung der Bayesregel erfolgt aus bereits dargelegten Gründen in diesem Unterrichtsgang nicht. Eine formale Behandlung in der Sekundarstufe I schafft aus meiner Sicht eher Verwirrung als Transparenz.¹⁰⁵

Überblick 2: Wahrscheinlichkeiten neu bewerten

Ein anderer verallgemeinernder Aspekt des Bayesianischen Denkens sollte auf jeden Fall in der Sekundarstufe I vorbereitet werden: Die Sichtweise des „Lernen aus Erfahrung“ (Neubewertung von Wahrscheinlichkeiten aufgrund neuer Informationen). Wie bereits deutlich gemacht wurde, dient diese Sichtweise nicht nur der Transparenz der vorliegenden Situationen. Durch die Betonung des hypothetischen Charakters der Wahrscheinlichkeit gewinnt man Anschluss an die beurteilende Statistik. Bei der späteren Behandlung von Hypothesentests in der Sekundarstufe II werden bedingte Wahrscheinlichkeiten als Signifikanz ($P(\text{Daten} \mid \text{Nullhypothese})$) berechnet. Dem Bayesianischen Zugang zur Statistik liegt aber die Neubewertungsdenkweise zugrunde (Berechnung von $P(\text{Hypothese} \mid \text{Daten})$). Diese curriculare Verbindung des Themenbereichs „Bedingte Wahrscheinlichkeit“ und „Testen von Hypothesen“ wurde als didaktischer Vorschlag immer wieder betont (vgl. Kap.II.2.2.2; z.B. Riemer, 1985; Krauss & Wassner, 2001) und auch in Lehrmaterialien für die Sek. I verwirklicht (z.B. Schmid & Weidig, 1996, S.180f.; Kuypers, Lauter & Wuttke, 1999, S.174f.). Als Fortsetzung dieses Vorschlages in der Sekundarstufe II sollten klassische *und* Bayesianische Methoden der beurteilenden Statistik behandelt und durch eine vergleichende Gegenüberstellung über Sinn und Grenzen nachgedacht werden (vgl. z.B. Götz, 2001). In keinem anderen Bereich der Mathematik hat man eine so gute Chance, Schülern einmal vorzuführen, dass es auch innerhalb der Mathematik Streitfragen gibt.

In der Reihe wird anhand eines Kriminalfalles (Indizien dafür, ob jemand der Täter ist; vgl. Arbeitsblatt 7) ein realistischer Kontext bearbeitet, der die Einführung der Neubewertungsdenkweise und der Begriffe („Hypothese“, „Indiz“, „A-priori- und A-posteriori-Wahrscheinlichkeit“; vgl. Überblick 2) gut unterstützt. In den zugehörigen Übungen (vgl. Arbeitsblatt 8) soll die Terminologie auf andere Beispiele angewandt und geübt werden.

¹⁰⁴ Wortwahl: Die Schüler sollen sich einen Überblick über das bisher Erfahrene verschaffen.

¹⁰⁵ Hierzu Schupp (1984, S.238): „Hilfsmittel sollten nur und erst dann gebracht werden, wenn sie beim Schüler wirklich als Hilfen ankommen, weil sie komplexe Geflechte transparent zu machen vermögen“.

1.2.6 Erweiterung des Anwendungsbezuges

Durch den bisherigen Unterrichtsfortgang sollte bei den Schülern nun ein tragfähiges Konzept entwickelt sein, mit Problemstellungen aus dem Bereich bedingte Wahrscheinlichkeit und Satz von Bayes fertig zu werden. Weitere Unterrichtszeit soll für die Erarbeitung weiterer realer Anwendungsbeispiele genutzt werden (verschiedene Themen zur Auswahl: Schwangerschaftstest, Vaterschaftstest, Krebsfrüherkennung, BSE-Krise, Drogenkontrollen im Straßenverkehr; vgl. Arbeitsblätter 9 bis 12, Anhang B).

Anhand dieser Anwendungen sollen die Schüler erworbenes Wissen weiter vertiefen. Dazu gehört etwa in Situationen zu denken, Material zu elaborieren, Fragen zu stellen, Fragestellungen vernünftig zu modellieren, Darstellungsmittel zur Lösungsfindung zu benutzen, numerische Ergebnisse zu interpretieren, die Modellbildung nachträglich zu überdenken und nicht zuletzt sich über die Hintergründe und Bedeutung der Kontexte zu informieren und darüber zu diskutieren. Das Material bietet hierzu umfassend Gelegenheit.

1.2.7 Übung

Es sollte ein Kernpunkt der Unterrichtsreihe sein, Schülern realistische Anwendung von Mathematik zu ermöglichen. Jedoch geht eine derartige Behandlung von Problemstellungen im Stochastikunterricht weit über das hinaus, was üblicherweise als „Kompetenz“ oder „Leistung“ gemessen wird. Es war nötig „geeignete“ Übungs- und Prüfungsaufgaben zu finden, so dass die Wesenszüge des angewandten Problemlösens aber auch die Bedingungen formalisierter Prüfungen beachtet werden (vgl. Blum, 1996, S.26). Deshalb beinhaltet die Unterrichtsreihe auch „kontextärmer“ realistische Übungsaufgaben, die einen Kompromiss insofern darstellen, dass sie im Sinne schulischer Leistungsmessung „prüfungsfähig“ sind, aber grundlegende Anwendungsmerkmale miteinbeziehen (vgl. Arbeitsblätter 13 und 14). Arbeitsblatt 14 enthält dabei Aufgabenstellungen, die auch strukturelle Veränderungen bei der Modellierung im Vergleich zu bisherigen Aufgaben erfordern (Z.B. nicht dichotome Struktur; vgl. Kap.III, Tab.III.1, Itemtyp A3).

2 Ziele und Methoden einer Unterrichtsstudie in der Sekundarstufe I

Die folgenden Teile dieses Kapitels beschreiben eine *Unterrichtsstudie* zur Untersuchung einer Realisierung der entworfenen Unterrichtsreihe „Authentisches Bewerten und Urteilen unter Unsicherheit“ im Stochastikunterricht der gymnasialen Sekundarstufe I. Das Ziel der Unterrichtsstudie war zunächst kein quasi-experimenteller Performanzvergleich mit anderen Klassen, die herkömmlichen Unterricht erhielten, sondern die Exploration des häufigkeitsbasierten Zugangs. Es sollten vor allem Aussagen gewonnen werden, ob und wie das Häufigkeitskonzept im Stochastikunterricht der Sekundarstufe I realisiert werden kann. Die Realisierung wurde videodokumentiert und unter Einbeziehung von Leistungsmerkmalen und Lehrer- und Schülermeinungen analysiert und bewertet. Die Analyse soll weiterführend einer Optimierung des Lehr-Lern-Szenarios dienen bzw. als Ausgangspunkt für vergleichende Untersuchungen.

Teilnehmende Klassen, Ablauf und Erhebungsinstrumente werden im Folgenden ausführlich beschrieben. Anschließend werden Ergebnisse dargestellt und diskutiert.

2.1 Stichprobe

Auf Grundlage der entwickelten Unterrichtsreihe wurden 144 Schüler (fünf 9. Klassen, Gymnasium Nordrhein-Westfalen¹⁰⁶; 71 weiblich, 73 männlich) von ihren jeweiligen Mathematiklehrerinnen und Lehrern im Themengebiet „Anwenden der Bayesschen Regel“ unterrichtet. Gemäß dem nordrhein-westfälischen Lehrplan für die gymnasiale Sekundarstufe I (Kultusministerium NRW, 1993) sollen die Schüler bereits auf folgendes Wissen zurückgreifen können:

Schülerinnen und Schüler ...

- *benutzen relative Häufigkeiten von langen Versuchsreihen zur Schätzung von Wahrscheinlichkeiten,*
- *verwenden ein- oder zweistufige Zufallsversuche zur Simulation zufälliger Erscheinungen in alltäglichen Situationen,*
- *bestimmen Wahrscheinlichkeiten bei einstufigen Zufallsexperimenten mithilfe der LAPLACE-Regel,*
- *bestimmen Wahrscheinlichkeiten bei zweistufigen Zufallsexperimenten mithilfe der Pfadregeln,*
- *nutzen Wahrscheinlichkeiten zur Beurteilung von Chancen und Risiken und zur Schätzung von Häufigkeiten (z.B. in Spielsituationen).*

2.2 Ablauf der Unterrichtsstudie

Der Unterricht erfolgte im 2. Halbjahr des Schuljahres 2002/03. Nach Diskussion und Fertigstellung der Unterrichtsreihe und des Materials (vgl. 1.) wurde der Entwurf den beteiligten Lehrenden in einer Fortbildungsveranstaltung Mitte Februar 2003 detaillierter vorgestellt und diskutiert. Die Inhalte und die Prozessplanung, Arbeitsmaterialien für Schüler (Arbeitsblätter, Medientexte, Übungsaufgaben, Hilfsmittel) und didaktische Hintergründe wurden genau fixiert. Die jeweilige Unterrichtsvorbereitung wurde von den Lehrenden in gegenseitiger Absprache auf Grundlage des Materials durchgeführt. Didaktisch-methodische Entscheidungen im Realisationsprozess waren jeweils in der Verantwortung der Lehrenden als Unterrichtsexperten.

Es wurde vereinbart, dass der Stochastikunterricht gemäß der vorliegenden Unterrichtsreihe in den fünf 9. Klassen des Gymnasiums parallel stattfinden soll. Für die Realisierung standen insgesamt 6 Schulwochen à 3 Schulstunden (Zeitraum 24.2.03 – 4.4.03) zur Verfügung. In diesem Zeitraum erfolgte auch eine außerschulische Sexualberatung, bei der der Kontext AIDS und die Diagnose vertieft wurden. In der letzten Schulstunde wurde zeitlich parallel eine Klassenarbeit (45 min.) geschrieben, die in allen Klassen identisch war. Insgesamt fanden maximal 15 Stunden Unterricht zu diesem Themengebiet statt.

Im Anschluss an den Unterricht wurde von den Schülern ein (nicht-leistungsbezogener) Fragebogen zum zurückliegenden Unterricht bearbeitet. Die Lehrenden wurden ebenfalls schriftlich zum zurückliegenden Unterricht befragt. Eine abschließende Gruppendiskussion mit den Lehrenden erfolgte im Anschluss an die Bearbeitung des Fragebogens. Der Leistungstest zur Überprüfung der Stabilität von Lernerfolgen wurde ca. 14 Wochen nach Beendigung des Unterrichts durchgeführt. Zwischen Unterrichtende und Leistungstest erfolgte kein weiterer Unterricht in Stochastik.

¹⁰⁶ Die Stichprobe bestand aus der gesamten 9. Jahrgangsstufe des Schuljahres 2002/03 des Freiherr-vom-Stein Gymnasiums in Bünde (Westfalen), an dem der Mitautor der Unterrichtsreihe Stefan Schweynoch als Mathematiklehrer unterrichtet.

Stichprobe	Unterricht	Klassenarbeit	Befragungen	Leistungstest
Fünf 9.Klassen, Gymnasium NRW (N=144)	ca. 15 Std. Beobachtung (Videoprotokolle) (+ 2 Std. Sexual- beratung)	45 Min. parallel in allen Klassen unmittelbar nach Beendigung der Unterrichtsreihe	Schülerfragebogen Lehrerfragebogen Gruppendiskussion	14 Wochen nach Unterrichtsreihe

Tab.IV.2: Ablauf der Unterrichtsstudie

2.3 Erhebungsinstrumente

2.3.1 Klassenarbeit

Am Ende der Unterrichtsreihe mussten sich alle Schüler einer Leistungsfeststellung mittels einer (für alle Klassen identischen) Klassenarbeit unterziehen, die in Absprache mit allen teilnehmenden Lehrern entwickelt wurde. Die Inhalte der Arbeit wurden so abgestimmt, dass sie die geforderten Kompetenzen umfassten (sachliche Bezugsnorm). Diese klassenübergreifende Leistungsprüfung musste auf die Leistungsfähigkeit der einzelnen Klassen (soziale Bezugsnorm) Rücksicht nehmen, deren Einschätzung der diagnostischen Kompetenz der Lehrenden unterlag (vgl. Schrader & Helmke, 2001).

Die Klassenarbeit bestand aus zwei kontextuell verschiedenen Aufgaben¹⁰⁷, die jeweils in 5 Teilaufgaben (a bis e) untergliedert waren. Die erwarteten Schülerleistungen und die Kriterien für die Bewertung wurden für jede Teilaufgabe in einer Beschreibung eines Erwartungshorizontes („Musterlösung“) genau festgelegt. In Anhang C findet sich die Klassenarbeit mit den jeweils erwarteten Leistungs- und Bewertungskriterien.

Um eine differenzierte Aufgabenstellung zu erreichen, wurden die Teilaufgaben so konstruiert, dass sie gemäß der Forderung in gymnasialen Richtlinien für Abiturprüfungsanforderungen drei Bereiche mit insgesamt steigenden Anforderungen abdecken (vgl. z.B. Kultusministerium NRW, 1999, S.70f.; Stern & Hardy, 2001, S.159):

- **Faktenwissen:** Reproduktion von aus dem Unterricht bekannten Standardverfahren (Anforderungsbereich 1), z.B. „Bestimmen von Wahrscheinlichkeiten in einfachen, vom Unterricht her vertrauten Zusammenhängen“
- **Anwendungswissen:** Anwenden der Kenntnisse auf andere Situationen (Anforderungsbereich 2), also „selbstständiges Übertragen des Gelernten auf vergleichbare neue Situationen, wobei es entweder um veränderte Fragestellungen oder um veränderte Sachzusammenhänge oder um abgewandelte Verfahrensweisen gehen kann“
- **Problemlösewissen:** Planmäßiges Verarbeiten komplexer Gegebenheiten mit dem Ziel zu selbstständigen Lösungen, Gestaltungen oder Deutungen etc. zu gelangen (Anforderungsbereich 3). „Dabei werden aus den gelernten Methoden oder Lösungsverfahren die zur Bewältigung der Aufgabe geeigneten selbstständig ausgewählt oder einer neuen Problemstellung angepasst“ (S.72, ebenda)

Die Gewichtung wurde entsprechend den Lehrplanvorschlägen vorgenommen. Tabelle IV.3 zeigt die Verteilung der Teilaufgaben auf die Anforderungsbereiche und die jeweilige Bewertungsgewichtung.

¹⁰⁷ Mehr als zwei Kontexte wären im vorgegebenen Zeitrahmen von 45 min. nicht möglich gewesen.

AFB 1		AFB 2		AFB 3	
Teilaufgabe	Punkte	Teilaufgabe	Punkte	Teilaufgabe	Punkte
1a	7	1d	6	1e	5
1b	3	2a	9		
1c	6	2b	4		
2c	3	2d	3		
		2e	2		
Summe	19 (40%)		24 (50%)		5 (10%)

Tab.IV.3: Verteilung der Teilaufgaben der Klassenarbeit auf Anforderungsbereiche (erreichbare Punkte und Anteil an der Gesamtpunktzahl)

2.3.2 Kompetenztest zur Überprüfung der Nachhaltigkeit des Lernerfolges

Das wichtige Kriterium der Nachhaltigkeit von Lernerfolgen sollte ebenfalls untersucht werden. Zu diesem Zweck und der Möglichkeit einer Verankerung mit entsprechenden Laborergebnissen (Kap.III.3) absolvierten alle Schüler *nach 14 Wochen* den Kompetenztest zum Bayesianischen Denken, der in den Trainingsstudien mit Schülern der Sekundarstufe I entworfen wurde (vgl. zum Testaufbau detailliert Kap.III.3.2.2.).¹⁰⁸

Die Schüler hatten in der Zeit zwischen dem Modellunterricht und dem Leistungstest keinen weiteren Stochastikunterricht. Hinsichtlich der kontextuellen Einkleidung der Problemstellungen unterschieden sich Testitems und im Unterricht behandelte Aufgabenstellungen, so dass alle Testitems auch die Fähigkeit einer Übertragung des Gelernten auf veränderte Sachzusammenhänge (im Sinne von AFB 2, s. oben) erforderten. Reines Reproduktionswissen war nicht Testinhalt.

2.3.3 Schülerbefragung nach Unterricht

Zur Erfassung nicht-leistungsbezogener Merkmale und Aussagen der Schüler zum Modellunterricht wurde ein Fragebogen entwickelt. Der Fragebogen enthielt Items zur Selbstbewertung und lernmotivationale Aspekte wie spezielles Sachinteresse, generelle Tätigkeits- und Folgenanreize (z.B. „Spaß“, „gute Noten“ etc.), Einschätzung der Wirksamkeit der eigenen Lernhandlung bezogen auf das Ergebnis (Situation- bzw. Handlung-Ergebnis), Selbststeuerung(-sprobleme) etc. Die Fragebogenitems (geschlossenes Antwortformat mit 4-stufiger Skala: 1 = volle Zustimmung, ..., 4 = volle Ablehnung) bezogen sich generell auf den erfolgten Unterricht, spezieller auf behandelte Problemsituationen und -kontexte und die Verwendung von Darstellungsformen. Abschließend sollten die Schüler auch die eigene Präferenz für die beiden vorkommenden Darstellungsformen Baumdiagramm und Tabelle bewerten. Am Ende hatten sie Gelegenheit, im offenen Antwortformat positive und negative Eindrücke über den zurückliegenden Unterrichtszeitraum zu äußern (im Detail vgl. Ergebnisdarstellung in 3.2.).

¹⁰⁸ Aufgrund unterrichtstechnischer Gegebenheiten war nur *eine* Leistungsmessung zusätzlich zur üblichen Klassenarbeit möglich. Der spätere „Nachhaltigkeitstest“ wurde gewählt, da mit der Klassenarbeit bereits eine unmittelbare Leistungsmessung erfolgte und die Befunde zu längerfristigem Erfolg relevanter eingeschätzt wurden.

2.3.4 Lehrerbefragung nach Unterricht

Zur Erfassung der Meinung und Bewertung der beteiligten Lehrenden zu ausgewählten Aspekten des durchgeführten Unterrichts wurde ein Fragebogen entwickelt. Der Hauptzweck des Fragebogens war eine unter speziellen Kriterien gezielte Beurteilung des erfolgten Unterrichts. Fragen bezogen sich auf Lehrinhalte bzw. -material und die Auswirkung auf Unterrichts- und Schülermerkmale. Es wurden insbesondere die Kernaspekte der Studie hervorgehoben: Die Rolle der Darstellungsmittel bzw. numerischen Formate und des Realitätsbezuges. Abschließend erfasst der Fragebogen noch einige personenbezogene Daten. Geschlossene Antwortformate (Rating, Multiple Choice) sind mit offenen Formaten ergänzt (z.B. „Gab es andere Besonderheiten bzgl. des Schülerverhaltens?“).

Der Fragebogen sollte auch als Anstoß zur weiteren Reflexion über den realisierten Unterricht dienen. Im Rahmen einer abschließenden Gruppendiskussion mit allen Beteiligten an der Schule konnten die Äußerungen in den Fragebögen erweitert und spezifiziert bzw. weitere Kommentare und Meinungen hinzugefügt werden. Diese wurden schriftlich protokolliert.

2.3.5. Beobachtung des Unterrichts

Der Unterricht wurde teilweise (in einer Klasse vollständig) per Videoaufzeichnung protokolliert und teilweise wurden schriftliche Verlaufs- und Inhaltsprotokolle erstellt.¹⁰⁹ Im Rahmen dieser Arbeit wurden anhand der Videoprotokolle ausgewählte Teile des Unterrichts *qualitativ* analysiert (im Sinne einer Ergänzung quantitativer Untersuchungsmethoden). Hauptsächlich sollte die Verwendung des didaktischen Häufigkeitskonzeptes anhand der Unterrichtsrealität genauer überprüft werden. Die Beobachtungsdaten dienten zunächst einer Beschreibung und Illustrierung des tatsächlich erfolgten Unterrichts, insbesondere von Schwierigkeiten oder Abweichungen von intendierten Zielen bzw. Prozessen bei der Realisierung. Es sei darauf hingewiesen, dass im Rahmen des Forschungsgegenstandes dieser Arbeit keine detailliertere qualitative Unterrichtsanalyse mit interpretativen oder semiotischen Ansätzen erfolgen sollte (vgl. für eine Übersicht in Birkner-Ahsbahs et al., 2003). Diese Zielsetzung wäre gegebenenfalls einer weiterführenden Analyse der vorliegenden Beobachtungsdaten vorbehalten (z.B. mit dem semiotischen Ansatz von Godino & Batanero, 1998).

¹⁰⁹ Eine vollständige Protokollierung der Unterrichtsstunden aller Lehrenden konnte nicht erfolgen, da keine weiteren Hilfskräfte zur Verfügung standen.

3 Ergebnisse der Unterrichtsstudie

Im Folgenden werden zunächst leistungsbezogene Ergebnisse der Klassenarbeit und des späteren Leistungstests dargestellt und diskutiert. Im Anschluss wird der Unterricht anhand der Äußerungen der Schüler und Lehrenden sowie ausgewählten Beobachtungsdaten beschrieben und qualitativ analysiert.

3.1 Klassenarbeit

3.1.1 Leistungsverteilung

Die Leistungen der 144 Schüler in der Klassenarbeit wurden anhand der in Anhang C dargestellten Kriterien bewertet, die mit den Lehrenden vorher festgelegt wurden. Die Bewertung erfolgte jeweils für jede Klasse von den jeweiligen Lehrenden. Unabhängig davon wurden alle Klassenarbeiten auch von mir selbst ausgewertet. Ich beziehe mich bei der folgenden Analyse auf *die eigene Auswertung* aus folgenden Gründen:

1. Eine einheitliche Beurteilung durch eine Person vermeidet ungewollte interpersonelle Unterschiede.
2. Die Lehrenden konnten nur innerhalb der eigenen Klasse vergleichen (klasseninternes vs. klassenübergreifendes Bezugssystem).
3. Die eigene Bewertung differenzierte stärker nach der Art des mathematischen Arbeitens (z.B. rechnerische od. begriffliche Modellierung und rein technische Fähigkeiten wie Prozentrechnen sind getrennt analysiert).

In der Klassenarbeit konnte eine *maximale Summe* von 48 *Bewertungspunkten* erreicht werden. Der folgende Boxplot (Abb.IV.4 a) zeigt die Verteilung über die Punkteskala und Tab.IV.4 enthält Kennwerte. Der Median der erreichten Bewertungspunkte lag bei 40 (83,3% der erreichbaren Gesamtpunktzahl), das erste bzw. das dritte Quartil lagen bei 35 bzw. 44 Bewertungspunkten.

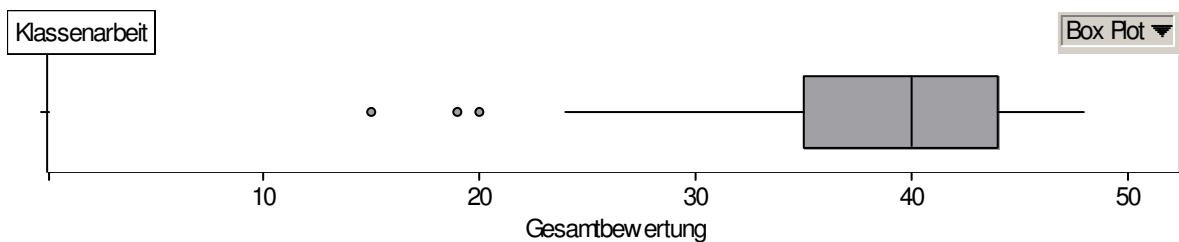
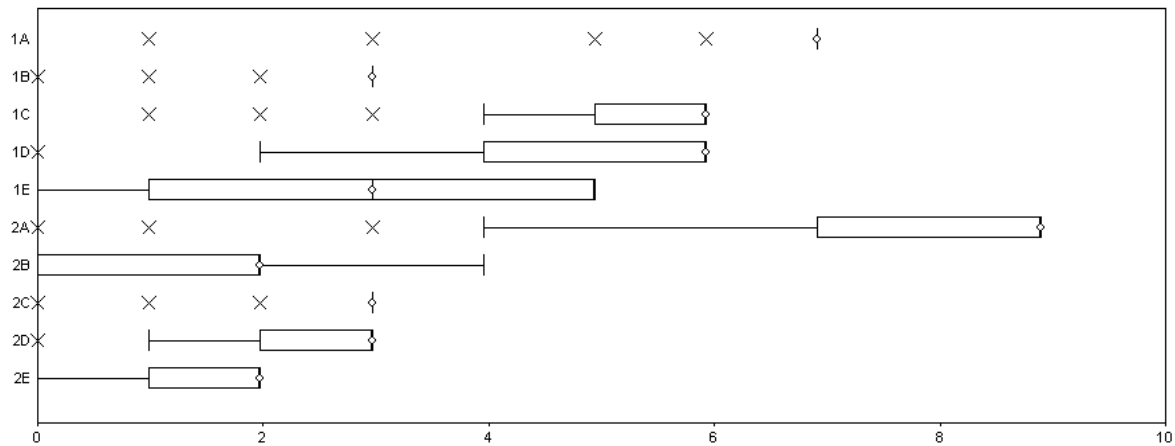


Abb.IV.4: Gesamtleistungsverteilung in der Klassenarbeit (in erreichten Bewertungspunkten)

Variable	Gesamtbewertung	SummeAufg1	SummeAufg2
Mittelwert	38,8	22,8	16,1
Median	40,0	23,0	18
Modus	42	27	19
Minimum	15	6	0
Maximum	48	27	21
Standardabweichung	6,54	3,89	4,10
Erstes Quartil	35,0	21,0	14,0
drittes Quartil	44,0	26,0	19,0

Tab.IV.4: Kennwerte der Leistungen in der Klassenarbeit, gesamt und differenziert nach Aufgaben

Die jeweilige Verteilung der Bewertungspunkte pro Teilaufgabe geben Tab.IV.5 und Abb.IV.5 wieder.



Punkte	1a	1b	1c	1d	1e	2a	2b	2c	2d	2e
0	0	3	0	7	6	2	43	6	19	32
1	1	3	4	0	44	2	11	3	5	9
2	0	14	5	1	16	0	73	4	20	103
3	2	124	5	17	17	5	1	131	100	-
4	0	-	15	16	20	4	16	-	-	-
5	8	-	15	13	41	5	-	-	-	-
6	3	-	100	90	-	15	-	-	-	-
7	130	-	-	-	-	4	-	-	-	-
8	-	-	-	-	-	10	-	-	-	-
9	-	-	-	-	-	97	-	-	-	-
Gesamt	144	144	144	144	144	144	144	144	144	144

Abb. u. Tab.IV.5: Leistungsverteilung in der Klassenarbeit differenziert nach Teilaufgaben (vgl. Anhang C)

Die Teilaufgaben waren Anforderungsbereichen zugeordnet (vgl. oben). Die Leistungsverteilung, nach Anforderungsbereichen differenziert, geht aus den Kennwerten in Tab.IV.6 hervor. Der Median der erreichten Bewertungspunkte im AFB 1 lag bei der maximalen Gesamtpunktzahl 19, im AFB 2 bei 19,5 (von der erreichbaren Gesamtpunktzahl 24 = 81,25%), im AFB 3 bei 3 (von der erreichbaren Gesamtpunktzahl 5 = 60%). Die Percentil-Plots zeigen (Abb.IV.6a-c), dass im AFB 1 über 60% die Maximalleistung erreicht hatten und im AFB 2 und AFB 3 entsprechend mehr Leistungsdefizite auftreten.

Variable	AFB1	AFB2	AFB3
Anzahl	144	144	144
Mittelwert	17,7	18,3	2,9
Median	19,0	19,5	3,0
Modus	19	22	1
Minimum	6	4	0
Maximum	19	24	5
Standardabweichung	2,36	4,51	1,73
Erstes Quartil	17,0	16,0	1,0
drittes Quartil	19,0	22,0	5,0

Tab.IV.6: Kennwerte der Leistungen in der Klassenarbeit, differenziert nach Anforderungsbereichen (AFB 1-3)

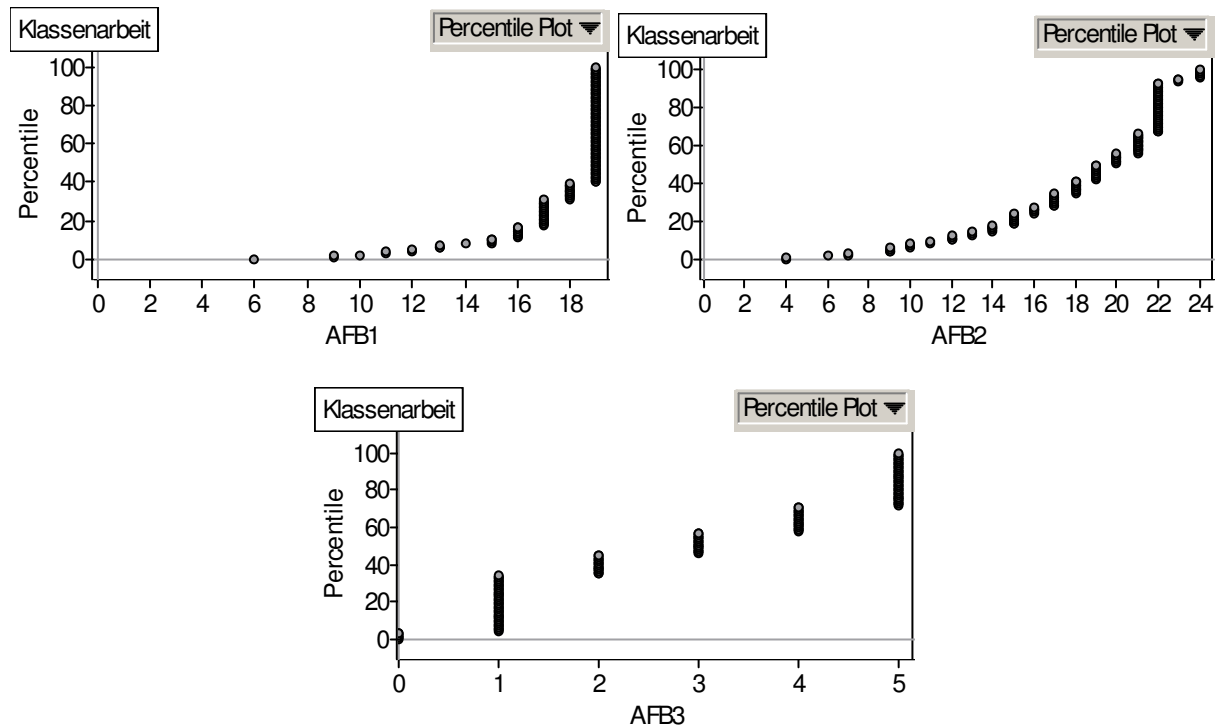


Abb.IV.6a-c: Leistungsverteilungen in der Klassenarbeit differenziert nach Anforderungsbereichen (AFB 1-3; in erreichten Bewertungspunkten)

Eine Differenzierung der Leistungen nach Geschlecht zeigt folgende Abb.IV.7. Der Median lag für Mädchen und Jungen gleichermaßen bei 40. Doch zeigte sich für die Jungen eine stärkere Streuung der Leistungen: Es gab mehr „sehr schlechte“, aber auch mehr „sehr gute“ Jungen als Mädchen.

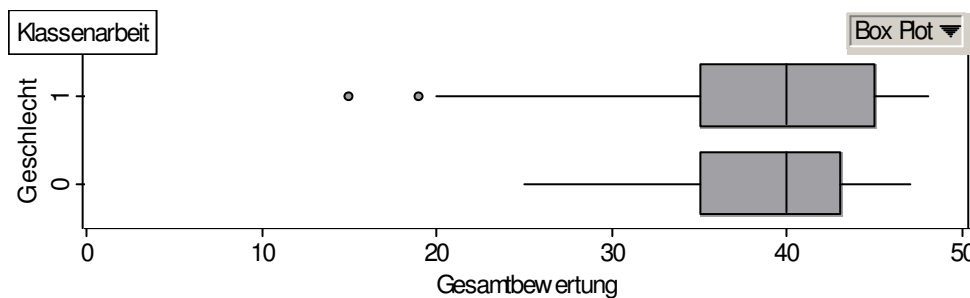


Abb.IV.7: Leistungsverteilung in der Klassenarbeit differenziert nach Geschlecht (0: Mädchen; 1: Jungen)

3.1.2 Benotung

Die Leistungsbeurteilung durch das eingeführte Bewertungssystem erfolgte nach inhaltlich verankerten und lernzielorientierten Kriterien (sachliche Bezugsnorm). Die eigene Bewertung war anonym und klassenübergreifend, so dass soziale und individuelle Bezugsnormen nicht gegeben waren. Schulische Zensuren unterliegen hingegen oft (bewusst) weiteren Bezugsnormen. Sie sind deshalb für die Bewertung des sachbezogenen Lernerfolgs nicht immer geeignet. Bei der Zensurenbestimmung ist es generell schwierig, „kriteriale“ Maßstäbe anzulegen, da der einzige inhaltliche „Anker“ meist nur die durchschnittliche Erwartung ist. Es besteht also die Gefahr, dass sie nur unzureichend inhaltlich verankert sind, sondern eher über den Vergleich innerhalb von Schülergruppen (soziale Bezugsnorm) oder über

individuelle Lernfortschritte (individuelle Bezugsnorm). Allerdings hat die rein sachliche Bezugsnorm auch Nachteile: Sie gibt keine Auskunft darüber, ob die fraglichen Kriterien schwer oder leicht zu erreichen waren und ob deshalb auf besondere Fähigkeiten und/oder Lernerfolge bei Schülern geschlossen werden kann. Dazu benötigt man den Vergleich mit anderen (vgl. Rheinberg, 2001, S.68). Ich diskutiere die in der Klassenarbeit durch die Lehrenden vergebenen Zensuren lediglich als Anhaltspunkte für eine Lernerfolgsbewertung bei den untersuchten Schülern und verweise auf die kontroverse Diskussion zur schulischen Leistungsbewertung (vgl. ausführlich in Weinert, 2001, Hrsg.).

Ein Vergleich setzt voraus, dass die Note „3“ in Mathematik im Zeugnis einen Schüler mit „durchschnittlichen“ Fähigkeiten in Mathematik kennzeichnet, entsprechend eine „2“ bzw. „1“ einen „überdurchschnittlichen“ und eine „4“ bzw. „5“ einen „unterdurchschnittlichen“ Schüler. Weiterhin sollte gelten, dass Schüler, die in der Klassenarbeit eine durchschnittliche Bewertung erhalten, auch etwa durchschnittlich kompetent im Sinne der definierten Ziele sind. Wenn man von diesen Annahmen ausgeht, zeigt sich eine deutlich „bessere Leistung“ in der Klassenarbeit im Vergleich zum Leistungsbild im letzten Zeugnis in Mathematik (vgl. Abb.IV.8). Leistungsdifferenziert ist aus der Kontingenztafel (Tab.IV.8) zu entnehmen, dass in Mathematik „überdurchschnittliche“ Schüler (Zeugnisnote 1 und 2) weiterhin zum größten Teil überdurchschnittliche Leistung in der Klassenarbeit zeigten. „Durchschnittliche“ Schüler (Zeugnisnote 3) verbesserten sich aber in der Klassenarbeit deutlich in den überdurchschnittlichen Leistungsbereich. „Unterdurchschnittliche“ Schüler (Zeugnisnote 4 und 5) zeigten ebenfalls eine deutliche Verbesserung zu durchschnittlichen bzw. sogar überdurchschnittlichen Leistungen.

KA_Lehrerbewert		Summary Table					
⇩	⇨	Klassenarbeit_Note					Row Summary
		1	2	3	4	5	
Zeugnis_Mathenote	1	4	5	1	0	0	10
	2	12	17	2	0	1	32
	3	5	17	14	3	1	40
	4	5	14	17	9	6	51
	5	0	1	3	6	1	11
Column Summary		26	54	37	18	9	144

S1 = count ()

Tab.IV.8: Verteilung von Zensuren in der Klassenarbeit, differenziert nach Zensuren in Mathematik im letzten Zeugnis (absolute Häufigkeiten)

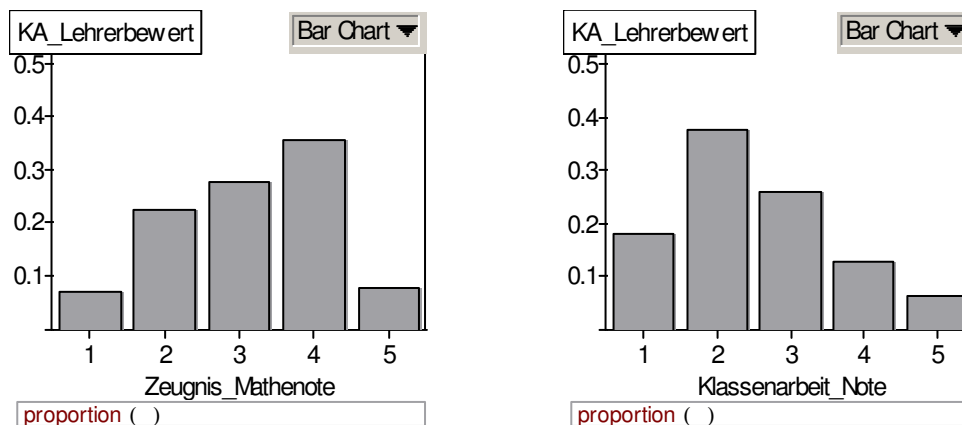


Abb.IV.8: Verteilung von Zensuren in Mathematik im letzten Zeugnis und in der Klassenarbeit im Vergleich (relative Häufigkeiten)

3.1.3 Leistungen differenziert nach Modellierungsschritten

Bereits bei Trainingsstudien wurde deutlich, dass ein Zugang zur differenzierteren Beurteilung von „Kompetenzen“ über die genauere Analyse der Modellierungsprozesse bei außermathematischen Problemstellungen führt. Auch Aufgaben in der Klassenarbeit können differenziert nach Modellierungsschritten (vgl. Kap.III.3.3.1) bewertet werden. Der Unterricht basierte wesentlich auf der Verwendung von Häufigkeitsbäumen und -tabellen zur (grafischen) Modellierung der Problemsituationen. Deshalb lagen grundlegende lernbereichsbezogene Anforderungen im Erstellen und Verwenden der grafischen Darstellungen. Für typische Bayesprobleme wurden bereits in Kap.III.3.3.1 Modellierungsschritte bei Verwendung von Häufigkeitsbäumen näher erläutert. Analog zu dieser Klassifikation sollen nun auch Leistungen in der Klassenarbeit differenziert werden:

- Darstellen in Häufigkeitsbäumen bzw. Häufigkeitstabellen (vgl. „*Mathematisieren i.e.S. und Strukturieren*“, Kap.III.3.3.1)
- Lösen mit Häufigkeitsbäumen bzw. Häufigkeitstabellen (vgl. „*Verknüpfen*“)
- Technische Fähigkeiten (vgl. „*Berechnen*“)
- Wahrscheinlichkeitsbezeichnung¹¹⁰

a) Darstellen in Häufigkeitsbäumen und Häufigkeitstabellen

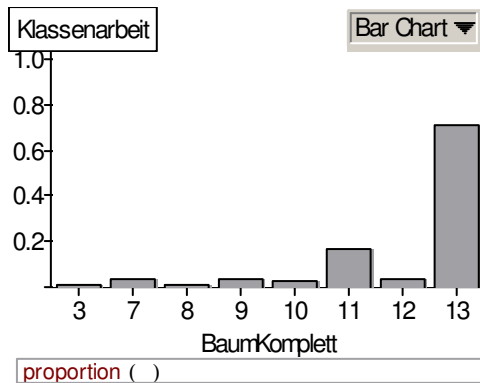
Zur Modellbildung mussten die relevanten Informationen aus dem Aufgabentext herausgefiltert und in Häufigkeiten übersetzt werden. Anschließend waren Strukturen im Baumdiagramm bzw. der Tabelle festzulegen (etwa durch Beschriften einzelner Felder) und die gegebenen Werte einzusetzen bzw. fehlende Werte zu ergänzen.

Das Darstellen in Häufigkeitsbäumen bei den Aufgaben 1a und 2a¹¹¹ erbrachte maximal 13 Bewertungspunkte (Abb. und Tab.IV.9a), Darstellen in Häufigkeitstabellen bei Aufgabe 2c maximal 3 Bewertungspunkte (Abb. und Tab.IV.9b). Über 72% der Schüler erreichten die volle Punktzahl beim Erstellen von Häufigkeitsbäumen und alle (bis auf einen) Schüler lagen mit 7 Punkten über der Hälfte der erreichbaren Bewertungspunkte. Die Häufigkeitstabelle wurde von 91% völlig korrekt ausgefüllt. Im Gegensatz zu Baummodellierungen waren jedoch die Strukturen („Beschriftung“) und teilweise bereits

¹¹⁰ war in den vorherigen Tests nicht als Bewertungskriterium enthalten.

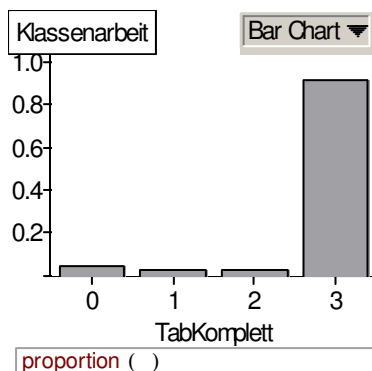
¹¹¹ 141 Modellierungen mit Baum. Nur 3 Schüler haben mit Tabellen modelliert (Leist.: 4, 4, 6 Pkt.).

Häufigkeitswerte vorgegeben (vgl. Aufgabe 2c, Klassenarbeit). Es handelte sich also nur um eine Ergänzung einer Häufigkeitstabelle. Jedoch zeigten auch immerhin 6 Schüler hier keine bewertbaren Leistungen, was bei Baummodellierung nicht vorkam.



BaumKomplett	Hfgk.	Proz.
0	0	0,00
1	0	0,00
2	0	0,00
3	1	0,71
4	0	0,00
5	0	0,00
6	0	0,00
7	5	3,55
8	1	0,71
9	4	2,84
10	3	2,13
11	21	14,89
12	4	2,84
13	102	72,34
Gesamt	141	100%

Abb. u. Tab.IV.9a: Leistungsverteilung beim Darstellen in Häufigkeitsbäumen („BaumKomplett“)



TabKomplett	Hfgk.	Proz.
0	6	4,17
1	3	2,08
2	4	2,78
3	131	90,97
Gesamt	144	100%

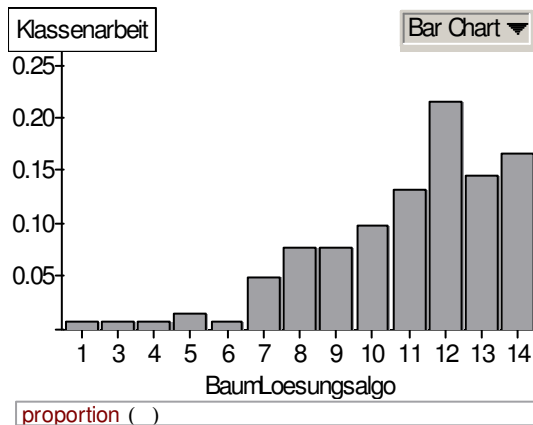
Abb. u. Tab.IV.9b: Leistungsverteilung beim Darstellen in Tabellen („TabKomplett“)

b) Lösen mit Häufigkeitsbäumen und Häufigkeitstabellen

Aus den gebildeten Modellen Häufigkeitsbaum bzw. Häufigkeitstabelle war schließlich ein „Lösungsterm“ zur Berechnung der gesuchten Wahrscheinlichkeit aufzustellen. Dieser Schritt erforderte Verknüpfung bereits errechneter Teilergebnisse, im Falle der Häufigkeitsmodellierung die Bildung eines Quotienten.

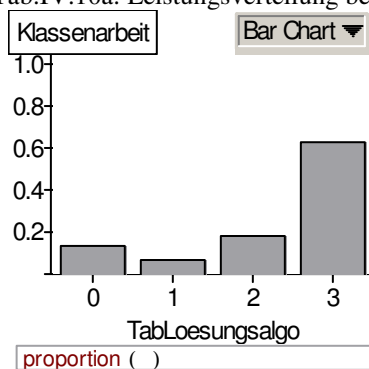
Dieser Lösungsschritt mit Bäumen bei den Aufgaben 1b, 1c, 1d, 2a¹¹² erbrachte maximal 14 Bewertungspunkte, mit Tabellen bei den Aufgaben 2d und 2e maximal 3 Bewertungspunkte. Zunächst wird deutlich, dass hier höhere Leistungsdefizite als bei der Modellbildung auftreten (vgl. Abb.IV.10a,b mit Abb.IV.9a,b). Bei der Lösung mit Tabellen zeigten immerhin 18 Schüler keine bewertbaren Leistungen, mit Bäumen aber kein Schüler (vgl. Tab.IV.10). Im unteren Drittel der Leistungsskala liegen bei Baumdarstellung ca. 2% der Schüler im Vergleich zu 18,75% bei Tabellendarstellung.

¹¹² 141 Modellierungen mit Baum. Nur 3 Schüler haben mit Tabellen modelliert (Leistung: 4, 7, 12 Pkt.).



BaumLösungsalgo	Freq.	Proz.
0	0	0,00
1	1	0,71
2	0	0,00
3	1	0,71
4	0	0,00
5	2	1,42
6	1	0,71
7	6	4,26
8	11	7,80
9	11	7,80
10	14	9,93
11	19	13,48
12	30	21,28
13	21	14,89
14	24	17,02
Gesamt	141	100%

Abb. u. Tab.IV.10a: Leistungsverteilung beim Lösen mit Häufigkeitsbäumen („BaumLösungsalgo“)



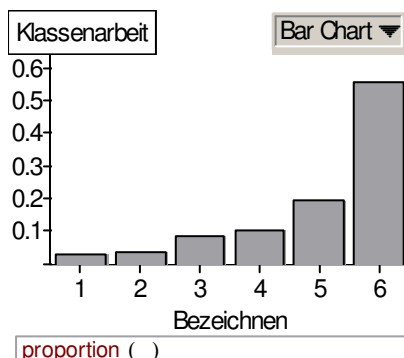
TabLösungsalgo	Freq.	Proz.
0	18	12,50
1	9	6,25
2	26	18,06
3	91	63,19
Gesamt	144	100%

Abb. u. Tab.IV.10b: Leistungsverteilung beim Lösen mit Tabellen („TabLösungsalgo“)

c) Bezeichnen von Wahrscheinlichkeiten

Kompetenzerwartungen der Unterrichtsreihe betrafen auch das Unterscheiden und Bezeichnen von Wahrscheinlichkeiten (vgl. Überblick 1, IV.1.2.5.). Insofern wurde auch in der Klassenarbeit von den Schülern das exakte Bezeichnen der Ergebnisse verlangt. Dabei lag die Bewertung nicht auf der Exaktheit mathematischer Symbolik, sondern auf einer eindeutigen Abgrenzung.

Insgesamt konnten bei den Aufgaben 1b, 1d, 2a, 2d und 2e maximal 6 Bewertungspunkte durch Bezeichnen der Lösungswahrscheinlichkeiten erreicht werden (vgl. Abb. und Tab.IV.11). 55,6% der Schüler konnten stets korrekt die gesuchten Lösungswahrscheinlichkeiten bezeichnen und zeigten deutliches begriffliches Wissen über vorkommende Wahrscheinlichkeiten.



Bezeichnen	Freq.	Proz.
0	0	0,00
1	4	2,78
2	5	3,47
3	12	8,33
4	15	10,42
5	28	19,44
6	80	55,56
Gesamt	144	100%

Abb. u. Tab.IV.11: Leistungsverteilung beim Bezeichnen der Lösungswahrscheinlichkeiten

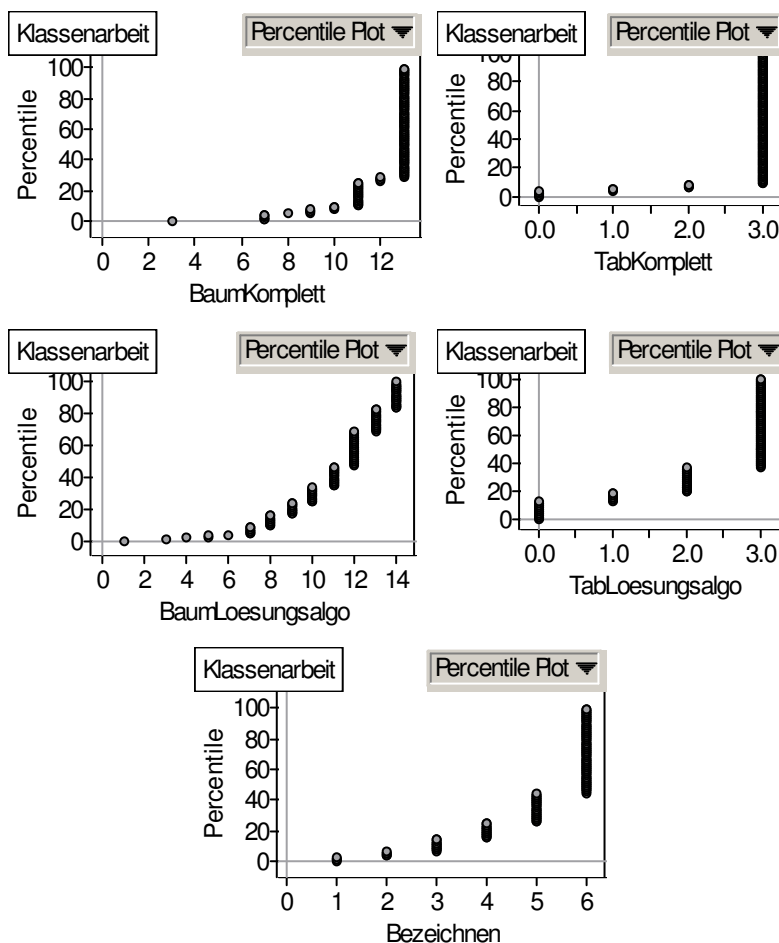
d) Berechnen (technische Fähigkeiten)

Schulische Kompetenzerwartungen betreffen auch rein algebraische Fähigkeiten des Umformens und Berechnens, z.B. eines Häufigkeitsquotienten in einen Prozentwert oder einen Dezimalbruch. Ich berücksichtigte jedoch in der eigenen Bewertung der Klassenarbeit reine technische Fehler nicht, wenn bereits ein korrekter Lösungsterm angegeben war. Der Grund lag in einer gewünschten Trennung von Modellierungskompetenz und rein technischen Fertigkeiten.

Eine getrennte Analyse rein technischer Lösungsschritte ergibt, dass bei 1130 Berechnungen bzw. Umformungen (meist Häufigkeitsquotient in Prozentwert) nur 23 (= 2,04%) fehlerhaft vorgenommen wurden. 166-mal wurde die gesuchte Wahrscheinlichkeiten nicht als numerischer Wert berechnet bzw. erfolgte keine Umformung. Insgesamt ergibt sich somit eine Leistungsrate bei „technischen Schritten“ von 85,4%.

e) Vergleich der Modellierungsschritte

Schritte bei der Modellierung mit Häufigkeitsbäumen bzw. -tabellen vergleichend, ergaben sich beim Bilden der Lösungsterme aus dem erstellten Modell („BaumLoesung“ und „TabLoesung“) deutlich geringere durchschnittliche Leistungen (77 bzw. 79%; vgl. Tab.IV.12) als bei der Modellbildung selbst („BaumKomplett“ und „TabKomplett“ jeweils ca. 94%; vgl. auch „Percentil-Plots“ in Abb.IV.12). Auch beim Bezeichnen von Wahrscheinlichkeiten („Bezeichnen“) waren vergleichsweise größere Leistungsdefizite (84,5%) zu erkennen.



Modellierungsschritt	Durchschnittl. Leistung*
BaumKomplett	93,9%
TabKomplett	93,5%
BaumLoesungsalgo	79,4%
TabLoesungsalgo	77,3%
Bezeichnen	84,5%

* als Anteil an den jeweils max. erreichbaren Bewertungspunkten

Variablenbezeichnungen:

BaumKomplett bzw. *TabKomplett*: Darstellen in Häufigkeitsbäumen bzw. -tabellen

BaumLoesungsalgo bzw. *TabLoesungsalgo*: Lösen mit Häufigkeitsbäumen bzw. -tabellen

Bezeichnen: Bezeichnen von Wahrscheinlichkeiten

Abb. u. Tab.IV.12: Leistungsverteilungen und durchschnittliche Leistungen bei verschiedenen Modellierungsschritten im Vergleich

3.1.4. Qualitative Ergänzungen der Leistungsanalyse

a) Lösungsstrategien

Der Befund von überproportionalen Leistungsdefiziten beim Bilden von Lösungstermen deutet auf nicht der Theorie entsprechende „Lösungsstrategien“ hin, die in diesem Zusammenhang auch in vorherigen Studien beobachtet wurden (vgl. Kap.I.4.1.1.; Gigerenzer & Hoffrage, 1995, „non-Bayesian algorithms“). Ich möchte im Falle der Klassenarbeit keine vollständige Analyse aller vorkommenden Strategien ansprechen, sondern vielmehr sich wiederholende und bei Schülern häufiger auftretende. Eine ausführlichere Betrachtung von Lösungsstrategien erfolgt auch bei der Analyse des späteren Kompetenztests (vgl. IV.3.1.2.).

Der häufigste Fehler bei der Lösung von dichotomen Bayesproblemen, also der Berechnung von $P(H | D)$ ¹¹³, war die Verwendung des Häufigkeitsquotienten $|H| / |D|$. Im Zähler steht anstatt der konjunktiven Häufigkeit von $|H \cap D|$ die Häufigkeit für das Ereignis H. Diese Fehlstrategie war in 5 Arbeiten vornehmlich zu erkennen. Als Folge wurden in 4 Arbeiten sogar Lösungswahrscheinlichkeiten von über 100% angegeben und gegen grundsätzliche Eigenschaften der Wahrscheinlichkeit verstoßen (vgl. Abb.IV.13, vgl. auch 3.2.2.b).

Fallbeispiel: Schüler/in Nr.55 verwendet bei Teilaufgabe 1b im Zähler die Häufigkeit für Allergiefälle (10000) anstatt der Häufigkeit für Fälle mit Allergie und positivem Test (9900). Auch bei Teilaufgabe c. wird an diesem Fehlkonzept festgehalten (Zähler: Häufigkeit für keine Allergie (990000) anstatt für keine Allergie und Test positiv (19800)), obwohl sich dadurch eine Wahrscheinlichkeit größer als 1 ergibt.

(b) *Aufg. 1*
 $P(\text{Allergie} | \text{Test positiv}) = \frac{10000}{29700} \cdot 100\% \approx 33,67\%$

c) $P(\text{Keine Allergie} | \text{Test negativ}) = \frac{990000}{970300} \cdot 100\% \approx 102,03\%$

Abb.IV.13: Fallbeispiel für Strategie $|H| / |D|$ und Verstoß gegen das Normierungsaxiom der Wahrscheinlichkeit

Eine in 2 Arbeiten durchgehend auftretende Fehlstrategie war die Verwechslung von $P(H|D)$ mit $P(D|H)$. Der Begriff A-posteriori-Wahrscheinlichkeit (Aufgabe 1d) wurde in 4 Fällen erkennbar falsch interpretiert: 2 Schüler verstanden darunter die konjunktive Wahrscheinlichkeit $P(\text{Allergie und Test positiv})$, 2 Schüler die (unbedingte) Wahrscheinlichkeit $P(\text{Test positiv})$.

b) Darstellungsverwendung

Zur Verwendung der Darstellungsform Häufigkeitsbaum sind folgende Befunde zu erwähnen, die aber teilweise auf eine Bedeutungsungleichheit im Unterricht zugunsten des Baumdiagrammes zurückzuführen sind. Bei Teilaufgabe 2c war die Darstellungsform Tabelle vorgegeben und zu vervollständigen. Die Häufigkeitstabelle diente als Darstellungsmittel für die Beantwortung der Folgeaufgaben (2d und 2e). 36 Schüler (von 144 = 25%) stellten - obwohl nicht ausdrücklich verlangt - zusätzlich einen Häufigkeitsbaum zu dem Problem auf

¹¹³ wobei H für das a-priori bewertete Ereignis (Hypothese) und D für das Datum steht, nach dem das Ereignis H neu bewertet wird.

und verwendeten ihn für die weitere Bearbeitung der Folgeaufgaben mit, obwohl dadurch weitere Arbeitszeit verloren ging. Andererseits kam eine zusätzliche Verwendung von Tabellen bei vorgegebener Baumstruktur (bei Teilaufgaben 1a, e) nicht vor. Bei Teilaufgabe 2a gab es keine explizite Vorgabe der Darstellungsform. Eine Darstellung mit Häufigkeitstabelle erfolgte dort nur von 3 Schülern. Von allen anderen 141 Schülern (= 98%) wurden Häufigkeitsbäume zur Modellierung der Aufgabe verwendet. Weitere Befunde zur Darstellungspräferenz folgten aus den Befragungen (vgl. IV.3.2).

c) Interpretation als Neubewertung von Wahrscheinlichkeiten

Ein weiterer Befund bezieht sich auf die Teilaufgabe 1d, in der eine A-priori-Wahrscheinlichkeit und eine dazugehörige A-posteriori-Wahrscheinlichkeit bezeichnet und berechnet werden sollten. Gerade diese Aufgabe sollte prüfen, ob die Schüler die zentrale Denkweise der Neubewertung von Wahrscheinlichkeiten nachvollziehen und die dafür eingeführten Begriffe „a priori“ und „a posteriori“ richtig interpretieren können. 131 Schüler (= 91%) konnten eine A-priori-Wahrscheinlichkeit richtig angeben, 106 (= 73,6%) Schüler bezeichneten eine zugehörige A-posteriori-Wahrscheinlichkeit korrekt und 101 (= 70,1%) gaben sie korrekt an (bzw. berechneten sie korrekt). Diese Aufgabe ließ verschiedene Lösungen zu, von denen die naheliegendste war, die Basisrate $P(\text{Allergie})$ als a priori und die berechnete $P(\text{Allergie} \mid \text{Test}+)$ als A-posteriori-Wahrscheinlichkeit anzugeben. 27 Schüler gaben auch andere a-priori-a-posteriori-Zusammenhänge bei der Lösung an (vgl. Abb.IV.14).

Fallbeispiel: Schüler/in Nr.97 stellt sehr übersichtlich dar, dass sich die a priori getroffene Wahrscheinlichkeitsschätzung je nach Testergebnis (positiv oder negativ) zu einer A-posteriori-Schätzung ändert.

d) a priori	Testergebnis	a posteriori
Katzenhaareallergie + $P(K+) = \frac{10.000}{1.000.000} = 1\%$	Test +	33,3%
	Test -	$P(K+ \mid T-) = \frac{100}{970.300} = 0,01\%$

Abb.IV.14: Fallbeispiel für eine Darstellung der Neubewertungsinterpretation

Bemerkenswerterweise wechselten 8 Schüler auch bewusst die Schlussrichtung (vgl. Abb.IV.15 bzw. Kap.IV.1.3.2., Arbeitsblatt 5). Ein Schüler fasste die Fragestellung offener auf und modellierte eine neue Ausgangssituation, indem er eine A-priori-Schätzung $P(\text{Allergie})$ von 40% annahm und dazu die A-posteriori-Wahrscheinlichkeit $P(\text{Allergie} \mid \text{Test}+)$ bei den gegebenen Testdaten berechnete.

Fallbeispiel: Schüler/in Nr.111 gibt als a priori $P(\text{Test}+)$ bzw. als a posteriori $P(\text{Test}+ \mid \text{Allergie})$ an.

d)	$P(T+) = \frac{29700}{1000000} = 2,97\%$	a priori W.
	$P(T+ \mid A+) = \frac{9900}{10000} = 99\%$	a posteriori W.

Abb.IV.15: Fallbeispiel für eine Umkehrung der stochastischen Schlussrichtung

3.2 Kompetenztest zur Überprüfung der Nachhaltigkeit des Lernerfolges

3.2.1 Leistungsstabilität und Vergleich mit Labortrainingsgruppen

Die Nachhaltigkeit von Lernerfolgen wurde nach 14 Wochen mit demselben Kompetenztest zum Bayesianischen Denken gemessen, der auch in den Trainingsstudien im Labor verwendet wurde (vgl. Kap.III.3.2.2). Abb.IV.16 zeigt die mittleren Testleistungen (als Anteile an den erreichbaren Bewertungspunkten) und Abb.IV.17a die Leistungsverteilungen von 125 teilnehmenden Schülern¹¹⁴ 14 Wochen nach *Unterricht* im Vergleich zu Leistungen der Schüler, die Training im Labor erhalten hatten (vgl. III.3.3.2). Die Schüler erreichten 14 Wochen nach ca. 15-stündigem Unterricht („Unterricht 9“; vgl. Abb.IV.16) *insgesamt* durchschnittlich 68% der maximalen Testleistung. Im Vergleich dazu, hatten die Schüler 12 Wochen nach 2-stündigem Häufigkeitstraining im Labor („Freq 10“) im Durchschnitt 42% der maximalen Testleistung gezeigt, nach Wahrscheinlichkeitstraining („Prob 10“) nur noch 8%. Einen weiteren Vergleichsmaßstab bieten z.B. die Leistungen der Leistungskurs-Schüler, die auch im Unterricht das Thema behandelt und im Labor Wahrscheinlichkeitstraining („Prob LK“) erhalten hatten. 12 Wochen später erreichten sie im Durchschnitt nur noch 21% der maximalen Testleistung.

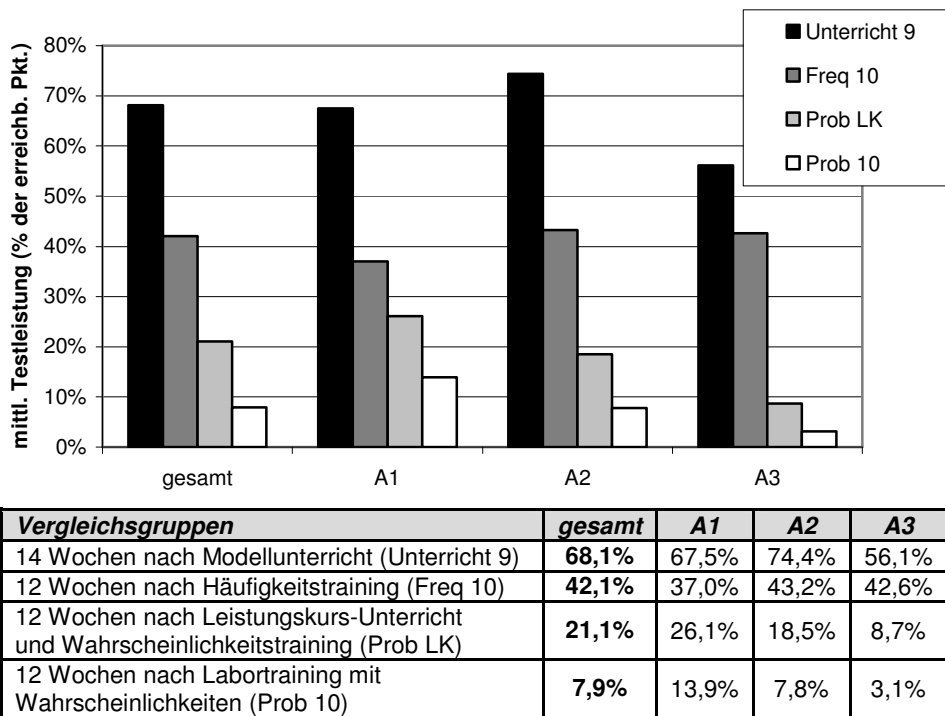


Abb.IV.16: Nachhaltigkeit des Lernerfolgs ca. 3 Monate nach dem Unterricht im Vergleich zu Trainingsgruppen der Laborstudie (Kap.III.3), gesamt und differenziert nach Itemtypen

¹¹⁴ 19 Schüler nahmen nicht teil.

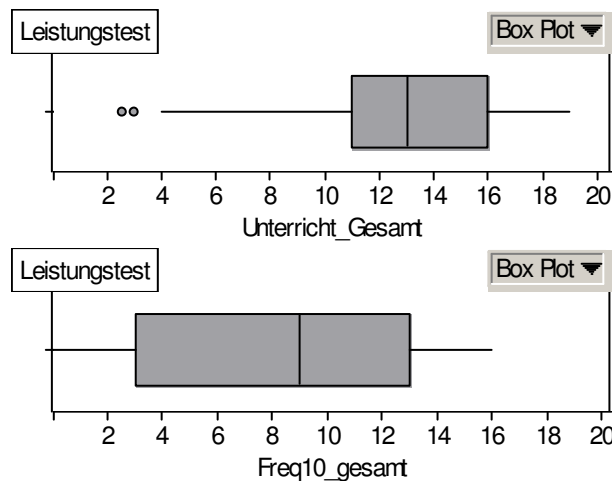


Abb.IV.17a: Nachhaltigkeit des Lernerfolgs mit dem Häufigkeitszugang ca. 3 Monate nach dem Unterricht (oben) bzw. nach Labortraining mit Zehntklässlern (unten, Verteilung der Bewertungspunkte, maximal 19)

Nach Itemtypen des Kompetenztests differenziert, wurden bei Items des *Typs A1* zur totalen Wahrscheinlichkeit 67,5%, des *Typs A2* zu üblichen dichotomen Bayes-Problemen 74% und des *Typs A3* zu trichotomen Bayes-Problemen 56% der maximalen Leistung erreicht (vgl. Abb.IV.16). Die Leistungen blieben nach Modellunterricht bei allen Itemtypen deutlich höher als in den verglichenen Trainingsgruppen, wobei bei A3-Items der Unterschied zu Häufigkeitstraining kleiner war.

Eine differenzierte Analyse nach Klassen sollte in dieser Arbeit nicht erfolgen.¹¹⁵ Leistungsunterschiede im Kompetenztest zwischen den fünf Klassen gehen aus Abb.IV.17b hervor. Die Unterschiede sind statistisch als nicht signifikant und im Rahmen der allgemeinen Leistungsstärke der Klassen in Mathematik als erwartungsgemäß zu bewerten.¹¹⁶ Eine Klasse (mittlerer Boxplot) zeigte ein deutlich polarisiertes Leistungsbild: 30% der Schüler erreichten die maximal erreichbare Punktsomme, aber auch 20% lagen unter der Hälfte der erreichbaren Punktsomme.

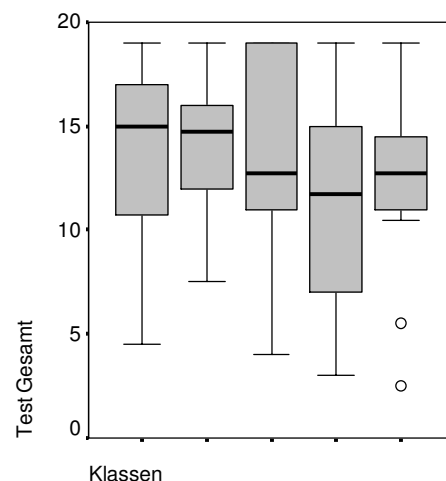


Abb.IV.17b: Nachhaltigkeit des Lernerfolgs nach ca. 3 Monaten in den mit dem Häufigkeitskonzept unterrichteten Klassen (Verteilung der Bewertungspunkte, maximal 19)

¹¹⁵ Das wurde mit den Lehrenden vereinbart.

¹¹⁶ Eine Varianzanalyse ergibt keinen signifikanten Unterschiedseffekt ($F=1,754$; $df=4$; $p=0,143$; $\eta^2=0,055$).

Aus Abb.IV.17c geht der Zusammenhang der Leistungen im Kompetenztest nach 14 Wochen mit den Leistungen in der Klassenarbeit (vgl. 3.1) genauer hervor.

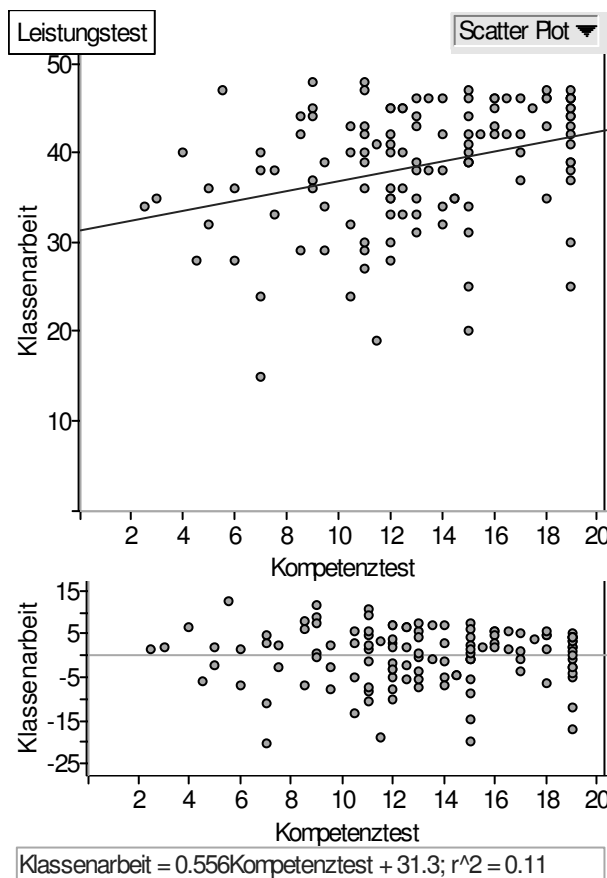


Abb.IV.17c: Zusammenhang der Leistungen im Kompetenztest nach 14 Wochen und den Leistungen in der Klassenarbeit (vgl. 3.1)

3.2.2 Qualitative Ergänzungen der Leistungsanalyse

Im Folgenden werden unterschiedliche individuelle Formen der Modellierung und insbesondere Lösungsstrategien, die zu unkorrekten Ergebnissen führten, an ausgewählten Beispielen genauer dargestellt und analysiert. Man beachte, dass insgesamt *hohe* Modellierungskompetenzen erreicht wurden und *überwiegend* korrekte Lösungsstrategien eingesetzt wurden – wie am Leistungsergebnis bereits abzulesen ist.

a) Darstellung und Modellierung

Zur Darstellung und Modellierung der Aufgabenstellungen des Kompetenztests wurden von 114 (von 125 = 91%) Schülern ausschließlich die im Unterricht entwickelten Häufigkeitsbäume verwendet („Doppelbaum“, vgl. Abb.IV.3). Modifikationen betrafen v.a. den Beschriftungsumfang (keine Beschriftung bis Beschriftung jedes Knotens) und Eintragung von Pfadwahrscheinlichkeiten in den Häufigkeitsbaum (vgl. Abb.IV.18). 3 Schüler modellierten ausschließlich mit Häufigkeitstabellen in der intendierten Form (vgl. Abb.IV.19). 5 Schüler verwendeten beide Darstellungsformen Baum und Tabelle im Test (jeweils eine pro Item) und 2 Schüler modellierten durchgehend verbal (ohne Darstellung) mit Häufigkeiten (vgl. Abb.IV.20).

Die Beispiele (Abb.IV.18-21) von Modellierungen bezogen sich auf folgendes Aufgabenitem:

ITEM 2a „PILOT“ (Typ A2, Antwortformat offen) :

Ein Pilot bemerkt während eines Fluges Rauch, der aus dem Motor aufsteigt. Er muss sich entscheiden, ob er den Flug vorzeitig abbricht oder ob er weiterfliegt, je nachdem wie wahrscheinlich es ist, dass der Motor aussetzt. Er verfügt über die folgenden Informationen aus der Flugstatistik:

- A** Wenn der Motor aussetzt, deutet in 90% der Fälle aufsteigender Rauch vorher darauf hin.
- B** Während 10% der Flüge, bei denen der Motor nicht aussetzt, kommt Rauch aus dem Motor.
- C** Der Motor seines Flugzeugtyps setzt in 1% aller Flüge aus.

Die für den Piloten (lebens-)entscheidende Frage ist nun:
Wie wahrscheinlich ist es, dass der Motor aussetzt, wenn Rauch aus ihm aufsteigt?

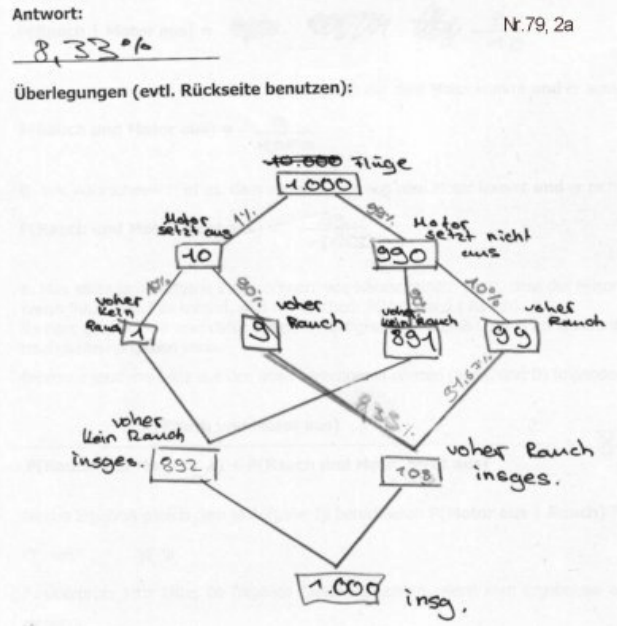


Abb.IV.18: Modellierung mit Häufigkeitsbaum und Pfadwahrscheinlichkeiten

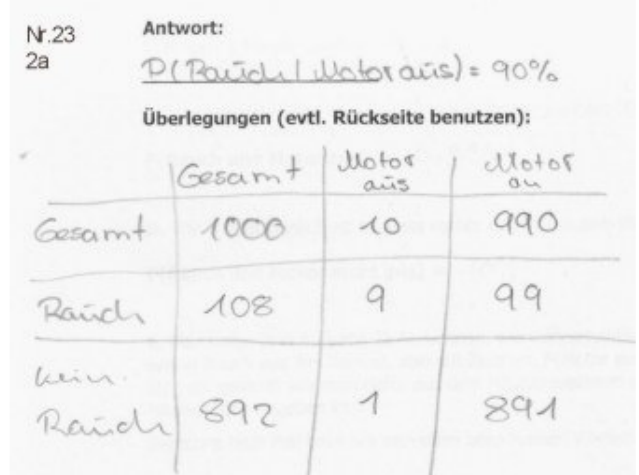


Abb.IV.19: Modellierung mit Häufigkeitstabelle (Lösungsantwort falsch)

Überlegungen (evtl. Rückseite benutzen): Nr.111, 2a

Zuerst 1: Der Motor setzt bei 10 von 1000 Flügen aus.

Zuerst 2: Bei 9 von 10 Flügen war Rauch zu sehen.

Zuerst 3: Obwohl er nicht aussetzt kommt bei 99 Flügen Rauch aus dem Motor.

Bei 108 Flügen von 1000 errichtet sich Rauch aber nur bei 10 setzt der Motor aus.

Abb.IV.20: Verbale Modellierung (ohne grafische Darstellung) mit Häufigkeiten

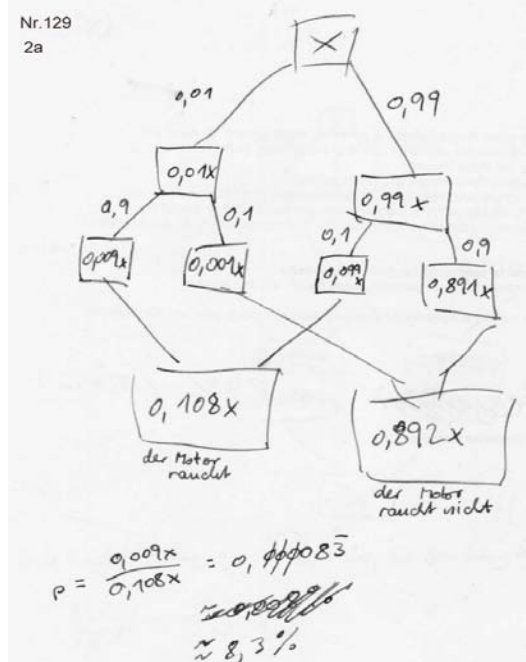


Abb.IV.21: Modellierung als Anteile an einer variablen Grundgesamtheit x

Ein Schüler verwendete eine interessante Form eines Baumdiagramms, in dem relative Anteile einer variablen Grundgesamtheit x angegeben wurden (vgl. Abb.IV.21). Die Modellierung ist so zu interpretieren: Z.B. die Häufigkeit der Motoraussetzer erhält man, wenn man die Gesamtzahl betrachteter Flüge x mit 0,01 multipliziert. Es werden in jedem Knoten relative Häufigkeiten h angegeben. Intuitive Pfadregelanwendung erfolgt z.B. bei $h(\text{Rauch}) = h(\text{Rauch und Motoraussetzer}) + h(\text{Rauch und kein Motoraussetzer})$. Der weitere Schritt zu einer allgemeinen Herleitung der Bayesregel mit Wahrscheinlichkeiten dürfte diesem Schüler nicht mehr schwer fallen.

Das Darstellen mit Häufigkeitsmodellen (Modellierungsschritte: „Mathematisieren und Strukturieren“, vgl. auch 3.1.3a¹¹⁷) erfolgte auch im Kompetenztest überdurchschnittlich gut. Bei A2-Items wurde in 88% aller Schülerlösungen eine korrekte und tragfähige Modellierung erstellt (bei 5% kein erkennbarer Lösungsversuch). Bei A3-Items war dies immerhin noch zu 73% der Fall (fast 15% kein erkennbarer Lösungsversuch). Bei nur 2 Schülern (ein „Baummodellierer“ und ein „Tabellenmodellierer“) zeigten sich tiefgreifende Probleme, genauer gesagt sie verfehlten immer das Ziel, ein korrektes Modell aufzustellen.

b) Lösungsstrategien

Es wurde bereits deutlich, dass v.a. das Aufstellen von Lösungstermen aus den erstellten Häufigkeitsmodellen bei den Schülern mehr zu Schwierigkeiten führte als die Modellerstellung selbst. Es stellt sich im Hinblick auf die kognitive Dimension der Arbeit die interessante Frage, welche Strategien maßgeblich *nach* dem Unterricht zu erkennen waren. In kognitionspsychologischen Studien (*ohne Training oder Unterricht*) zum Umgang mit Bayesproblemen (vgl. Kap.I.4.1.1.) wurde deutlich, dass mehrere „nicht-Bayesianische“ Strategien („Non Bayesian algorithms“; Gigerenzer & Hoffrage, 1995, S.694) gleichsam als „Primärintuitionen“ des Menschen gehäuft auftreten. In aller Kürze zusammengefasst, waren besonders häufige „nicht-Bayesianische“ Strategien bei dichotomen Bayesproblemen im Häufigkeitsformat, die konjunktive statt der bedingten Wahrscheinlichkeit zu berechnen („Joint occurrence“, formal: $P(H \cap D)$ statt $P(H | D)$ ¹¹⁸) bzw. die Bedingung zu vertauschen, d.h. die Likelihood-Wahrscheinlichkeit anzugeben („Fisherian“, formal: $P(D | H)$ statt $P(H | D)$). Auch nur die Angabe der Basisrate („Base-rate only“, formal: $P(H)$ statt $P(H | D)$) als Ergebnis war häufiger als intuitive Lösungsstrategie anzutreffen (vgl. genauer Kap.I, Tab.I.2). Studien von Zhu & Gigerenzer (eingereicht, vgl. I.4.1.4) zeigten, dass schon Kinder (4. bis 6. Klasse) bei in Häufigkeiten formulierten Bayesproblemen nach gewissen primären Intuitionen vorgehen. Neben der einfachen Angabe der Basisrate, wurde auch die Strategie „Joint occurrence“ oder die Bildung der totalen Wahrscheinlichkeit des Datums („Positives only“, formal: $P(D)$ statt $P(H | D)$) häufiger verwendet. Ein Ergebnis war allerdings, dass die Verwendung von Fehlstrategien im Laufe der Entwicklung der Kinder abnimmt und sich „Bayesianische“ Lösungen durchsetzen.

In Tab.IV.13 sind im Kompetenztest 14 Wochen nach Unterricht auftretende Lösungsstrategien bei Bayesproblemen zusammengefasst. Bei 41 Schülern (von 125) waren *ausschließlich* korrekte („Bayesianische“) Lösungsstrategien erkennbar. Die folgenden Häufigkeiten geben an, wie viele Schüler mindestens einmal eine jeweilige unkorrekte („nicht-Bayesianische“) Strategie verwendeten. Die Verwechslung mit der Konjunktion

¹¹⁷ Die Schritte „Mathematisieren i.e.S.“ und „Strukturieren“ entsprechen dem Darstellen mit Häufigkeitsmodellen (vgl. auch Laborbefunde in Kap.III.3.3.5).

¹¹⁸ wobei H für das a priori bewertete Ereignis (Hypothese) und D für das Datum steht, nach dem das Ereignis H neu bewertet wird.

$P(H \cap D)$ kam auch bei 11 Schülern, mit der Likelihood-Wahrscheinlichkeit $P(D | H)$ bei 19 Schülern vor. Die einfache Angabe der Basisrate $P(H)$ erfolgte nach Unterricht nur noch in einem Fall. Zumindest diese Fehl intuition war gewissermaßen verschwunden. Bei 7 Schülern war eine Strategie zu finden, die im Ergebnis zur totalen Wahrscheinlichkeit für das Datum führte.

Strategie	Lösungsterm	Ergebnis entspricht...	Häufigkeit ¹¹⁹
Bayes (normative Lösung)	$ H \cap D / D $	$P(H D)$	41
Konjunktion („joint occurrence“)	$ H \cap D / \text{Gesamtzahl}$	$P(H \text{ und } D)$	11
Likelihood („Fisherian“)	$ H \cap D / H $	$P(D H)$	19
Bayes ohne Likelihood	$ H / D $	keiner Wahrscheinlichkeit	30
Basisrate („base-rate only“)	$ H / \text{Gesamtzahl}$	$P(H)$	1
Totale Wahrscheinlichkeit („positives only“)	$ D / \text{Gesamtzahl}$	$P(D)$	7
1 minus Bayes	$ \text{nicht } H \cap D / D $	$1 - P(H D)$	2
?	$ D / H $	keiner Wahrscheinlichkeit	2
Bayes invers	$ D / H \cap D $	keiner Wahrscheinlichkeit	6
Likelihood invers	$ H / H \cap D $	keiner Wahrscheinlichkeit	4

Tab.IV.13: Strategien zur Lösung von Bayesproblemen ca. 3 Monate nach Unterricht

Eine fehlerhafte Lösungsstrategie, die in der Klassenarbeit und u.a. auch in Befunden von Zhu und Gigerenzer (eingereicht) am häufigsten bei Fünft- und Sechstklässlern auftrat, ist die Bildung des Quotienten aus den Häufigkeiten $|H| / |D|$, der nicht als Wahrscheinlichkeit interpretierbar ist. Im Kompetenztest etwa 3 Monate nach Unterricht trat diese Strategie („Bayes ohne Likelihood“) auch erstaunlich häufig auf (bei 30 Schülern und zweimal invers: $|D| / |H|$, vgl. Tab.IV.13). Die Strategie wurde bei Zhu & Gigerenzer als eine Art „Prototyp“ der korrekten „ $|H \cap D| / |D|$ “ angesehen. Demnach werden zwar die „möglichen Fälle“ (Nenner) korrekt erkannt – nämlich alle, auf die das Datum zutrifft. Doch für die Ermittlung der „günstigen Fälle“ (Zähler) bleibt die Likelihood-Wahrscheinlichkeit unbeachtet. Dies geschehe allerdings – so die Argumentation – unbewusst, also nicht als sogenannter bewusster „Shortcut“, der Rechenaufwand erspart. Die Strategie, so die Ergebnisse bei Zhu und Gigerenzer, werde im Laufe der Entwicklung ebenfalls zunehmend durch die „Bayesianische“ ersetzt.

Aus den eigenen Beobachtungen ergibt sich als weiterer Grund für unkorrekte Lösungsstrategien das ungenaue Bezeichnen von Häufigkeiten. Ich will dies an der „Pilotaufgabe“ (s. oben) demonstrieren: Z.B. wurden „konjunktive Häufigkeiten“ mit ungenauen Beschriftungen „Rauch“ bzw. „kein Rauch“ versehen (vgl. Abb.IV.22). Dadurch wurden sie eventuell nicht als solche („Rauch *und* Motor setzt aus“) interpretiert. Diese Bezeichnungsungenauigkeiten führten möglicherweise zur Vernachlässigung des wesentlichen Punktes bei der Bildung einer bedingten Wahrscheinlichkeit: Der Wahrscheinlichkeitsraum wird auf ein Ereignis reduziert (im Beispiel „D: Rauch kommt aus dem Motor“). Ereignisse unter der Bedingung D müssen immer Teilmengen in D sein. H ist aber i.A. nicht Teilmenge in D, aber natürlich $D \cap H$ (vgl. Buth, 2003; Abb.IV.23).

¹¹⁹ Die Häufigkeitsangabe bei „Bayes“ bedeutet: Anzahl der Schüler, bei denen nur (normativ-richtige) „Bayesianische“ Strategien erkennbar waren. Die Häufigkeitsangaben bei den „Nicht-Bayesianischen“ Strategien sind so zu interpretieren: Anzahl der Schüler, die mindestens einmal im Leistungstest diese Strategie benutzten. Man beachte, dass ein Schüler mehrere Strategien im Test angewandt haben könnte.

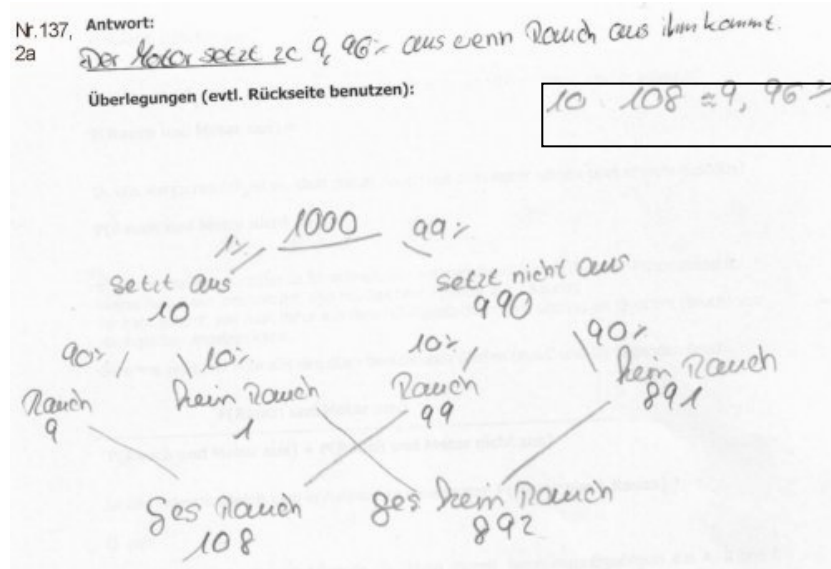


Abb.IV.22: Fallbeispiel für die Lösungsstrategie „Bayes ohne Likelihood“

Im Beispiel wird der Quotient „Häufigkeit von Motoraussetzern geteilt durch Häufigkeit von Rauch aus dem Motor = 10 / 108“ als Lösung angegeben, der keiner bedingten Wahrscheinlichkeit entspricht. Eine wesentliche Fehl intuition bei der Bildung bedingter Wahrscheinlichkeiten, die auch durch den Unterricht nicht völlig überwunden werden konnte, betrifft wohl die Reduktion des Wahrscheinlichkeitsraumes.

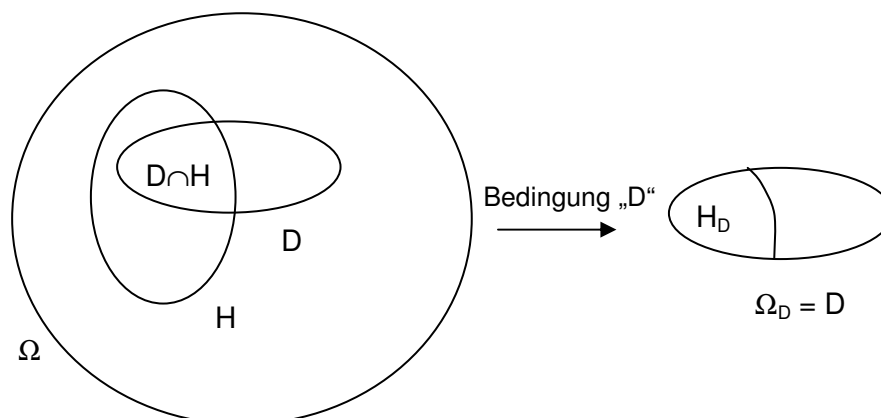


Abb.IV.23: Reduktion des Wahrscheinlichkeitsraumes zum bedingten Wahrscheinlichkeitsraum (vgl. Buth, 2003)

Im Falle obiger Aufgabe unterscheidet sich das Ergebnis zwar kaum vom korrekten „9 / 108“, was natürlich an der relativ hohen Likelihood-Wahrscheinlichkeit von 90% liegt. Jedoch kann die Strategie auch leicht zum völligen Scheitern führen: Angenommen die Basisrate wäre $P(\text{„Motoraussetzer“}) = 10\%$ und $P(\text{„Rauch“} \mid \text{„kein Motoraussetzer“}) = 1\%$, dann würde bei 1000 Flügen „Bayes ohne Likelihood“ zum Quotienten 100 / 99 führen. Ich werde weiter unten noch an Beispielen die tiefgreifenden Probleme von Schülern bei der Prozentrechnung belegen, die sogar einen Schüler in diesem Beispiel zu folgender Fehlinterpretation verleitete: „Die gesuchte Wahrscheinlichkeit ist $100 / 99 = 1,01\%$ “.

Sorgfältiges Beschriften der Darstellungen, so demonstriert auch Abb.IV.24, ist als integraler Bestandteil im Unterricht streng zu beachten und zu üben. Im ungünstigsten Fall

treten sonst bei Schülern Deutungsprobleme ihrer eigenen Modellbildung zu Tage, welche die Vorteile einer grafischen Darstellung ins Gegenteil verkehren würden.

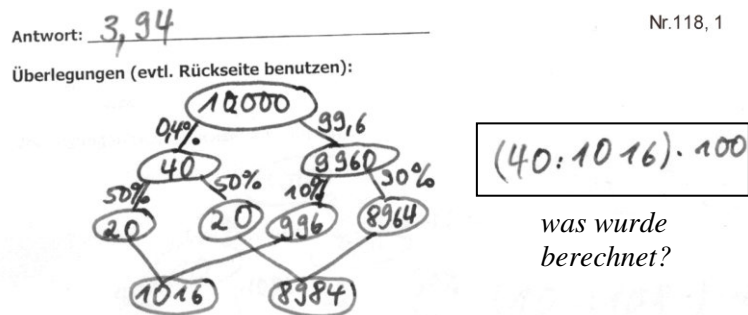


Abb.IV.24: Fallbeispiel für fehlende Beschriftung des Häufigkeitsbaumes (Lösungsstrategie „Bayes ohne Likelihood“)

Die genauere Betrachtung der Lösungsstrategien belegte tiefgreifende Defizite einiger Schüler im Prozentrechnen bzw. stochastischen Grundwissen. In 10 Fällen wurden die größeren Häufigkeiten der Gesamt mengen |H| bzw. |D| durch die Häufigkeit der Konjunktion $|H \cap D|$ geteilt, so dass sich ein Ergebnis größer als 1 ergab („Bayes bzw. Likelihood invers“, vgl. Tab.IV.13). Einige Schüler erkannten nicht, dass dieser Quotient (z.B. „108 / 9“, vgl. Abb.IV.25) keine Wahrscheinlichkeit sein kann (da > 1), weil sie das Ergebnis falsch umwandelten (Antwort: „= 12%“). Grundsätzlich war die Anwendung von Prozentrechnung eine Fehlerquelle, die häufig für das Scheitern bei der Modellierung verantwortlich war (vgl. Abb.IV.26).

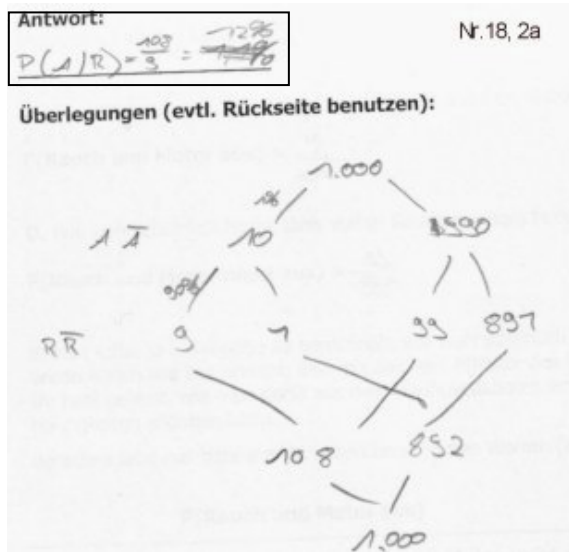


Abb.IV.25: Fallbeispiel für tiefgreifende Fehlvorstellung (Strategie „Bayes invers“)

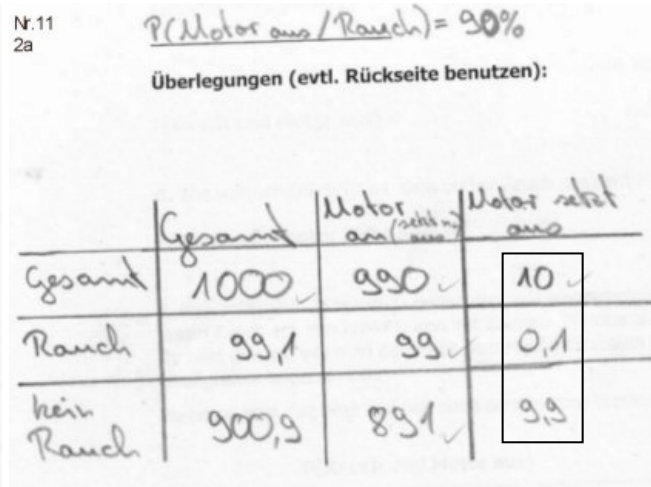


Abb.IV.26: Beispiel für Fehlvorstellung bei der Prozentrechnung („10% von 10 = 0,1“)

Diese qualitativen Befunde zu individuellen Lösungsstrategien im Kompetenztest zeigen, dass nach dem Unterricht weiterhin Fehlvorstellungen, wenn auch in erheblich geringerem Ausmaß, existierten. Die Häufigkeit des Auftretens gibt Hinweise darauf, welche intuitiven Fehlvorstellungen von Schülern besonders schwer abzubauen waren. Im Sinne einer Korrektur bzw. eines Abbaus theoretisch inkonsistenter primärer Intuitionen und eines Aufbaus und einer Verankerung theoretisch konsistenter sekundärer Intuitionen (vgl. Fischbein, 1975) gibt die Analyse konkrete Handlungsempfehlungen.

Aus dem gehäuften Auftreten der Strategien „Bayes ohne Likelihood“ und „Bayes invers“ kann gefolgert werden, dass die Wirkung einer Bedingung als Reduktion des Wahrscheinlichkeitsraumes im Unterricht zu bedingter Wahrscheinlichkeit sorgfältiger zu beachten ist. Sehr deutlich herausgearbeitet werden muss zum einen, dass eine Bedingung die Gesamtzahl der möglichen Fälle einschränkt. Zum anderen, dass nur noch Fälle für die Betrachtung in Frage kommen, die auch in dieser Einschränkung enthalten sind (vgl. Abb.IV.23, S.124). Zu empfehlen sind hier Training und Übungen mit „Häufigkeitsrastern“, wie bei Sedlmeier und Köhlers (2001; vgl. Kap. II.3.2.). Der Zusammenhang, dass die Konjunktionshäufigkeiten immer kleiner oder gleich der Häufigkeit der korrespondierenden Einzelereignisse sein müssen, wird mit dem Raster ebenfalls anschaulich. Demnach kann bei der Bildung von (bedingten) Wahrscheinlichkeiten, die einen Wert ≤ 1 annehmen müssen, nur ein Quotient „Konjunktionshäufigkeit durch Häufigkeit für Einzelereignis“ sinnvoll sein und nicht umgekehrt.

Bei einigen Schülern fehlten völlig grundlegende Kompetenzen über Wahrscheinlichkeitseigenschaften und Prozentrechnung, die natürlich für den Unterricht elementare Grundvoraussetzungen darstellen. Zu empfehlen wäre daher, vor Beginn der Unterrichtsreihe derartige Wissensstände zu prüfen und gegebenenfalls zu ergänzen.

3.3 Schülerbefragung

Von 140 Schülern der 9.Klassen wurde der Fragebogen bearbeitet (69 Schülerinnen, 71 Schüler), wobei in einigen Fällen Items nicht beantwortet wurden. Zur allgemeinen Einschätzung (nichtleistungsbezogener) Wirkungen des Unterrichts bzw. speziell des betonten Realitätsbezuges und der Häufigkeitsmodellierungsformen erfolgte eine Befragung mit Ratingskalen (1: volle Zustimmung, ..., 4: volle Ablehnung).

Das allgemeine Sachinteresse (U1, U2, U11, vgl. Tab.IV.14) speziell an den zurückliegenden Unterrichtsstunden war bei über 70% höher als sonst im Mathematikunterricht. Eine generelle Ablehnung des Faches Wahrscheinlichkeitsrechnung (U3) wird von nur 10% ausdrücklich bzw. von 30% tendenziell bekundet. Generelle Tätigkeitsanreize im Fach Mathematik (U10) sind bei etwa der Hälfte der Schüler positiv bewertet, beim zurückliegenden Stochastikunterricht (U9) bei ca. 72%. Als Folgenanreiz für Lernbemühungen im Unterricht sehen 90% der Schüler gute Noten (U5) als wichtig, aber nur 52% die Lebensrelevanz des Themas. 70% der Schüler bewerteten ihren Erfolg als Folge des Unterrichts auch ohne eigenes Handeln (U7). Dieses Item korreliert am deutlichsten mit der Leistung in der Klassenarbeit ($r = 0,43$). Die Schüler sahen zu 78% einen guten Lernerfolg als Folge der eigenen Lernbemühungen (U8). Selbststeuerungsprobleme wurden nur selten in diesem Zusammenhang angegeben (U12).

Unterricht <i>Item</i> (Kodierung: 1: stimmt genau, 2: stimmt eher, 3: stimmt eher nicht, 4: stimmt gar nicht)	Häufigkeiten					Mittel (Median)
	N	1	2	3	4	
U1. Verglichen mit anderen Stunden in Mathe fand ich den Unterricht interessanter	140	28	71	32	9	2,2 (2)
U2. Die Probleme (Aufgaben) fand ich spannender als sonst in Mathe	140	29	76	24	11	2,1 (2)
U3. Generell würde ich Wahrscheinlichkeitsrechnung abwählen, wenn ich könnte	140	15	27	61	37	2,9 (3)
U4. Ich habe im Unterricht gut aufgepasst, weil ich das im Leben gebrauchen kann	140	9	64	54	13	2,5 (2)
U5. Ich habe im Unterricht gut aufgepasst, weil ich eine gute Note wollte	140	71	55	13	1	1,6 (1)

U6. Ich habe im Unterricht gut aufgepasst, weil es mich interessiert hat	140	19	65	44	12	2,4 (2)
U7. Generell brauchte ich für den Unterricht nicht viel zu tun, weil ich alles leicht verstand	140	37	61	39	3	2,1 (2)
U8. Wenn ich mir Mühe gab, konnte ich im Unterricht alles verstehen	138	47	61	24	9	1,9 (2)
U9. Der Unterricht hat mir mehr Spaß gemacht verglichen mit anderen Stunden in Mathe	140	39	62	28	11	2,1 (2)
U10. Generell macht mir die Beschäftigung mit Mathe Spaß	139	23	50	44	22	2,5 (2)
U11. Ich hatte das Gefühl, dass der Unterricht abwechslungsreicher war als sonst in Mathe	140	25	57	43	15	2,3 (2)
U12. Ich hatte Probleme, im Unterricht meine Gedanken bei der Sache zu halten	139	8	22	74	35	3,0 (3)

Tab.IV.14: Allgemeine Unterrichtswirkungen (Häufigkeitsverteilung, Mittelwerte)

Speziellere Items betrafen die Wirkung von realitätsbezogenen Problemkontexten im Unterricht (vgl. Tab.IV.15). Knapp die Hälfte der Schüler konnte die Lebensrelevanz der Inhalte weitgehend erkennen (PK1). Bei ca. 64% war ein erhöhtes Sachinteresse aufgrund der realen Problemkontexte festzustellen (PK2). Für die meisten Schüler (88%) waren die Problemstellungen verständlich (PK3). Flowspezifische Anreize durch diese Problemsituationen wurden eher negiert (PK4). Fehlende mathematische Klarheit und Exaktheit bzw. Ablenkung durch viele nichtmathematische Informationen vom eigentlichen mathematischen Problem geben aus Schülersicht meist keinen Grund für Negativbewertung (PK5, PK6).

Problemkontexte <i>Item</i> (Kodierung: 1: stimmt genau, 2: stimmt eher, 3: stimmt eher nicht, 4: stimmt gar nicht)	Häufigkeiten					Mittel (Median)
	N	1	2	3	4	
PK1. Ich konnte gut erkennen, warum man das alles im echten Leben brauchen kann	140	12	50	59	19	2,6 (3)
PK2. Mich interessierte der Unterricht mehr als sonst, da es um reale Probleme ging	140	26	64	36	14	2,3 (2)
PK3. Ich hatte oft das Gefühl, dass ich nicht wusste, worum es eigentlich geht	138	4	13	62	59	3,3 (3)
PK4. Die Probleme im Unterricht haben mich richtig gefesselt verglichen mit sonstigen Matheproblemen	140	7	32	61	40	3,0 (3)
PK5. Ich habe die Klarheit und Exaktheit vermisst, die es sonst in Mathe gibt	140	10	25	66	39	3,0 (3)
PK6. Die vielen Informationen (über AIDS etc.) haben mich von der eigentlichen Mathematik abgelenkt	140	10	16	67	37	2,9 (3)

Tab.IV.15: Wirkungen des betonten Realitätsbezugs (Häufigkeitsverteilung, Mittelwerte)

Der dritte Teil des Fragebogens enthielt spezielle Items zur Wirkung der im Unterricht verwendeten häufigkeitsbasierten Modellierungs- und Darstellungsformen (vgl. Tab.IV.16). Wenn die Schüler sich selbst bewerten, sehen 96% das Ziel des verständigen Umgangs mit Häufigkeitsbäumen als erreicht an (D1) und ca. 90% das Modellieren von Problemstellungen in Häufigkeitsbäumen (D2). 89% beurteilen Häufigkeitsbäume als für sich hilfreich für die Problemlösung (D3), auch aufgrund der Förderung der Anschauung durch Häufigkeiten (D5). Generelle positive Anreize durch die Verwendung von Häufigkeitsbäumen sind sehr deutlich geäußert worden (D4).

Darstellungs- und Modellierungsform <i>Item</i> (Kodierung: 1: stimmt genau, 2: stimmt eher, 3: stimmt eher nicht, 4: stimmt gar nicht)	Häufigkeiten					Mittel (Median)
	N	1	2	3	4	
D1. Die Verwendung des Baumes habe ich bis jetzt nicht richtig verstanden	140	5	1	16	118	3,8 (4)
D2. Das Übersetzen der Textangaben in Werte im Baum habe ich leicht verstanden	140	66	59	11	4	1,7 (2)
D3. Ich konnte gut sehen, wie der Baum mir half, die Probleme zu lösen	139	67	57	9	6	1,7 (2)
D4. Es hat mehr Spaß gemacht, mit einem Baum zu arbeiten als mit irgendwelchen Formeln wie sonst in Mathe	140	73	39	18	10	1,8 (1)
D5. Im Baum wurde alles klarer, weil man da jeweils die Anzahlen von Personen erkennen konnte	140	57	66	11	6	1,8 (2)

Tab.IV.16: Wirkungen der häufigkeitsbasierten Modellierung (Häufigkeitsverteilung, Mittelwerte)

Bei der Frage einer Präferenz für eine der beiden im Unterricht verwendeten Darstellungsformen Häufigkeitsbaum bzw. -tabelle sprachen sich 69,1% für den Baum, 5,8% für die Tabelle aus, und 25,2% legten sich nicht fest (Tab.IV.17).

Darstellungsform, Präferenz <i>Item</i> (M.C.: 1 Baum / 2 Tabelle / 3 beide etwa gleich)	Häufigkeiten			
	N	1	2	3
PRÄ. Hast Du lieber den Baum oder die Tabelle verwendet?	139	96	8	35

Tab.IV.17: Verwendungspräferenz einer Darstellungsform

Am Ende des Fragebogens hatten die Schüler die Möglichkeit den Unterricht im offenen Antwortformat zu bewerten („Was hat Dir am Unterricht gefallen, was nicht?“). Für viele Schüler war es offensichtlich eine neue Erfahrung, sich über den zurückliegenden Unterricht (anonym) äußern zu dürfen und sie nahmen diese Gelegenheit rege in Anspruch. In der folgenden Übersicht (Tab.IV.18) wurde versucht, alle konstruktiven Äußerungen¹²⁰ der Schüler positiver und negativer Art zu erfassen und zu kategorisieren, wobei für inhaltlich analoge Kommentare nur jeweils ein Beispiel angegeben wurde (jeweils Angaben in Klammern dahinter: Codenummer, Geschlecht, Note in der Klassenarbeit, Note im letzten Zeugnis in Mathematik).

Kategorie/Bezug	Positive Äußerung	Negative Äußerung
1. Modellierungs- und Darstellungsformen, Grad der Formalisierung	Die Bäume haben geholfen, die Aufgaben besser zu verstehen und zu lösen (Nr.100, weiblich, 3 / 3) Gut: Texte in Baum und Tabelle umgewandelt (Nr.80, weiblich, 2 / 3) Der Baum hat viel Klarheit verschafft. Man konnte das Thema so schneller und besser verstehen (Nr.139, weiblich, 2 / 4) Mir hat gut gefallen, dass man nicht mit komplizierten Formeln arbeiten musste, wie es bei anderen Themen der Fall ist (Nr.117, männlich, 1 / 2). Ich fand das Thema gut, weil man nicht wie sonst mit Formeln rechnen musste (Nr.48, weiblich, 3 / 4) Mir war es wichtig, dass es sich immer nur um ein Thema handelte, nicht so viele Formeln und verschiedene Wege gab (Nr.97, männlich, 1 / 4).	Mir hat es nicht gefallen, dass wir uns 4 Wochen mit Bäumen und Tabellen beschäftigt haben. Es hatte nichts mit logischer Mathematik zu tun und man musste kaum rechnen. (Nr.51, männlich, 2 / 3)

¹²⁰ Die Kommentare sind aus den Schülerfragebögen wörtlich entnommen, lediglich orthografisch und grammatikalisch verbessert.

Kategorie/Bezug	Positive Äußerung	Negative Äußerung
2. Realitäts- und Lebensbezug	<p>Es ging nicht einfach nur um Mathe, sondern um Dinge, die das wirkliche Leben betreffen. Normalerweise frage ich mich in Mathe immer, wofür ich das jemals im späteren Leben brauchen werde, wenn ich nicht gerade Informatiker o.ä. werden will. [...] Doch diesmal war es etwas realistischer (Nr.1, weiblich, 4 / 3).</p> <p>Es waren realere Probleme, zu denen man mehr Bezug hatte und wo man besser erkennen konnte, wofür man so etwas lernt und auch später brauchen kann (Nr. 56, weiblich, 4 / 5).</p> <p>Es war aber nicht schlecht, dass man etwas über AIDS gelernt hat und dass alles an konkreten Beispielen war, denn sonst weiß man manchmal gar nicht, wozu man manche Dinge in Mathe braucht (Nr.15, weiblich, 1 / 2)</p> <p>Ich konnte in der ersten Stunde nicht genau erkennen, was das Thema mit Mathematik zu tun hat, da es am Anfang eher über die Krankheit AIDS ging und dieses ein wenig zu ausführlich für das Fach Mathe war. Aber gerade das machte das Thema auch interessant, weil man nicht starr irgendwelche Fragen beantworten musste, ohne zu wissen, worum es ging (Nr.31, weiblich, 2 / 2)</p> <p>Ich fand es gut, wie wir die Angaben über AIDS berechnet haben. Das war sehr erschreckend. [...] Man denkt immer 99% sind eigentlich ganz sicher, aber jetzt weiß man, dass es nicht so ist (Nr.70, weiblich, 2 / 2)</p>	<p>Manche Aufgaben waren allerdings unrealistisch (Nr. 42, weiblich, 5 / 4)</p> <p>Außerdem finde ich, dass das Thema Mord und Vergewaltigung nicht in einen Matheunterricht über Wahrscheinlichkeitsrechnung passt (Nr.31, weiblich, 2 / 2)</p> <p>Auch makaber und pervers war die Frage zum Testen, in denen gesagt wurde, dass nach positivem Aidsstest 22 Menschen Selbstmord begangen haben oder dass ein Mensch namens David einen Horror nach einem positiven Aidsstest durchlebt hat. Die Frage zu stellen: „Warum wurde der Gedanke eines Pflicht-AIDS-Tests in Deutschland verworfen?“ und auf eine mathematische Antwort zu hoffen, ist sehr pervers. (Nr.20, männlich, 2 / 2)</p>

Kategorie/Bezug	Positive Äußerung	Negative Äußerung
3. Komplexität und Anspruch von Problemtexten bzw. Fragestellungen	<p>Außerdem erschienen sie mir einfacher als andere Aufgaben, da wir Stochastik schon einmal in der 8.Klasse durchgenommen haben (Nr.31, weiblich, 2 / 2)</p> <p>Allerdings fand ich, dass man, wenn man sich angestrengt hat, schnell mit den Hausaufgaben fertig war (Nr.71, männlich, 2 / 2)</p> <p>Es [meint die Probleme] war leicht zu verstehen (Nr.132, männlich, 4 / 5)</p>	<p>Es hat mir auch nicht gefallen, dass manche Aufgaben so kompliziert formuliert waren und dass sie erst nach Erklärungen des Lehrers verstanden wurden. (Nr.15, weiblich 2 / 3).</p> <p>Die Texte waren teilweise zu lang, zu viele Infos, die nichts mit Mathe zu tun hatten. (Nr.116, männlich, 1 / 3)</p> <p>Mir hat es nicht gefallen, dass manche Formulierungen schwer zu verstehen waren (Nr.12, weiblich, 3 / 4)</p> <p>Allerdings kamen mir die Fragen oft kompliziert vor (Nr.31, weiblich, 2 / 2)</p>

Kategorie/Bezug	Positive Äußerung	Negative Äußerung
4. Interessantheit der Themen, Abwechslungsreichtum	<p>Der Unterricht, oder speziell die Aufgaben, waren interessanter, weil man u.a. auch auf die Lösungen und Ergebnisse gespannt war. Man wurde auch gleichzeitig über AIDS informiert (Nr.8, weibl., 2 / 3)</p> <p>Mir hat die Unterrichtsreihe eigentlich gut gefallen, weil sie abwechslungsreich war (Nr.31, weibl., 2 / 2).</p> <p>Der Unterricht war für mich nicht viel anders als sonst, da ich mich nicht direkt und persönlich damit befasse, sondern tue was von mir verlangt wird, um es dann geistig wegzulegen. Der Informationsschub über AIDS war allerdings doch recht interessant (Nr.26, weiblich, 2 / 2).</p> <p>Es hat mir gefallen, dass sehr genau auf die Fragestellungen eingegangen wurde und so vieles</p>	<p>Mir hat nicht gefallen, dass die Aufgaben oft das gleiche Thema hatten und man oft das Gleiche tun musste (Nr.117, männlich, 1 / 2)</p> <p>Nicht gefallen: immer dasselbe Schema nur mit anderen Texten und Zahlen (Nr.74, männlich, 1 / 1)</p> <p>Ich fand es nicht so gut, dass es eigentlich immer das gleiche Prinzip war (Nr. 139, weiblich, 2 / 4)</p> <p>Schlecht: Es wurde langweilig am Ende, weil wir es dann manchmal doch zu lang besprochen hatten (Nr.101, weiblich, 3 / 3)</p>

	<p>erklärt wurde. Die verschiedenen Beispiele der Texte brachten Abwechslung (Nr.65, weiblich, 3 / 5)</p> <p>Außerdem fand ich es gut, dass die Aufgaben in spannenden Geschichten und Beispielen verpackt waren (Nr.92, männlich, 2 / 3)</p> <p>Das Thema hat mir insgesamt gefallen, da es eine andere Art von Unterricht war (Nr. 42, weibl., 1 / 2)</p>	
--	---	--

Kategorie/Bezug	Positive Äußerung	Negative Äußerung
5. Arbeitsmaterial, Aufgaben	<p>Mir hat am Unterricht gefallen, dass er nicht vom Buch abhängig war (Nr.20, männlich, 3 / 3)</p> <p>Mir hat gefallen, dass nicht das Buch benutzt wurde. Es ging darin mehr um alltägliche Probleme, in die man sich hineinversetzen konnte. (Nr.90, weiblich, 3 / 4)</p> <p>Gut war auch, dass es viele Zettel zur Erklärung gab (Nr.97, männlich, 1 / 4)</p> <p>Mir hat es gefallen, dass wir vieles wiederholt haben und wirkliche Probleme angesprochen haben. Die Zusammenfassungen fand ich recht praktisch (Nr.12, weiblich, 3 / 4).</p> <p>Ich fand die Wiederholungszettel auch gut, denn man konnte so alles nochmal in Kürze nachlesen. (Nr.15, weiblich, 2 / 3)</p>	<p>Es war nicht gut, dass so viele einzelne Zettel dazu gehörten. Man konnte sie leicht verlieren und es war nicht übersichtlich. (Nr.11, weiblich, 2 / 3)</p> <p>Die Aufgaben waren oft nicht ganz eindeutig formuliert und haben einen so verwirrt. (Nr.56, weiblich, 3 / 3)</p> <p>Aufgabenstellungen waren aber nicht immer eindeutig (Nr.8, weiblich, 2 / 3)</p> <p>Mir haben die Aufgabenstellungen nicht gefallen, weil sie oft sehr undeutlich waren. Mir hat besonders die Arbeit nicht gefallen, weil ich nur eine 2 geschrieben habe (Sonst schrieb ich nur Einsen) (Nr.129, männlich, 2 / 1)</p>
6. Sozialformen, Unterrichtsklima	<p>Mir hat gefallen, dass wir mit unserem Partner zusammengearbeitet haben (öfter als im normalen Unterricht). (Nr.16, männlich, 5 / 4)</p> <p>Mir hat am Unterricht gut gefallen, dass er individueller gestaltet wurde als sonst (Nr.1, weiblich, 4 / 3)</p> <p>Der Unterricht lief lockerer ab (Nr.103, männlich, 1 / 3)</p>	<p>Negativ war, dass wir gefilmt wurden (Nr.110, weiblich, 1 / 1)</p>
7. Parallelisierung des Unterrichts in Klassenstufen	<p>Mir hat es gefallen, dass alle Klassen der Stufe 9 den Matheunterricht gleich gestaltet haben. (Nr.33, weiblich, 3 / 4)</p>	

Tab.IV.18: Freie Schüleräußerungen zur Frage: „Was hat Dir am Unterricht gefallen, was nicht?“ (In Klammern jeweils: Codenummer, Geschlecht, Note in der Klassenarbeit / Note im letzten Zeugnis in Mathematik)

Neben deutlich positiven Äußerungen der Schüler über die Formen der mathematischen Modellierung und insbesondere der zurückhaltenden Entwicklung des mathematischen Kalküls, gab es eine Äußerung, dass die gewohnte analytische Vorgehensweise vermisst wurde (vgl. Tab.IV.18, 1.).

Der Realitäts- und Lebensbezug wurde von den Schülern am häufigsten als positives Merkmal des Unterrichts bewertet. Hieraus geht deutlich hervor, dass es Schülern sehr wichtig war, den Anwendungsbezug der Mathematik im realen Leben erkennen zu können. Manchmal wurden sogar noch höhere Realitätsgrade gefordert, die natürlich nicht immer in einer für den Unterricht geeigneten Weise zu verwirklichen sind. Deutlich wurde, dass einige reale Themen (z.B. Mordfall) vielleicht für 9.Klassen zu heikel sind (vgl. 2.).

Eine größere Realitätsnähe bringt natürlich komplexere Problemsituationen und evtl. auch Fragestellungen mit sich, die oft von den Schülern als schwer bewertet wurden. Meines Erachtens ist das auch darauf zurückzuführen, dass derartiges Erarbeiten von komplexen außermathematischen Situationen eher nicht im Erfahrungsbereich der Schüler im

Mathematikunterricht liegt. Diese Vorgehensweise stellt sehr hohe Anforderungen an die Lehrenden, um allen Schülern diesen Weg zu ermöglichen (vgl. 3.).

Der Unterricht wurde aufgrund der Themen von vielen Schülern als interessanter und abwechslungsreicher als sonst empfunden, selbst von Schülern, die eine eher negative Einstellung zu mathematischen Inhalten haben (vgl. 4, Nr.26). Dennoch scheint bei einigen Schülern bei zu ausführlicher Beschäftigung mit einer Problemsituation bzw. mit einem Problemtyp die Motivation zu sinken. Die Frage des Umfangs der Problembehandlung ist wohl zu diskutieren.

Die Arbeit mit umfangreichen Materialien in Arbeitsblattform wurde meist als positiv bewertet, wenn auch manchmal eine gebundene Form vermisst wurde. Deutlich ist, dass Schüler Wiederholungsphasen und übersichtliche Rückblicke wünschen und brauchen. Bemängelt wurde die fehlende Eindeutigkeit mancher Fragestellungen. Neben der Kritik an einigen noch beinhalteten Unklarheiten im Erprobungsmaterial, war diese „Offenheit“ durchaus auch intendiert. So resultieren meines Erachtens auch Verwirrungen über relativ offen gestellte Aufgaben aus der Tatsache, dass die Schüler zu selten mit derartigen Problemstellungen konfrontiert sind. Z.B. der Schüler Nr.129 (vgl. 5.), der „sonst nur Einsen schrieb“, kommt offenbar normalerweise sehr gut mit eindeutigen Vorgehensweisen in der Mathematik zurecht, die aber im realitätsbezogenen Stochastikunterricht oftmals gar nicht so eindeutig vorgegeben sind.

Die individuelleren und freieren Sozialformen wurden allgemein begrüßt, Partnerarbeit v.a. auch von schlechteren Schülern (vgl. 6, z.B. Nr.16). Der Aspekt der Parallelisierung von Inhalten in einem Jahrgang wird auch von Schülern für positiv gehalten (vgl. 7). Diese Parallelisierung ist auch im Hinblick auf eine Angleichung von Leistungsbewertungen an sachlichere Bezugsnormen interessant (vgl. Rheinberg, 2001, S.70).

3.4 Lehrerbefragung

3.4.1 Standardisierte Befragung

Die Auswertung des Lehrerfragebogens sollte einen Überblick über Einschätzungen der fünf Lehrenden zu Lehrinhalten, -material und zum erfolgten Unterricht ergeben (vgl. folgende Tabellen IV.19, Teile 1 bis 10).¹²¹ Die verschiedenen Bildungswirkungen des erfolgten Unterrichts bei Schülern (Leistung, Interesse, etc.) wurden tendenziell höher eingeschätzt als im sonst üblichen Mathematikunterricht. Am deutlichsten gilt das für die Leistung in der Klassenarbeit, am wenigsten für Aufmerksamkeit bzw. Leistung bei Hausaufgaben.

1. Schülermerkmale	Wie beurteilen Sie Ihre Schüler in den zurückliegenden Stunden – im Vergleich zu sonstigem Mathematikunterricht – unter folgenden Kriterien?				
	deutlich höher	höher	etwa gleich	niedriger	deutlich niedriger
Interesse	-	3	2	-	-
Aufmerksamkeit	-	-	5	-	-
Verstehen	-	3	1	1	-
Motivation	-	2	3	-	-
Leistung (KA)	1	4	-	-	-
Leistung (Hausaufgaben)	-	2	1	2	-

Tab.IV.19.(Teile 1-10): Ergebnisse der Lehrerbefragung (Bewertungshäufigkeiten)

¹²¹ Im Folgenden werden aufgrund der kleinen Stichprobe (n=5) nur deskriptiv Bewertungshäufigkeiten angegeben.

Zwei der fünf Lehrenden hatten bereits Lehrerfahrung im Themengebiet „Bedingte Wahrscheinlichkeit und Satz von Bayes“ in der 13. Jahrgangsstufe, in der ca. 7 Unterrichtsstunden zur Verfügung standen. Sie konnten damals auf Vorwissen der Schüler zu „Elementarer Wahrscheinlichkeit“, „Kombinatorik“, „Pfadregeln“ und „Summenregel“ zurückgreifen.

2. Unterrichtserfahrung im Themengebiet	ja / nein	Klassenstufen	Stunden	Themenbereiche vorher
Bedingte Wahrscheinlichkeit / Satz von Bayes	2 / 3	13	ca. 7	„Elementare Wahrscheinlichkeit“, „Kombinatorik“, „Pfadregeln“, „Summenregel“

Zwei Lehrbücher wurden von den Lehrenden dabei verwendet:

3. Verwendete Schulbücher / Lehrmaterial
1. Griesel, H. & Postel, H. (2000, Hrsg.). Elemente der Mathematik – Grundkurs NRW. Hannover: Schroedel 2. Kuypers, W. & Lauter, J. (1984). Mathematik Sek.II: Wahrscheinlichkeitsrechnung und Statistik. Düsseldorf: Schwann

Die Bewertung der zugrundeliegenden Unterrichtsentwürfe und Arbeitsmaterialien (vgl. S.132) durch die Lehrenden war hauptsächlich positiv v.a. bei Anschaulichkeit, Übersichtlichkeit, Interessantheit und verwendete Darstellungsmittel. Übungsmöglichkeiten technischer Fähigkeiten und Umfang außermathematischer Vertiefung wurde auch negativer bewertet. Drei sehr negative Bewertungen eines Lehrenden bei Verständlichkeit, verwendete Anwendungsbeispiele und Übungsmöglichkeiten für Schlussfolgern und Interpretieren wurden zum Anlass genommen, das entworfene Material in dieser Hinsicht nochmal kritisch zu überarbeiten und zu verbessern.

4. Materialbewertung: Wie bewerten Sie die vorliegenden Unterrichtsentwürfe zum Thema „Bedingte Wahrscheinlichkeit und Satz von Bayes“ in der Sek. I unter folgenden Aspekten?					
	1*	2	3	4	5
Anschaulichkeit	1	3	1	-	-
Übersichtlichkeit	2	3	-	-	-
Verständlichkeit	1	2	1	-	1
Interessantheit	2	2	1	-	-
Gestaltungsspielraum (2**)	-	2	1	-	-
Verwendete Darstellungsmittel (Baum etc.)	4	1	-	-	-
Verwendete Anwendungsbeispiele	1	2	1	-	1
Umfang mathematisch-formaler Vertiefung	-	2	3	-	-
Umfang begrifflicher Vertiefung	-	2	3	-	-
Umfang der Vertiefung außermath. Kontexte	1	2	1	1	-
Übungsmöglichkeiten mathematisch-technischer Fertigkeiten	-	1	2	2	-
Übungsmöglichkeiten des Modellierens realitätsnaher Probleme (1**)	1	1	2	-	-
Übungsmöglichkeiten des Schlussfolgerns und Interpretierens	1	1	2	-	1
* Kodierung: 1: sehr gut, 2: gut, 3: befriedigend, 4: ausreichend, 5: mangelhaft **Anzahl der fehlenden Angaben (missings)					

Die Lehrenden hielten das Darstellungsmittel „Häufigkeitsbaum“ als Modellierungshilfe für wichtig. Sie sahen den Häufigkeitsbaum als weitgehend leicht von den Schülern zu begreifen und anpassbar an neue Problemstellungen.

5. Bewertung zur Rolle des Darstellungsmittels „doppeltes Baumdiagramm“	1*	2	3	4
Ohne die Darstellung hätten die Schüler die Probleme nicht mathematisch durchdringen können.	3	-	2	-
Schüler können auch ohne ein Darstellungsmittel richtige Lösungen finden.	1	1	1	2
Die Darstellung war als Modellierungshilfe für die Schüler unbedingt nötig.	3	2	-	-
Die Schüler konnten die Systematik und Gesetzmäßigkeit der Darstellung leicht begreifen.	1	4	-	-
Die Schüler konnten die Darstellung leicht an neue Problemstellungen anpassen.	2	3	-	-
* Kodierung: 1: stimmt genau, 2: stimmt weitgehend, 3: stimmt eher nicht, 4: stimmt gar nicht				

Nach Schätzung der Lehrenden verwendeten mindestens 70% der Schüler im Unterricht oder bei Hausaufgaben das Baumdiagramm zur Lösung der Probleme. Das Baumdiagramm wurde als leichter für Schüler begreiflich und besser anpassbar an neue Problemstellungen als die Tabelle bewertet.

6. Vergleich der Darstellungsmittel „Doppelbaumdiagramm“ und „Tabelle“	Baum	etwa gleich	Tabelle
Welche Darstellung begreifen die Schüler bzgl. ihrer Systematik besser?	4	1	-
Welche Darstellung kann leichter von den Schülern an neue Problemstellungen angepasst werden?	4	1	-
Welche Darstellungen haben Sie bisher im Unterricht verwendet und haben Sie jeweils Vor- bzw. Nachteile festgestellt?	einfache Baumdiagramme mit Wahrscheinlichkeiten		

Die betonte Verwendung von Häufigkeitsformaten in dieser Unterrichtsreihe bewerteten die Lehrenden als Erleichterung für die Schüler beim Modellieren und Rechnen. Die Übersetzung der Informationen von Prozent- oder Dezimalzahlen in entsprechende Häufigkeiten wurde als weitgehend problemlos angesehen, wobei die Rückübersetzung nicht von allen Lehrenden als problemlos für die Schüler bewertet wurde.

7. Bewertung zur Rolle des Häufigkeitsformates	1*	2	3	4
Ohne die Überlegung in Häufigkeiten hätten die Schüler die Probleme nicht mathematisch durchdringen können.	1	3	-	1
Schüler können auch ohne Übersetzung von Prozentangaben oder Dezimalzahlen in Häufigkeiten richtige Lösungen finden.	-	3	2	-
Häufigkeiten erleichterten den Schülern das Modellieren der Problemstellungen.	2	3	-	-
Häufigkeiten erleichterten den Schülern das Rechnen.	2	3	-	-
Die Schüler konnten Angaben in Prozent- oder Dezimalzahlen leicht in entsprechende Häufigkeiten übersetzen.	1	4	-	-
Die Schüler konnten aus gefundenen Häufigkeiten leicht Wahrscheinlichkeiten berechnen.	-	4	-	1
* Kodierung: 1: stimmt genau, 2: stimmt weitgehend, 3: stimmt eher nicht, 4: stimmt gar nicht				

Das Häufigkeitsformat wurde von den Lehrenden tendenziell als vorteilhafter als „Prozent- oder Dezimalformate“ beim Finden, Durchdringen und Erinnern eines Lösungsweges beurteilt.

8. Vergleich des numerischen Formates „Häufigkeit“ mit „Prozent-/Dezimalzahl“	Häufigkeit	etwa gleich	% / Dez.
Welches numerische Format ermöglicht leichter das Finden des richtigen Lösungsweges?	3	2	-
Welches numerische Format ermöglicht leichter das Durchdringen der Lösung?	4	1	-
Welches numerische Format unterstützt besser das Erinnern des Lösungsweges?	4	1	-

Die Bedeutung realistischer Kontexte auf motivationale Merkmale wurde von den Lehrenden weitgehend bestätigt. Eine Erschwerung des mathematischen Modellierens durch realistische Kontexte wurde eher nicht gesehen. Ansonsten wurden die Aussagen kontroverser beurteilt. Meiner Meinung nach positiv ist, dass 4 Lehrende die Verwendung realer Kontexte im Unterricht weitgehend nicht für zu aufwendig halten und die Schüler hinsichtlich Sprache oder Modellbildung nicht überfordert sehen. Drei Lehrende präferieren überwiegend realistische Kontexte im Stochastikunterricht, zwei sehen ein ausgewogenes Verhältnis realistischer und konstruierter Problemstellungen als optimal an.

9. Bewertung zur Rolle der realistischen Kontexte	1*	2	3	4
Ohne die Einkleidung in realistische Kontexte wären die Schüler weniger motiviert gewesen. (1**)	2	2	-	-
Schüler können auch ohne Einkleidung in realistische Kontexte die Sinnhaftigkeit von Mathematik erkennen. (1**)	1	1	2	-
Die Einkleidung in realistische Kontexte erschwerte den Schülern das mathematische Modellieren. (1**)	-	-	4	-
Die Einkleidung in realistische Kontexte lenkte die Schüler zu sehr von der formalen Mathematik ab. (1**)	1	-	3	-
Die Einkleidung in realistische Kontexte ist zwar wünschenswert, war aber im Unterricht zu aufwendig.	1	-	3	1
Die Einkleidung in realistische Kontexte förderte das logische Denken und kritische Reflektieren der Schüler. (1**)	2	1	-	1
Die Einkleidung in realistische Kontexte überforderte die Schüler hinsichtlich sprachlicher Kompetenzen bzw. Genauigkeit.	-	1	4	-
Die Schüler konnten Probleme in realistischen Kontexten schlechter mathematisch modellieren als in konstruierten Kontexten. (1**)	-	2	2	-
* Kodierung: 1: stimmt genau, 2: stimmt weitgehend, 3: stimmt eher nicht, 4: stimmt gar nicht **Anzahl der fehlenden Angaben (missings)				
	mehr realistisch	ausgewogen	mehr konstruiert	
Eigene Präferenz	3	2	-	

Die in der Reihe mit Zusatzmaterial ausgearbeiteten realitätsbezogenen Problemstellungen (vgl. IV.1.2.6) wurden nicht alle für den Unterricht als gleichermaßen geeignet beurteilt. Die medizinisch-diagnostischen Situationen wie AIDS-Test, Schwangerschaftstest und Vaterschaftstest wurden vergleichsweise als geeigneter bewertet als andere. Nach einer Einschätzung hatten aber v.a. die reichlichen Beispiele zu den recht guten Leistungen geführt. Im zur Verfügung stehenden Unterrichtszeitraum wurden hauptsächlich die Kontexte Mammografie, BSE-Krise und Drogen im Straßenverkehr nicht bearbeitet.

10. Bewertung der Eignung der realistischen Beispiele	gut geeignet	bedingt geeignet	nicht geeignet
AIDS-Test	4	1	-
AIDS-Test aus „Sicht der Testentwickler“	3	2	-
Mordfall	-	5	-
Schwangerschaftstest	3	2	-
Vaterschaftstest	2	2	1
Mammografie (und Ultraschall) (2**)	1	-	2
BSE-Krise (2**)	1	-	2
Drogen im Straßenverkehr (2**)	2	-	1

Die personenbezogenen Angaben der Lehrenden zur fachlichen Ausbildung in Stochastik zeigten kontroverse Einschätzungen (sehr zufrieden bis sehr unzufrieden). Eine persönliche Einschätzung, welche Faktoren die heutige Kompetenz der Lehrenden in Stochastik bestimmen, ist ebenfalls sehr uneinheitlich, wobei als wesentlichere Faktoren die Ausbildung, das Lernen aus eigener Erfahrung während der Berufstätigkeit und die Lektüre fachlicher

bzw. fachwissenschaftlicher Literatur angegeben werden. Fortbildungsveranstaltung, der Austausch mit Kollegen und Kolleginnen und didaktische Literatur scheinen tendenziell eine unmaßgeblichere Rolle zu spielen.

3.4.2 Nichtstandardisierte Befragung

Einige interessante Ergänzungen zum standardisierten Fragebogen ergaben sich in der abschließenden offenen Gruppendiskussion (vgl. genauer 2.3.4), die im Folgenden stichpunktartig zusammengefasst sind:

- Offene Fragestellungen, wie etwa: „Was kann passieren, wenn ein HIV-Infizierter getestet wird?“ (vgl. Arbeitsblatt 1), können problematisch sein, da sie zu weit von mathematischen Zielen wegführen können.
- Offene Fragen, wie: „Wann brachten die Neubewertungen keine ausreichende Sicherheit und was wurde dann getan?“ (vgl. Arbeitsblatt 8, Nr.1c), können aber auch Anstoß zum Nachdenken sein, z.B.: „Was wäre eine ausreichende Sicherheit? Mache eigene Annahmen!“.
- Das Themengebiet heißt zwar „Anwenden des Satzes von Bayes“, doch er wird nicht formal entwickelt. Die formale Entwicklung wäre auch völlig unnötig und unangebracht in der 9.Klasse. Die Schüler wären nur verwirrt. Selbst im Leistungskurs in der Oberstufe wären mindestens 5-6 Stunden dafür nötig.
- Die Reihe ist sehr geeignet für Schüler mit geringeren algebraischen Fähigkeiten. Es ist generell wenig Vorwissen nötig. Für viele Schüler bedeutete das eine „Rettung“ des bisherigen schlechten Leistungsstandes.
- Die Baumdiagramme mit Häufigkeiten machen vieles transparenter.
- Gruppenarbeit oder Partnerarbeit ist an vielen Stellen sinnvoll (z.B. bei Arbeitsblatt 4 zum Einfluss der Basisrate; Arbeitsblatt 7, Nr.4)
- Die Überblicke folgen zeitlich zu kurz hintereinander.
- Der Zusammenhang bzw. die Abgrenzung der Begriffe in Überblick 1 (Merkmal und Bedingung) und in Überblick 2 (Hypothese und Indiz) muss sehr deutlich entwickelt werden. Diese Begriffe werden sonst von den Schülern womöglich synonym verwendet und vertauscht.
- Die A-priori-Schätzung 1 von 10000 möglichen Tätern im Mordfallbeispiel (Arbeitsblatt 7) ist eine problematische Hürde. Es wird nicht leicht erkannt, dass es sich nur um eine mehr oder weniger plausible Annahme handelt. Diese Denkweise ist sehr vorsichtig und genau mit den Schülern zu entwickeln. Insbesondere muss hier beachtet werden, dass die Aufgabenstellungen aufeinander aufbauen.
- Es wäre wünschenswert, dass Aufgaben und Arbeitsaufträge nach Schwierigkeitsgraden gekennzeichnet wären (z.B. „Sternchenaufgaben“ etc.).
- Einige Übungsaufgaben sollten eher in gemeinsamer Diskussion bearbeitet werden (z.B. Arbeitsblatt 8, Nr.1).
- Es sind nicht alle Anwendungsbeispiele in der verfügbaren Unterrichtszeit behandelbar.
- Es ist im Hinblick auf Leistungsfeststellungen nötig auch verstärkt (unrealistischere) Übungsaufgaben zu behandeln.

3.5 Unterrichtsbeobachtung

Im Folgenden werden tatsächliche Unterrichtsprozesse in *einer* Klasse mittels Transskriptausschnitten der Videoprotokolle genauer beschrieben und illustriert. Ich möchte mich jedoch in dieser Arbeit auf ausgewählte Teile beschränken, die ich für Schlüsselstellen des Unterrichtsgangs halte oder die meines Erachtens Besonderheiten gut verdeutlichen. Darüber hinaus werden solche Unterrichtsszenen analysiert, die besondere Schwierigkeiten des Lehr-Lern-Prozesses oder Abweichungen von intendierten Inhalten bzw. Prozessen kennzeichnen. Die entsprechenden transkribierten Unterrichtsszenen werden jeweils inhaltlich anhand des verwendeten Materials (vgl. Anhang B) eingeordnet. Die Transskripte greifen hierbei nur *offizielle* Äußerungen auf und werden durch eigene erklärende Kommentare in Klammern [...] sowie ggf. durch Videobilder ergänzt. Die Zielsetzung dieser Analyse ist *nicht* die detaillierte Unterrichtsanalyse mit qualitativ-interpretativen oder semiotischen Ansätzen, wie in Kap.IV.2.3.5 bereits erläutert worden ist.

3.5.1 Einstieg und erste Modellierung mit Häufigkeitsbäumen


Es erfolgt zunächst eine kurze Beschreibung des Verlaufs der *ersten Unterrichtsstunde*, die durch das Arbeitsblatt 1 (vgl. Anhang B) bereits sehr genau vorstrukturiert wurde. Es wurde sehr eng nach dieser Struktur vorgegangen. Nach dem lauten Vorlesen des Einstiegstextes durch einen Schüler, wurde ohne mathematischen Hintergrund über den Inhalt kurz diskutiert. Dabei haben einige Schüler bereits die numerischen Informationen herausgestellt, andere eher die kontextuellen Informationen betont.

Der Lehrer lenkte die Diskussion schließlich auf eine spezielle Situation, die in Aufgabe 1 genau formuliert ist. Der Zusammenhang, den Begriff der Sicherheit mit der Angabe einer Wahrscheinlichkeit zu spezifizieren, schien den Schülern plausibel zu sein. Jeder Schüler konnte so intuitiv eine Schätzung der Sicherheit abgeben. Je sicherer der Test eingeschätzt wurde, desto höher fiel die angegebene Wahrscheinlichkeit aus. Nicht verwunderlich war, dass alle Schüler über 90% als Wahrscheinlichkeit angaben (29-mal 99,9%, 1-mal 99%).

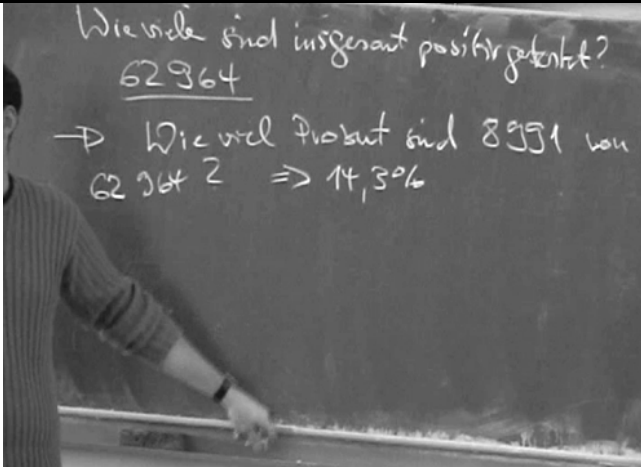
Die Aufgabe 2 war offen formuliert: „Was kann passieren, wenn...“. Es war zunächst nicht offensichtlich, dass die vier möglichen Ergebnisse dieser Zufallssituation Ziel der Frage waren. Eine Hinführung des Lehrers war aber relativ problemlos. Die Aufgabe 3 wurde zunächst selbstständig in etwa 10-minütiger Partner- bzw. Gruppenarbeit bearbeitet. Die Ergebnisse führten zur Aufstellung eines Häufigkeitsbaumes (Aufgabe 4) an der Tafel in einem vom Lehrer gelenkten Gespräch (vgl. Abb.IV.27).

Die Mathematisierung der Situation mit dem Häufigkeitsbaum schien für die meisten Schüler subjektiv offensichtlich. Schwierigkeiten waren vornehmlich technischer Art bei Prozentrechnung (z.B. $0,05\% \cdot 18000000 = 9000$), so dass eine kurze Wiederholungsphase eingebaut wurde. Differenzbildung bei zu ergänzenden Baumwerten schien auch intuitiv klar, z.B. „Nicht-infiziert“ sind $18000000 - 9000 = 17991000$. Die Gültigkeit einer bereits bekannten Verzweigungsregel in Baumdiagrammen wurde in der Häufigkeitsanschauung auch intuitiv begründet, z.B. $P(\text{nicht HIV und Test negativ}) = 99,7\%$ bzw. der entsprechende Häufigkeitsquotient $17937027 / 17991000 = 99,7\%$; die Summe aus $P(\text{nicht HIV und Test positiv}) = 0,3\%$ (war gegeben) und $P(\text{nicht HIV und Test negativ}) = 99,7\%$ ist 1, wie in der Verzweigungsregel gefordert.

Der nun folgende entscheidende Lösungsschritt bei dieser ersten Modellierung (Aufgabe 5) soll anhand des entsprechenden Transskriptausschnitts (vgl. Tab.IV.20) genauer beschrieben werden.¹²²

Unterrichtsstunde 1 Materialgrundlage: Arbeitsblatt 1 Thema: Wie sicher ist der AIDS-Test	
Zeit	Unterrichtsdialog <i>Abkürzungen: L: Lehrer, S: Schüler/in, LE: Lehrerecho, AB: Arbeitsblatt</i>
	<p>[Der Häufigkeitsbaum wurde gemeinsam an der Tafel erstellt (Aufgabe 4, vgl. Abb.IV.27). Das folgende Teilziel ist die Lösung der Aufgabe 5 mit dem aufgestellten Baumdiagramm]</p>  <p style="text-align: right;">Abb.IV.27</p>
37:08	L: Wofür wir dieses Baumdiagramm noch benutzen wollen diese Stunde, ist zu kucken, ob eure Häufigkeiten [deutet auf den häufigsten Schätzwert der Schüler aus Aufgabe 1: 99,9%] richtig sind. – So, was war die Frage? Kucken wir noch mal nach: [liest vor aus AB] Wahrscheinlichkeit für eine HIV-Infektion, wenn der Test positiv ist. [schreibt es an die Tafel als Aufgabe 5.]. Wie berechnen wir diese Wahrscheinlichkeit?
38:00	S: Erst mal kucken, wie viele der Getesteten überhaupt insgesamt positiv sind. Wenn man diesen Wert hat, kann man kucken, was 1% entspricht und dann natürlich alle Tests, die positiv sind, wie viel % das dann sind. L: Hm.
38:30	L: [zeigt auf Tafel] Wie viele sind insgesamt positiv getestet? Und das sind?
38:55	S: 8991 L: Und das sind alle, die positiv getestet wurden? S: plus 53973 [LE: L führt an der Tafel beide Knoten zu einem zusammen mit Bezeichnung „Test positiv“]
39:13	L: Und wie viele sind das? S: 62964 [LE: L schreibt es in den Knoten und als Antwort auf die Frage an die Tafel] L: So, was ist jetzt die nächste Frage?
39:36	S: Na, man muss kucken, wie viel 1% davon ist [meint von 62964] und dann hab ich [...unverständlich...] multipliziert und gekuckt, wie viel das dann sind.
39:48	L: Welche beiden Zahlen müssen wir jetzt ins Verhältnis setzen? Wir wollen wissen die Wahrscheinlichkeit für HIV-Infektion, wenn der Test positiv ist.
40:13	S: Wie viel jetzt wirklich positiv waren [meint von 62964]. L: Und wie viel sind das? S: Die 9000. L: Nein, nicht alle 9000?
40:25	S: Na die 8991. L: Warum die? S [fährt fort]: Weil hier die falschen Testergebnisse abgezogen sind.
40:41	L: Also, ich möchte ja wissen, wie viele meiner positiv Getesteten sind wirklich HIV infiziert? Warum muss ich diese 9 [zeigt drauf] dann rausschmeißen?
40:58	S: Na, weil die falsch sind. L: Was heißt falsch? S [fährt fort]: Negativ getestet sind die.
41:02	L: Ja, stimmt, die haben ja gar kein positives Testergebnis. Ich möchte ja nur wissen: Wie viele von meinen positiv Getesteten. Also wie viele von diesen [zeigt auf die Fälle HIV+ und Test+] und von diesen [zeigt auf die Fälle HIV- und Test+] sind wirklich HIV positiv.

¹²² Aus Gründen der Anonymität wird bei allen in der Arbeit abgedruckten Transskripten auf Namen oder sonstige personenbezogenen Angaben verzichtet. Das Transskript gibt den offiziellen Unterrichtsdiallog [erklärende Bemerkungen in Klammern] wieder. Die zeitliche Kodierung ist bezogen auf den Anfang der jeweiligen Unterrichtsstunde.

41:28	L: Mit anderen Worten muss ich jetzt rausfinden: Wie viel Prozent sind 8991 von 62964? [schreibt es an die Tafel] - Und wie viel Prozent sind das?
42:01	S: 14,3 L [LE]: 14,3% [vgl. Abb.IV.28]
	 <p style="text-align: right;">Abb.IV.28</p>
42:09	S: nein?! [Unruhe und Verblüffung bei vielen Schülern]
42:13	L: Was besagt jetzt diese 14,3%?
42:27	S: Dass von allen positiv Getesteten 14,3% HIV positiv sind.
42:32	L: Was besagt das über eure geschätzten 99,9%?
42:39	S: [Gelächter, Unruhe] ja, ist vollkommen falsch! Total daneben. L: M. [ein S. der niedriger schätzte] lag also am Anfang besser.
42:50	L: Also nur 14,3%, die am Ende ein positives Testergebnis kriegen, sind auch wirklich HIV-infiziert.
43:12	S: Also ich verstehe jetzt: Man muss die beiden positiven Tests addieren, dann kommt man auf die 62964. Und wie kommt man dann auf den nächsten Schritt? L: Auf welchen, das hier? [zeigt auf die Frage nach dem Anteil der HIV+ und Test+ an allen Test+] S: Ja, wie viel Prozent das da genau sind.
43:25	S [anderer]: Man weiß ja, dass die 53973 gar nicht in Wirklichkeit HIV-positiv sind. Und das muss man halt dann wegrechnen. [wird durch Stundengong unterbrochen]
43:35	L: Wir vertiefen das nächste Stunde weiter. Eure Hausaufgabe: Baum nochmal zeichnen aus dem Gedächtnis, Beantwortung der Frage von O. [Schülerfrage, siehe 43:12]: „Wie kommt man auf die 14,3%“, Rechenweg nochmal überlegen!

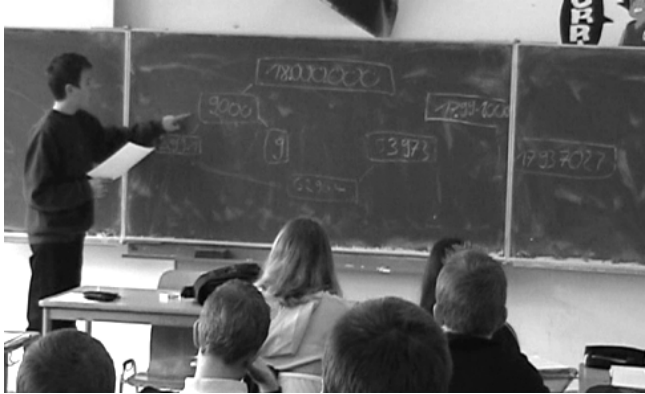
Tab.IV.20: Unterrichtsszene zur ersten Modellierung mit Häufigkeitsbäumen

Aus dem gelenkten Unterrichtsgespräch geht hervor, dass die Summenbildung aller positiv Getesteten vom Schüler intuitiv erfolgte (Timecode 38:55). Dieser entscheidende Schritt zur Lösung wäre in einem üblichen Wahrscheinlichkeitszugang deutlich mühsamer durch Einführung und Begründung des Satzes von der totalen Wahrscheinlichkeit (2.Pfadregel) abzuleiten. Die Frage, welches Häufigkeitsverhältnis im Baumdiagramm die gesuchte Wahrscheinlichkeitslösung ergibt, konnte mit Führung des Lehrers beantwortet werden (39:48 bis 42:01). Allerdings traten bereits hier (unerwartete) Schwierigkeiten mit dem Prozentbegriff auf. Der Lehrer übergang zunächst in 39:48 den Ansatz des Schülers, indem er von „ins Verhältnis setzen“ spricht, wobei dies aber wahrscheinlich gar nicht die „typische“ Schülerstrategie ist, um Prozente zu bestimmen. Auf die Klärung dieser Fehlvorstellungen zum Prozentbegriff wird an anderer Stelle näher eingegangen (vgl. 3.5.2). Festzustellen ist, dass bereits nach 45 Minuten im Unterrichtsprozess das Ziel der mathematischen Modellierung eines realitätsbezogenen Bayesproblems erreichbar war. In traditionellen Wahrscheinlichkeitszugängen kann dieses Ziel im Allgemeinen erst nach Stunden der Regelherleitung und -begründung erreicht werden.

3.5.2 Vertiefung und Interpretation

Der Vergleich mit den Schätzungen der Schüler führte zur erwarteten Verwunderung und auch zu Zweifeln an der Richtigkeit des Lösungsweges (Stunde 1, 43:12). Damit wurden bereits Fragen zur Interpretation des Ergebnisses und Validierung des Modells aufgeworfen, die in folgenden Unterrichtsstunden zu vertiefen waren. Das überraschende Ergebnis schaffte die Motivation dafür, sich nochmal genauer zu besinnen.

Die folgenden Transskriptausschnitte (vgl. Tab.IV.21a-c) haben eine Vertiefung des rechnerischen Ergebnisses von Arbeitsblatt 1 in der 2. Unterrichtsstunde zum Inhalt. Ein Schüler entwickelte und erklärte kompetent und eigenständig den in der 1. Stunde gemeinsam erstellten Häufigkeitsbaum an der Tafel (2:30 bis 6:35, Abb.IV.29). Die Beschriftung der einzelnen Baumknoten, die für die sachgerechte Interpretation der Darstellung nötig ist, wurde vom Lehrer durch Zuruf der Schüler ergänzt (6:48 bis 8:22). Dabei wurde auch kontextuell der Unterschied zwischen „HIV-infiziert“ und „AIDS-erkrankt“ deutlich gemacht (7:01). Von Schülerseite wurden nochmal Erstaunen und Zweifel über die hohe Anzahl an falsch-positiv Getesteten laut (8:22).

Unterrichtsstunde 2 Materialgrundlage: Arbeitsblatt 1 und 2 Themen: Wie sicher ist der AIDS-Test und das AIDS-Testverfahren im Detail	
Zeit	Unterrichtsdialog (Abkürzungen: L: Lehrer, S: Schüler/in, LE: Lehrerecho, AB: Arbeitsblatt)
2:00	L: So, ich möchte jetzt einen von euch bitten, das Baumdiagramm, das wir letzte Stunde entwickelt haben und das ihr zuhause nochmal alleine hingemalt habt, an die Tafel zu zeichnen und zu erklären, welche Überlegungen da alle drinstecken.
2:30	[Schüler erklärt sich bereit und beginnt an der Tafel einen Häufigkeitsbaum ohne Bezeichnungen zu erstellen. Dann erklärt der Schüler die eingesetzten Werte, vgl. Abb.IV.29] <div style="text-align: center;">  </div>
4:52	S [deutet jeweils auf die entsprechenden Werte im Baumdiagramm]: So, das ist erstmal die gesamte Einwohnerzahl Nordrhein-Westfalens laut der Statistik. Und dann hatten wir ja die Prozentual-, äh die Prozentsätze von denen. Also 0,05% haben davon AIDS, statistisch gesehen. Das sind dann 9000. Und wir hatten ja Wahrscheinlichkeiten. Die Wahrscheinlichkeit, dass der Test, wenn einer AIDS hat, positiv auch ausgeht waren 99,9%. Das sind dann bei den 9000, die auch AIDS haben, 8991, bei denen der Test das entdeckt. Und bei 9 wird der Test HIV-negativ, obwohl die AIDS haben, gehört also auch zu den 9000. So [geht auf die andere Hälfte des Baumes], das sind die 17991000, die kein AIDS haben von den 18 Millionen. Hm, bei 53973 von denen wurde aber trotzdem AIDS entdeckt, obwohl die keins haben.
	L: Was heißt entdeckt?
	S: Also, der Test sagt das. Die haben in Wirklichkeit kein AIDS, aber der Test sagt, sie hätten es.
	L: Der Test ist also positiv.
	S [fährt fort]: Ja, der Test ist positiv. Und bei 17937027 sagt der Test kein AIDS, das waren 99,7%. Das war die Wahrscheinlichkeit. So, dann war ja die Frage, bei wie vielen schlägt der Test positiv an, das sind die 53973 von denen, die kein AIDS haben und die 8991, die AIDS haben.
6:35	L: Gut, danke schön. Fantastisch.
6:48	L: Also, noch einmal. Auf der ersten Stufe, wonach wird da unterschieden?
	S: Die, die AIDS hatten und die, die kein AIDS hatten.
7:01	L: Ja, wir sollten uns vielleicht einigen auf HIV-infiziert [ergänzt die Bezeichnung im Baumdiagramm]. Weil AIDS haben und HIV-infiziert sein, ist ja nochmal ein Unterschied. Wer weiß schon, wo der Unterschied liegt?

	S: Ja, AIDS ist die Krankheit und HIV ist der Virus.
	S [anderer]: Ja, aber das hat doch jeder in sich. [Unruhe bei den anderen Schülern]
	L: Nein, nein. Das ist eindeutig nicht richtig!
	S [anderer, erklärend]: Das sieht man doch, nur 9000 von 18 Millionen sind HIV-infiziert.
	L: Wenn alle HIV-infiziert wären, bräuchte man doch den Test gar nicht mehr machen, oder? Es hat nicht jeder den HIV-Virus in sich. Es ist aber möglich, wenn man den HIV-Virus in sich hat, dass die Krankheit AIDS noch nicht ausgebrochen ist.
7:55	L: So, auf der zweiten Stufe, wie kann man das jetzt clever bezeichnen.
	S: Mit Testergebnis positiv.
	L: Ja, Test positiv, negativ, positiv, negativ [schreibt die Testergebnisse an die entsprechenden Knoten im Baum]
8:22	S: Also, ich versteh das jetzt gar nicht so richtig. Dass da fast 54000 den Test positiv haben, wenn sie nicht HIV haben.
	S [anderer]: Ja, aber von 18 Millionen gesehen ist das ja schon gar nicht mehr so viel.
	L: O.k., vielleicht kommen wir da nochmal auf die Frage von letzter Stunde zurück und dann kommen wir nochmal auf das Problem zurück, das du gerade angesprochen hast.
	S [wie in 8:22]: Ja, ich fand das mit den 14,3% eigentlich unglaublicher als mit den 99,9%, was wir vorher angenommen hatten.
	L: O.k., ich denke, wir sollten uns jetzt nochmal genau überlegen: Wie kamen wir denn auf diese 14,3%? Was geben die 14,3% an?
10:52	S: Wie viele von den positiv Getesteten auch HIV-infiziert sind.

Tab.IV.21a: Unterrichtsszene zur Vertiefung der ersten Modellierung (Aufgabe 5 und 6, AB 1)

Da auch technische Schwierigkeiten bei den Schülern erkennbar waren, entschied sich der Lehrer zunächst für eine *Wiederholung der Prozentrechnung* (10:55 bis 18:17). Die Richtigkeit des Ergebnisses konnte nun mit den nötigen technischen Mitteln (Berechnung von Prozentwerten aus Häufigkeiten), die eigentlich bereits geläufig sein sollten, verifiziert werden (18:17 bis 22:00). Meiner Meinung nach zeigt das Beispiel, dass das „Formelwissen“ zur Berechnung eines Prozentsatzes den an sich einfachen und intuitiven rechnerischen Ansatz der Division zweier natürlicher Zahlen eher behindert (21:51).

Fortsetzung Unterrichtsstunde 2	
10:55	L: Ja. [LE, schreibt dies als Frage an die Tafel]. So, jetzt wurden letzte Stunde wilde Rechnungen über Dreisätze vorgeschlagen. Ich würde sagen, wir wiederholen nun einfach kurz an der Tafel die Prozentrechnung. „Schnellkurs“ Prozentrechnung.
...	[<i>Wiederholung der Prozentrechnung</i>] [Begriffsklärung Grundwert G, Prozentsatz p, Prozentwert W; $W = G \cdot p / 100$, Umformung zu $p = W / G \cdot 100$]
18:17	L: Was ist der Grundwert? Worauf beziehen wir uns.
	S: Auf alle positiv Getesteten, die 62964.
	L: Also, das ist der Grundwert: „Wie viele von diesen...“ [Rahmt den Ausdruck in der Frage ein und schreibt an: $G = 62964$]
	L: Was ist mein Prozentwert? Los, da müsst ihr alle wissen.
	S: Alle, die davon wirklich HIV infiziert sind.
	L: Was heißt wirklich?
	S: Also das sind irgendwie alle, die den Test positiv haben und wirklich HIV infiziert sind.
	L: Also die 9000.
	S: Nein, die sind ja noch nicht getestet worden.
	L: Was müssen also Bedingungen für „wirklich positiv“ sein.
	S: Die müssen positiv getestet sein.
	S [anderer]: Die müssen auch wirklich infiziert sein.
	L: Ja, das sind also hier die 8991. Also was ist der Prozentwert?
	S: 14,3
	L: Wie kommt man nun auf 14,3%?
	S: 8991 durch 62964.
21:51	L: Ja, das tippt ihr einfach in den Taschenrechner und dann steht da rund 0,143. Und dann wisst ihr entweder, dass das 14,3% sind oder ihr nehmt mal 100, dann habt ihr den Prozentwert von 14,3. Klar?

Tab.IV.21b: Unterrichtsszene zu Problemen bei der Prozentrechnung (zu Aufgabe 5 und 6, AB 1)

Das vermeintlich niedrige Ergebnis wurde schließlich durch Rückkopplung mit der außermathematischen Situation interpretiert (ab 22:00). Aus dem vom Lehrer gelenkten Unterrichtsgespräch entwickelte sich allmählich eine sekundäre Bewertung des Ergebnisses, das vorher lediglich als „erstaunlich niedrig“ von den Schülern beurteilt wurde. Nun wurde herausgearbeitet, dass das Ergebnis relativ zu dem vorherigen Informationsstand einen deutlichen Zuwachs bedeutet (ab 25:40 bis 29:40; vgl. Aufgabe 6, AB1). Bereits jetzt fanden die Begriffe A-priori- und A-posteriori-Wahrscheinlichkeit Erwähnung (29:40), die als Schlüsselbegriffe im Laufe der Unterrichtsreihe immer wieder verwendet wurden. Dass die gewonnene Sicherheit über eine HIV-Infektion nach einem positiven Test zu gering ist, war ein Schülerfazit. Auch eine Folgerung aus dieser Einschätzung, nämlich die Notwendigkeit zu weiteren Tests, wurde eigenständig genannt (30:39). Das Konzept „Informationsgewinn“ durch weitere Tests (Arbeitsblatt 2) folgerten die Schüler gewissermaßen intuitiv aus der Bewertung der Ergebnisse.

Fortsetzung Unterrichtsstunde 2	
22:00	L: So, und jetzt kommen die Fragen von S. und O. [Schüler] von vorhin.
	S [wiederholt]: Also, ich finds verwunderlich, dass wenn man einen Test macht und der Test ist positiv, dass man meist gar nicht HIV hat, also nur zu 14,3%.
	L: Das geht in dieselbe Richtung wie die Frage von O. von der letzten Stunde. Was sagen die anderen dazu?
	S: Die Frage ist ja, von den positiv Getesteten, wie viele haben HIV.
	L: Ja, aber was bedeutet das jetzt. Ihr geht da hin, habt ein positives Testergebnis, und?
	S: Ja, dann kann man sich nur zu 14,3% sicher sein, dass man überhaupt HIV hat.
	S [mehrere andere]: Das ist aber wenig.
	L: Das ist wenig, oder? Wer findet das viel? [keine Meldung]
	L: Kann man den einzelnen Test noch verbessern?
	S: Ich denke nicht, dass man ihn verbessern kann, weil der ist ja schon bei 99,9%. Ich glaub kaum, dass man da viel besser sein kann.
	L: Ja, also die Werte des Tests sind ja super, 99,9% und 99,7%. Aber trotzdem kommt heraus, dass nur 14,3%, also 14 von 100 Leuten, HIV-infiziert sind, wenn sie ein positives Ergebnis haben.
	S: Das liegt daran, dass ich 18 Millionen teste.
	L: Du meinst die Quote kommt nur raus bei 18 Millionen Menschen?
	S [anderer]: Nein, die würde immer rauskommen, das ist egal.
	L: Warum kommt die immer raus?
	S [fährt fort]: Ja, wenn wir weniger Menschen haben, dann auch weniger Infizierte und auch weniger positiv Getestete.
	L: Ja, also die Werte des Tests sind ja super, 99,9% und 99,7%. Aber trotzdem kommt heraus, dass nur 14,3%, also 14 von 100 Leuten, HIV-infiziert sind, wenn sie ein positives Ergebnis haben.
	L: Soll man den Test denn überhaupt machen? Was bringt der denn dann? Also, stimmt ja, 14,3% sind ja nicht so besonders toll.
	S: Ja, das würde ja bedeuten, dass bei 6 positiven Menschen nur einer wirklich HIV hat.
	L: Naja, durchschnittlich kommt das hin. Also bringt das gar nix?
	S: Eigentlich schon, denn wenn man HIV hat, kann man sich ja zu 99,9% sicher sein, dass der Test dann positiv ist.
25:30	L: Ja, aber weißt du, ob du HIV hast? S: Nö. L: Weißt du es hinterher? S [fährt fort]: Nö.
25:40	S [anderer]: Ja, aber man weiß es besser als vorher.
	L: Aha, das ist schonmal gar nicht schlecht. Was heißt, man weiss es besser als vorher?
	S [fährt fort]: Vorher hat man ja gar keine Ahnung.
	L: Gar keine Ahnung ist übertrieben. Wie groß war die Wahrscheinlichkeit vorher?
	S [fährt fort]: Ja, o.k., 0,05%. Und jetzt ist der Wert schon ein bisschen höher.
	L: Ein bisschen?
	S [fährt fort]: Ja, o.k., vorher hat man zu 0,05% AIDS und wenn er dann positiv ausfällt, kann man schon sagen, dass man zu 14,3% AIDS hat.
	L: Und das ist ja schon mal auch ein Informationsgewinn.
26:56	[Es wird an der Tafel nochmal die Basisrate und die Wahrscheinlichkeit für HIV-Infektion nach positivem Test angeschrieben]
	L: Wie hat sich die Wahrscheinlichkeit verändert, dass man HIV-infiziert ist?
	S: Die ist höher
	L: Und um wie viel ist sie höher?
	S: Ca. 300mal

	S [anderer]: 286mal
	L: Also, die Wahrscheinlichkeit ist um das 286fache gestiegen. Wie bist du auf die 286 gekommen?
	S [fährt fort]: 14,3% durch 0,05%
29:40	[Die Begriffe „a priori Wahrscheinlichkeit“ und „a posteriori Wahrscheinlichkeit“ werden eingeführt]
30:39	S: Wenn man jetzt aber noch einen Test macht und der fällt wieder positiv aus, dann ist man sich aber nicht zu 28,6% sicher, dass man AIDS hat, oder?
	L: Ja das ist eine sehr gute Frage. R. hatte vorhin schon bemerkt, dass man hinterher einen zweiten Test machen sollte. Natürlich ist es so, man erhält ein positives Testergebnis und wird natürlich relativ beunruhigt sein. Und möchte dann natürlich genau wissen: Bin ich HIV-infiziert oder bin ich nicht HIV-infiziert. Und das ist auch richtig, dass da ein zweiter Test kommt. Dazu kriegt ihr jetzt einige Informationen
32:33	[L. teilt AB 2 aus] L: Welche Leute sollten denn jetzt noch einen zweiten Test machen? [.....]

Tab.IV.21c: Unterrichtsszene zur Vertiefung des Konzepts „Informationsgewinn“ (zu Aufgabe 6, AB 1)

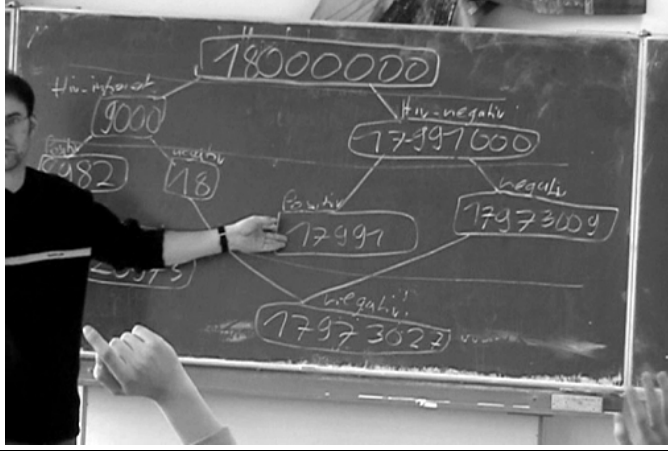
3.5.3. Spontaner Anstoß von realitätsbezogenen Fragestellungen

Eine besondere Gelegenheit zur Motivation entsteht, wenn Schüler von sich aus sinnvolle Fragen aufwerfen. Folgende Transkriptausschnitte aus Unterrichtsstunde 4 geben ein Beispiel (vgl. Tab.IV.22a-d). Ein Schüler wunderte sich, warum die Reihenfolge des Testablaufs gerade „ELISA“ und dann „Western-Blot-Test“ ist (vgl. Arbeitsblatt 2). Er vermutete, dass eine hohe Sicherheit bei richtig-positivem Ergebnis womöglich bereits durch einen einzelnen Western-Blot-Test zu erreichen sei. Die korrespondierende Fragestellung, mit welcher Wahrscheinlichkeit ein Bewohner aus NRW HIV-infiziert ist, wenn er bei einem Western-Blot-Test (ohne vorherigen ELISA-Test) ein positives Ergebnis erhält, konnte spontan gestellt werden. Die mathematische Modellierung war nun bereits als Reproduktionswissen abrufbar (ein Schüler an der Tafel, 2:05 bis 5:50). Schwierigkeiten beim Zeichnen an der Tafel reduzierten eventuell die Anschaulichkeit für andere Schüler, so dass ein wenig Hilfestellung von Lehrerseite nötig war (5:50).

Unterrichtsstunde 4 Materialgrundlage: Arbeitsblatt 2 Thema: Das AIDS-Testverfahren im Detail	
Zeit	Unterrichtsdiallog <i>(Abkürzungen: L: Lehrer, S: Schüler/in, LE: Lehrerecho, AB: Arbeitsblatt)</i>
1:20	L: Letzte Stunde hat R. [Schüler] die Frage gestellt: „Warum macht man den Western-Blot Test nicht sofort?“ Eure Aufgabe war nun, als erstes zu berechnen: Wie groß ist die Wahrscheinlichkeit, dass man HIV-infiziert ist, wenn man beim Western-Blot Test ein positives Ergebnis erhält? Wer zeigt das an der Tafel?
2:05	L: M., würdest du das übernehmen. Das ist schön. [Schüler M. vervollständigt Baumdiagramm zur gestellten Aufgabe an der Tafel, ausgehend von 18 Mio. Einwohnern in NRW und schreibt die Ergebnisberechnung rechts daneben]
5:50	L: Der Baum ist ja etwas verschoben, ich werde das mal in die verschiedenen Stufen einteilen, damit das bisschen klarer ist [zieht horizontale Linien, welche die Baumebenen abteilen, vgl. auch Abb.IV.30]
6:15	L: Gehen wir das mal schrittweise durch, was hast du gemacht als erstes?
	S: Erst hab ich 18 Mio. Einwohner aus NRW genommen, davon sind 0,05% infiziert, also 9000. Und 17991000, der Rest halt, nicht. 99,8% von den 9000 sind dann positiv.
	L: Genau, Western-Blot Test: 99,8% der HIV-infizierten erhalten ein positives Testergebnis, und 99,8% von 9000 sind 8982. Weiter!
	S [fährt fort]: Von den 9000 werden also 18 negativ getestet. Auf der anderen Seite sind dann halt von den 17991000 nicht HIV-infizierten 17991 mit positivem Testergebnis.
	L: Wie kommst du auf diese 17991?
	S [fährt fort]: Weil der Test ist zu 99,9% da negativ auch. Also 0,1%.
	L: Genau, 0,1% von 17991000 sind natürlich 17991. Gut, jetzt hat M. [der Schüler, der an der Tafel war] das schon mal weiter fantastisch zusammengefasst.

Tab.IV.22a: Unterrichtsszene zur Modellierung einer spontanen realitätsbezogenen Fragestellung


Modellierung und Ergebnisberechnung wurden reflektiert, insbesondere sollten die vorkommenden Ereignisse genauer abgegrenzt werden (vgl.Tab.IV.22b, Abb.IV.30). Es sollte vorgebeugt werden, dass Fehlvorstellungen durch bezeichnerische Ungenauigkeit entstehen (7:56 bis 9:54). In dieser Phase zeigte sich sehr deutlich, dass eine strenge Beachtung der sprachlichen Genauigkeit besonders in diesem realitätsbezogenen Kontext unbedingt erforderlich ist und Schüler bei den begrifflichen Vertiefungen besonders zu unterstützen sind. In diesem Zusammenhang sind einfache Ereignisse „positiver Test“ und konjunktive Ereignisse „postiver Test und HIV“ strikt zu unterscheiden (8:58 bis 9:54).

Fortsetzung Unterrichtsstunde 4	
7:56	L: Wie entsteht diese Feld? [deutet auf alle positiv Getesteten] S: Alle Positiven.
8:08	L: Hm, jetzt müssen wir aber ein bisschen aufpassen an dieser Stelle. S: Alle, die dem Test nach positiv getestet wurden. L: Ja, aber hier steht ja auch überall positiv dran [deutet auf die Felder für positiv und HIV-infiziert bzw. positiv und nicht HIV-infiziert]
8:23	L: Wo ist der Unterschied zwischen diesem positiv, diesem positiv und diesem positiv [deutet nacheinander auf die Felder, die sich auf positive Testergebnisse beziehen, vgl. Abb.IV.30]. Oder, gibt's überhaupt einen Unterschied? 
	S: Die 8982 sind wirklich positiv. L: Ja, was heißt wirklich positiv? S: Ja, die sind HIV-infiziert. L: Aber, das sind doch die? [deutet auf das Feld mit allen 9000 HIV-Infizierten] S [fährt fort]: Ja, aber ich meine nach dem Test HIV-infiziert. L: Ja, die 9000 sind aber nach dem Test auch immer noch HIV-infiziert. Man wird doch nicht dadurch geheilt, dass man ein negatives Testergebnis kriegt. Wäre zwar schön, aber...Ihr müsst das präzise formulieren!
8:58	S: Also, da gibt's die HIV-infizierten und davon wurden welche positiv getestet, und dann gibt's die nicht HIV-infizierten und davon wurden auch welche positiv getestet, und dann nimmt man alle zusammen. L: Was ist das hier also? [deutet auf das Feld mit 8982] S: Die sind HIV-infiziert und wurden auch positiv getestet. [LE, L ergänzt an dieser Stelle an der Tafel „und HIV-infiziert“] L: Und das? (deutet auf das Feld mit 17991) S [fährt fort]: Positiv und nicht HIV-infiziert [LE, L ergänzt an dieser Stelle „und nicht HIV-infiziert“]
9:44	L: Und das sind dann? (deutet auf das Feld mit 26973) S [fährt fort]: Alle positiv Getesteten.
9:54	L [LE]: Ja, alle mit positivem Testergebnis.

Tab.IV.22b: Unterrichtsszene zur Präzisierung von Bezeichnungen

Besondere Strukturen der Häufigkeitsbäume konnten nun herausgestellt werden: Die Häufigkeiten in jeder „Baumebene“ summieren sich immer zur Gesamtzahl 18 Millionen (9:55 bis 11:00; Tab.IV.22c). Bei der Ergebnisberechnung (11:00 bis 13:02) zeigten sich wiederum die bereits erwähnten deutlichen Schwierigkeiten im Bereich der Prozentrechnung, die auf nicht gefestigtes Wissen aus früheren Stufen zurückzuführen sind (12:03 bis 13:02).

Es stellt sich die Frage, welche Grundvorstellungen in diesem Bereich vorhanden und ob diese durch formale Behandlung hinreichend förderbar sind. Es ist offensichtlich, dass eine umfangreiche Vertiefung der Prozentrechnung als Voraussetzung stochastischer Themen dringend erforderlich war.

Fortsetzung Unterrichtsstunde 4	
9:55	<p>L: Was ergibt sich denn, wenn wir dieses hier noch zusammenfassen? [ergänzt an der Tafel den Baum weiter nach unten, also zeichnet Pfade von den untersten beiden Feldern nach unten]</p>  <p style="text-align: right;">Abb.IV.31</p>
	[Die Schüler überlegen, einige beginnen zu rechnen]
	S: Ja, da kommt man wieder auf die 18 Millionen.
	L: Warum das denn?
	S [fährt fort]: Ja, weil das sind ja wieder alle zusammen, also die positiv und die negativ Getesteten.
10:32	L: Ja klar, sozusagen muss ich ja entweder ein negatives oder ein positives Testergebnis kriegen. Also hier habe ich wieder meine 18 Millionen [ergänzt „18000000“ unter das bisherige Diagramm an der Tafel]
10:34	L [zusammenfassend]: Im Prinzip hab ich also jetzt 5 Stufen [deutet auf die Baumebenen], wobei jeweils oben und unten 18 Millionen stehen. Warum?
10:45	S: Also, ich würde sagen, auf jeder Stufe sind 18 Millionen, es werden nicht mehr oder weniger.
	L: Richtig, also wenn ich alle 4 Kästchen hier zusammenzähle [deutet auf die 3. Ebene], muss ich natürlich auch auf 18 Millionen kommen. Mit anderen Worten: Die 18 Millionen teilen sich immer nur auf die unterschiedlichen Felder auf.
11:00	L: So, jetzt war die Frage: Wie groß ist die Wahrscheinlichkeit, dass jemand, der ein positives Testergebnis kriegt, auch wirklich HIV-infiziert ist. Welche beiden Größen muss ich also dabei in Beziehung setzen?
	S: Die, die infiziert sind und den Test auch positiv haben.
	L [unterbricht]: Ja gut, da steckt das Wörtchen „und“ drin und das ist ganz wichtig an dieser Stelle! Wo sind die?
	S [fährt fort]: Die 8982. [Lehrer markiert dieses Feld mit Farbe]
	S [fährt fort]: Und dann die Zahl, die einen positiven Test hatten.
	L: Wo sind die?
	S [fährt fort]: Die 26973. [Lehrer markiert dieses Feld mit Farbe]
12:03	L: Mit anderen Worten, ich möchte wissen, wie viel Prozent sind 8982 von 26973? Und M. hat uns hier eine fantastische Rechnung hingeschrieben. [Der Schüler M. hatte bereits vorher eine Lösung an die Tafel geschrieben: $(8982 / 26973) \cdot 100 = 33,3\%$]
	L: Wir wollen einen Prozentwert ausrechnen? Wie geht das?
12:12	S: $P\% = W$ durch G mal 100.
	L: Das ist aber eigentlich falsch, was da steht. Warum darf ich da nicht mal 100 hinschreiben? [Unruhe bei den Schülern]
	S: Prozent heißt ja „Hundertstel“, also die 100 müssen dann weg.
	L: Wenn ihr das Prozentzeichen hinschreiben wollt, dann müsst ihr die 100 weglassen.
	S [wie 12:12]: Das wurde uns aber falsch erklärt.
	L: Würd ich so nicht sagen, vielleicht erinnert ihr euch da nicht ganz richtig. [Gelächter]
13:02	L: Also das mal 100 muss hier weg und dann sind das 33,3% [verbessert das Ergebnis und schreibt die richtige Formel an die Tafel: $p\% = W / G$]

Tab.IV.22c: Unterrichtsszene zur Vertiefung von Strukturen und Lösung mit Häufigkeitsbäumen

Die Rückkopplung des Ergebnisses mit der (von Schülern) gestellten Frage zeigte, dass die Schüler bereits sehr kompetent interpretativ weiterdenken konnten (13:05 bis 14:23; vgl. Tab.IV.22d). Die Folgerungen von Schülerseite waren, dass der Western-Blot-Test zwar „besser“ als der ELISA-Test (bei der Erkennung richtig-positiver Testergebnisse), doch keinesfalls „gut genug“ ist, so dass nicht auf einen Bestätigungstest verzichtet werden kann. Diese Unterrichtsszene belegt, dass sehr vernünftig von Schülern interpretiert werden konnte, wenn der Kontext entsprechende realistische Fragestellungen anstößt.


Fortsetzung Unterrichtsstunde 4	
13:05	S: Ja 33,3% ist ja nicht soviel wie 99,4% von vorhin. [meint die errechnete Sicherheit bei nacheinander positivem ELISA und Western-Blot-Test]
	L: Ja, aber wie haben wir die 99,4% erreicht?
	S [fährt fort]: Durch beide Tests.
	L: Ja, aber was können wir auf jeden Fall sagen?
	S: Western-Blot-Test ist zwar besser als ELISA-Test, aber ist auch nicht gut genug.
	L: Was müsste man also auch hier auf jeden Fall machen?
	S: Noch einen Test mit den Positiven.
	L: Also, um eine hohe Sicherheit zu kriegen, muss man auf jeden Fall zwei Tests machen. Warum fange ich dann mit dem ELISA-Test an?
	S: Er ist billiger. [wurde bereits vorher erwähnt]
14:23	L: Ja, er ist billiger, das ist der Grund.

Tab.IV.22d: Unterrichtsszene zur Interpretation von Ergebnissen

3.5.4 Vertiefungen zum Einfluss der Basisrate und Darstellungserweiterung

Eine wichtige Erkenntnis bei Bayesianischen Problemen ist der Einfluss von Basisraten. Im Kontext „AIDS und Diagnose“ ist eine darauf zielende Frage, woran es liegt, dass die Testsicherheit nicht ausreicht. Der *Einfluss der Basisrate* wurde anhand des Arbeitsblattes 4 erarbeitet (vgl. Tab.IV.23a). Es wurde nach erneuter Modellierung mit Werten einer sehr hohen Risikogruppe (z.B. intravenös Drogenabhängige in Berlin) deutlich, dass das Ergebnis bei vergleichsweise hoher Basisrate stark ansteigt (28:35 bis 31:17, vgl. Abb.IV.32). Das Ergebnis 97,7% konnte nun sehr genau interpretiert werden (32:50). Schüler verbesserten von sich aus sprachliche Ungenauigkeiten (33:12 bis 33:42). Der entscheidende Einflussfaktor wurde bereits in der Schülerformulierung „848 ist relativ hoch gegenüber 6652“ (34:11) sehr genau erkannt, nämlich der Anteil der HIV-Infektionen an der Gesamtzahl (34:11 und 35:26).


Unterrichtsstunde 4 (Fortsetzung) Materialgrundlage: Arbeitsblatt 4, Aufgaben 1-4 Thema: Einfluss der Basisrate	
Zeit	Unterrichtsdialog (Abkürzungen: L: Lehrer, S: Schüler/in, LE: Lehrerecho, AB: Arbeitsblatt)
23:10	[AB 4 wird verteilt und der Text laut von einem Schüler vorgelesen. Es folgt Diskussion über die Frage, warum intravenös Drogenabhängige besonders oft HIV-infiziert sind. Es wird deutlich, dass sie besonders gefährdet sind, sich mit HIV anzustecken. Hauptgrund: Ansteckung über das Blut durch die mehrfache Verwendung von Spritzen]
24:50	[Aufgabe 1 wird vorgelesen, wurde als Hausaufgabe vorbereitet]
25:38	[Ein Schüler erstellt an der Tafel einen zugehörigen Häufigkeitsbaum, jedoch ohne Beschriftung der Ereignisse. Lehrer überprüft Hausaufgaben und Lösungen der Schüler in den Heften]
28:35	L: Ich würde euch empfehlen, immer an die einzelnen Stufen genau dranzuschreiben, was da unterschieden wird. L: Was steht ganz oben? [Beschriftung wird nochmal gemeinsam entwickelt]
28:55	S: Das sind alle, also alle IVDAs. [LE, ergänzt ganz oben „IVDA“] S [fährt fort]: Dann HIV infiziert oder nicht. [LE, ergänzt rechts neben 2.Ebene „HIV-infiziert“ und an die Knoten „ja“ bzw. „nein“]
	S [anderer]: Dann infiziert und positiver ELISA-Test, infiziert und negativer Test, und nicht infiziert und positiver Test und nicht infiziert und negativer Test. [LE, ergänzt entsprechende Bezeichnung an der Tafel]
29:53	S [fährt fort]: Alle mit positivem und negativem Test. [LE, ergänzt entsprechend]
30:35	S: 99,9% von 848 kommt aber 846,5 oder so was raus. Kann man da auch runden, eigentlich auf 847. L: Ja, das ist jetzt die Frage, ob wir auf 846 oder auf 847 runden. Wer ist für 847, wer für 846? (Mehrheit für 847)

<p>30:55</p>	<p>[L. verbessert Werte entsprechend im Baum an der Tafel]</p>  <p style="text-align: right;">Abb.IV.32</p>
<p>31:17</p>	<p>L: Bis dahin Fragen? [keine Meldungen] Jetzt die Antwort! Wie viele der positiv Getesteten sind wirklich HIV-infiziert?</p>
<p></p>	<p>[L schreibt gesuchte Wahrscheinlichkeit an die Tafel] S [diktiert Lösung]: 847 durch 867 mal 100 ist gleich 97,7%. [LE, ergänzt die Lösung]</p>
<p></p>	<p>L: Nein das ist 97,7. Lass doch das „mal 100“ hier weg [wischt „ · 100“ wieder weg]. Dann sind das 97,7%.</p>
<p>32:30</p>	<p>L: So, was heißt das jetzt also, wenn ich als intravenös Drogenabhängiger zum ELISA-Test gehe?</p>
<p>32:36</p>	<p>S: Ich kann davon ausgehen, dass ich auch HIV-infiziert bin.</p>
<p></p>	<p>L: Das ist aber jetzt zu radikal formuliert. Geht aber in die richtige Richtung. Ich würde es etwas vorsichtiger formulieren.</p>
<p>32:50</p>	<p>S: Da kann man sich zu 97,7% sicher sein, HIV-infiziert zu sein. [Unruhe unter den Schülern, es wird reingerufen]</p>
<p></p>	<p>L: Ja, die Wahrscheinlichkeit ist sehr groß, HIV-infiziert zu sein.</p>
<p></p>	<p>L: Was denn? Nochmal bitte.</p>
<p>33:12</p>	<p>S: Ja sie haben ja gefragt, wie sicher kann man sich sein HIV-infiziert zu sein, wenn man zum ELISA-Test geht, aber ...</p>
<p></p>	<p>L: O.k., haste völlig recht, die Frage war eigentlich relativ mies. Kannst Du sie besser formulieren?</p>
<p>33:26</p>	<p>S [fährt fort]: Wie sicher kann man sich sein, dass man HIV-infiziert ist, wenn der ELISA-Test positiv ist. [Andere Schüler bekunden gleiche Meinung]</p>
<p></p>	<p>L: Ja, natürlich. Sehr gut, gut aufgepasst.</p>
<p>33:42</p>	<p>L: Aber, wo ist denn der Unterschied jetzt. Warum kommt denn jetzt bei den IVDA so ne relativ hohe Wahrscheinlichkeit raus und warum kommt bei der Gesamtbevölkerung eine geringe Wahrscheinlichkeit raus?</p>
<p>34:02</p>	<p>S: Ja, die können das schneller gegenseitig an sich weitergeben.</p>
<p></p>	<p>L: Hm, ich möchte das jetzt aber mal konkreter an den Zahlen erklärt haben.</p>
<p>34:11</p>	<p>S: Weil eben 848 HIV-infiziert sind und das ist ja relativ hoch gegenüber den 6652.</p>
<p></p>	<p>L: Aha, also scheint irgendwie diese Zahl 848 bzw. der prozentuale Anteil zu allen eine große Rolle zu spielen [deutet auf die Verbindung der Felder mit 848 und 7500].</p>
<p>34:35</p>	<p>S: Also bei der Gesamtbevölkerung haben ja von über 17 Millionen nicht Infizierten 0,3% ein positives Testergebnis gekriegt, aber das waren ja total viele. Aber da [meint im Falle der IVDA] haben ja auch total viele im Verhältnis HIV zu denen die nicht HIV haben und deswegen sind das so viele von denen die ein positives Ergebnis gekriegt haben.</p>
<p></p>	<p>L: Im Prinzip ist das noch ein bisschen anders aufgerollt, was S. vorher schon gesagt hat. Aber wir werden das noch die nächsten Stunden weiter vertiefen.</p>
<p>35:26</p>	<p>L: Das wäre die Antwort auf Frage 2 schon: Der entscheidende Punkt liegt hier oben [deutet auf 848]. Weil der Anteil der HIV-infizierten bei den IVDA viel höher ist als bei der Gesamtbevölkerung.</p>

Tab.IV.23a: Unterrichtsszene zum Konzept „Einfluss der Basisrate“

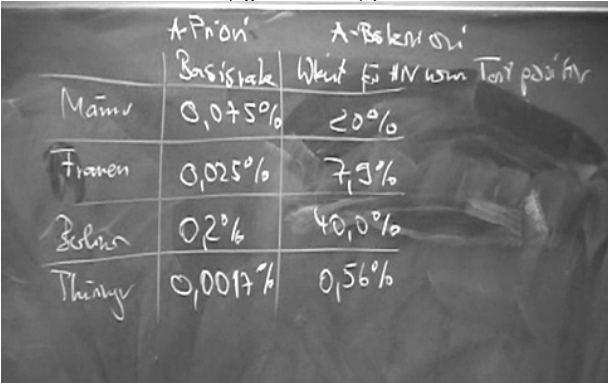
Anhand dieser Aufgaben wurde nun auch eine zweite Darstellungsform, die *Häufigkeitstabelle*, an der Tafel im direkten Vergleich mit dem erstellten Häufigkeitsbaum entwickelt (ab 35:50; vgl. Tab.IV.23b und Abb.IV.33). Eine Übertragung von Häufigkeitswerten zwischen Baum- und Tabellenstruktur war für die aufgerufenen Schüler problemlos möglich. Farbhervorhebungen konnten den systematischen Zusammenhang der beiden Darstellungsformen gut veranschaulichen (z.B. Welche Tabellenteile entsprechen welchen Teilen im Baumdiagramm?, vgl. 39:18). Insbesondere wurde herausgearbeitet, dass die Summe bestimmter Häufigkeiten jeweils die Gesamtzahl ergibt, also in den Darstellungen gewissermaßen verschiedene Aufspaltungen der Gesamtmenge nach unterschiedlichen Gesichtspunkten erfolgen (42:33, Abb.IV.33). Dass mit beiden Darstellungen unterschied-

liche Schlussrichtungen möglich sind, war weiter zu vertiefen (43:30; Thema in Unterrichtsstunde 7, vgl. Arbeitsblatt 5).

Fortsetzung Unterrichtsstunde 4	
35:50	[L beginnt <i>Tabelle</i> an die linke Tafelhälfte zu zeichnen, mit Beschriftung der Spalten und Zeilen, aber ohne Werte; vgl. Aufgabe 3 und Abb.IV.33]
36:47	L: O.k., also Bäume findet man ja nicht nur, sondern auch häufig Tabellen. Ziel ist jetzt, einen Zusammenhang zwischen dieser Tabelle und Bäumen zu entdecken. Füllen wir mal aus. S: Die erste Spalte. Gesamt sind 7500, und aus dem Text können wir entnehmen dass 848 HIV-infiziert sind. [L trägt jeweils die von Schülern genannten Zahlen in Tabelle ein] L [korrigiert]: Dann willst du die erste Zeile machen, seh ich das richtig? S [fährt fort]: Ja, ja. Jetzt wissen wir, dass 6652 nicht infiziert sind. L: Wo finde ich das in meinem Baum wieder? [Das vorher aufgestellte Baumdiagramm steht noch auf der Tafelmitte] S [fährt fort]: Ja, ganz oben ist die Gesamtzahl und links die Infizierten und rechts die nicht Infizierten. [L markiert die entsprechenden Häufigkeiten in Baum und Tabelle mit roter Farbe] L: O.k., weiter.
38:16	S: Also dann zweite Zeile: 867 testen gesamt positiv. Und dann in der Zeile 847, und dann 20. L: Letzte Zeile?
38:48	S: Test negativ sind 6633, davon sind infiziert, äh, einer, und nicht infiziert 6632.
39:18	L: O.k. Ich habe hier jetzt im Prinzip eine „Gesamt“-Zeile [zeigt auf Randsummen in der 1. Zeile], hier hab ich eine „Gesamt“-Spalte [zeigt auf Randsummen in der 1. Spalte]. Wo finde ich denn im Baum die Werte der „Gesamt“-Spalte wieder?
39:38	S: Also die 867 finde ich unten als die alle positiv Getesteten wieder, und daneben die 6633 negativen. [L markiert diese Häufigkeiten in Baum und Tabelle mit grüner Farbe] [L markiert die restlichen inneren 4 Felder der Tabelle mit gelber Farbe] L: Und diese Werte hier? S: Ja, das sind die einzelnen Testergebnisse. Stehen im Baum in der Mitte [L markiert diese Häufigkeiten in Baum und Tabelle auch mit gelber Farbe] [L markiert nun die Gesamtheit 7500 in Baum und Tabelle mit blauer Farbe]. L: Und was ist das Besondere an der Zahl 7500? S: Dass oben und unten wieder 7500 rauskommen. L: Ja, das ist auf den Baum bezogen. Steht einmal oben, und ganz unten können wirs nochmal hinschreiben [ergänzt unten Feld mit „7500“] S [fährt fort]: Und jede Farbe gibt auch 7500. L: Richtig, alle Werte mit der gleichen Farbe zusammengezählt geben auch 7500, warum?
41:43	S: Es müssen ja immer wieder alle sein, positiv und negativ Getestete, HIV-Infizierte und nicht HIV-Infizierte müssen immer zusammen 7500 sein.  Abb.IV.33
	L: Und die 4 Felder? [deutet auf die inneren gelb markierten Felder der Tabelle] S [fährt fort]: Die Test positiv und HIV, Test negativ und HIV, Test positiv und nicht HIV und Test negativ und nicht HIV sind die einzigen Fälle, die es gibt. Also diese 4 Felder müssen auch 7500 ergeben.
42:33	L [fasst zusammen]: Mit anderen Worten: man kann jetzt die Gesamtzahl auf unterschiedliche Weise aufspalten. Man kann sagen, ich betrachte die HIV-Infektion, dann wird in infiziert und nicht infiziert aufgeteilt. Im Baum starte ich von oben. Man kann auch aufteilen nach Testergebnis, positiv und negativ, dann starte ich im Baum unten und gehe nach oben. D.h. wie ich die Tabelle auf unterschiedliche Art lesen kann, von rechts nach links oder von oben nach unten, kann ich auch den Baum unterschiedlich lesen: Einmal von oben nach unten, einmal von unten nach oben.
43:30	S: Das geht doch nicht. Das unten muss man doch erst ausrechnen. L: Warum? Es ist die Frage, was ich gegeben habe. Es kann auch sein, dass ich zunächst nur weiß, wie viele positiv bzw. negativ getestet haben und schließen will, mit welcher Wahrscheinlichkeit ein HIV-infizierter ein positives Ergebnis erhält. Doch das machen wir nächste Stunde genauer.

Tab.IV.23b: Unterrichtsszene zur Darstellungserweiterung mit Häufigkeitstabellen

Aufgabe 8 des Arbeitsblattes 4 wurde in Gruppenarbeit bearbeitet. Jede Gruppe übernahm für je eine der vier verschiedenen Bevölkerungsgruppen (entspricht 4 unterschiedlichen Basisraten) die Lösung des ELISA-Test-Problems. Die Gruppenarbeit dauerte etwa 15 Minuten. Ergebnisse und zugehörige Basisraten wurden in der nächsten Stunde zusammengetragen und an der Tafel festgehalten (vgl. Tab.IV.24; vgl. Abb.IV.34). Die Basisrate wurde auch durch den bereits in der 2. Stunde eingeführten Begriff „A-priori-Wahrscheinlichkeit“ und die Wahrscheinlichkeit für „HIV-infiziert, wenn Test positiv“ mit „A-posteriori-Wahrscheinlichkeit“ bezeichnet. Mit Hilfe der Gegenüberstellung von A-priori und A-posteriori-Wahrscheinlichkeiten gelang es, den Zusammenhang herauszuarbeiten (12:05).

Unterrichtsstunde 6 Materialgrundlage: Arbeitsblatt 4, Aufgabe 8 Thema: Einfluss der Basisrate	
Zeit	Unterrichtsdialog (Abkürzungen: L: Lehrer, S: Schüler/in, LE: Lehrrecho, AB: Arbeitsblatt)
	[In Gruppenarbeit wurde die Aufgabe 8, AB4 bearbeitet. An der Tafel wurden alle A-priori-Wahrscheinlichkeiten bzw. Basisraten der verschiedenen Bevölkerungsgruppen (Männer, Frauen, Berliner, Thüringer, etc.) den errechneten A-posteriori-Wahrscheinlichkeiten gegenübergestellt. Die Begriffe „a-priori“ und „a-posteriori“ wurden auch wieder verwendet (vgl. Stunde 2)]  <p style="text-align: right;">Abb.IV.34</p>
9:30	L: Fällt euch irgendetwas in dieser Tabelle auf? S: Die Wahrscheinlichkeiten links sind ziemlich unwahrscheinlich. S [anderer]: Rechts sind die Wahrscheinlichkeiten alle größer als links. L: Ja, ist das inhaltlich vielleicht auch logisch, dass die Wahrscheinlichkeiten rechts alle größer sind als links? Wofür stehen die linken Werte, wofür die rechten Werte? S: Für die Wahrscheinlichkeit vor dem Test und nach dem Test. L: Nochmal die Frage: Ist es logisch, dass die rechten Wahrscheinlichkeiten größer sind als die linken. (L. fordert, dass alle diese Frage beantworten können müssen. Die meisten Schüler melden sich) Was meint ihr? L [spezifiziert nochmal die Frage]: Ist es logisch, dass die Wahrscheinlichkeit für HIV nach dem Test, wenn der Test positiv ist, größer ist als vorher? S: Ja, sonst würde doch der Test nix bringen.
12:05	S: Ich würde sagen, dass je höher die Basisrate ist desto höher die Wahrscheinlichkeit für HIV, wenn man ein positives Testergebnis hat. L [beweist das mit den Tafelwerten]: Thüringen hat die kleinste Basisrate, das ist auch die kleinste Wahrscheinlichkeit rechts. Dann alle Frauen in Deutschland [zeigt auf die a posteriori Wahrscheinlichkeit, die die zweitgrößte ist]. Dann diese, dann diese [zeigt nacheinander auf die Werte für Männer in Deutschland und auf Berliner]. So, und genau diesen Zusammenhang wollen wir jetzt mal intensiver untersuchen [schaltet eine Bildschirmprojektion an mit einem Excel-Arbeitsblatt; im Anschluss wird der Zusammenhang funktional veranschaulicht; vgl. Aufgabe 9]

Tab.IV.24: Unterrichtsszene zum Zusammenhang von Basisrate und Ergebnis

Eine Vertiefung dieses Zusammenhangs erfolgte im Anschluss anhand eines Tabellenkalkulationsprogramms, das eine funktionale Veranschaulichung durch dynamische Veränderung der Basisrate im Baumdiagramm ermöglicht. Wenn die Basisrate im

Baumdiagramm verändert wurde, berechnete das Programm entsprechende Auswirkungen im Baum und auf Ergebnisse in einer Tabelle und die Veränderungen wurden als neues Wertepaar in einem Koordinatensystem (x-Achse: Basisrate; y-Achse: Wahrscheinlichkeit für HIV-Infektion bei positivem Test) festgehalten (vgl. Abb.IV.35). Die Arbeitsweise des Programms schien für die meisten Schüler nachvollziehbar, für eine nutzbringende Erarbeitung müsste aber mehr Zeit aufgewendet werden.




Abb.IV.35 und 36: Arbeiten mit einem Excel-Arbeitsblatt zum Zusammenhang von Basisrate und Ergebnis

Mit der gewonnenen Darstellung konnten nun Zusammenhänge beschrieben werden: Es wurde bestätigt, dass nach schrittweiser Erhöhung der Basisraten auch die A-posteriori-Wahrscheinlichkeit monoton ansteigt. Auch sehr gut erkannt wurde, dass dieser Zusammenhang unterproportional ist. Einzelne Wertepaare, die vorher in Gruppenarbeit errechnet wurden, wurden überprüft und die Richtigkeit bestätigt (vgl. Abb.IV.36).

Anschließend wurde zu klären versucht, warum dieser Zusammenhang unterproportional ist. Diese detailliertere Untersuchung des funktionalen Zusammenhangs war deutlich anspruchsvoller, die offensichtlich viele Schüler der 9. Jahrgangsstufe überfordert hat. Ein Grund könnte sein, dass die Schüler bisher nur proportionale Zusammenhänge kennen gelernt hatten.

3.5.5 Vergleich zwischen Darstellungsformen und begriffliche Vertiefungen

Eine Hauptfrage dieser Arbeit war, welchen Nutzen Schüler aus Darstellungen für die Modellierung allgemeiner Bayesprobleme ziehen können. Nachdem die Schüler mehrfach Häufigkeitsbaum und Häufigkeitstabelle im Zusammenhang mit realistischen Diagnoseproblemen erstellen mussten, diente Arbeitsblatt 6 zur „realitätsärmeren“ Übung. Aufgabe 2 (vgl. Schmid & Weidig, 1996, S.188) wurde im Unterricht benutzt, die Vertrautheit der Schüler mit beiden Darstellungsformen festzustellen und zu vertiefen. 2 Schüler erstellten parallel an der Tafel eine Baum- und eine Tabellenmodellierung für diese Aufgabe. Im Klassenverbund wurden beide Modellierungen detailliert verglichen und diskutiert (vgl. Tab.IV.25 und Abb.IV.37).

Unterrichtsstunde 9 Materialgrundlage: Arbeitsblatt 6, Aufgabe 2a Thema: Übungen zum Rückblick 1	
Zeit	Unterrichtsdialog (Abkürzungen: L: Lehrer, S: Schüler/in, LE: Lehrerecho, AB: Arbeitsblatt, ...)
2:30	[2 Schüler modellieren parallel nebeneinander an der Tafel Aufgabe 2a, AB 6 mit Baumdiagramm bzw. Tabelle] 
7:19	L: Wie kann man jetzt, wo die Darstellungen so schön nebeneinander stehen, überprüfen, ob es richtig ist? S: Am Ende beim Baum müssen auch wieder 10 Mio. rauskommen L: Ja, da müssen 10 Mio. rauskommen, das ist ja schon mal gut. S: Und in den Zeilen [meint Ebenen im Baumdiagramm] müssen auch jeweils 10 Mio. rauskommen L: Da meinst du, ja? [zeigt die Bauebenen nacheinander] L: Und wie kann man das jetzt anhand von diesen beiden Darstellungen verbinden und überprüfen? S: In der Tabelle, einmal nach rechts ergibt die Gesamtzahl und nach unten, und auch alle 4 Felder in der Mitte. L: Also, wo find ich alle, die Xelo haben, in der Tabelle. S: Bei gesamt – Xelo plus. [L. zeigt auf diesen Wert] Ja. L: Und die restlichen 990 000 bei S [fährt fort]: Gesamt – Xelo minus. [L. zeigt auf diesen Wert] Ja. L: Diese 9600 [L. zeigt auf die Häufigkeit im Baum für Xelo-erkrankt und positiver Test]? S: Also, die haben Xelo und bei denen ist auch der Test positiv.
8:57	L: [zeigt in der Tabelle erst die Spalte für Xelo, dann die Zeile für Test positiv und dann auf den Wert]. Xelo plus und Test plus. Hier. [entsprechend wird weiter bei den anderen Häufigkeitswerten verfahren. Ein Fehler in der Tabelle an der Tafel wird entdeckt und verbessert]

Tab.IV.25: Unterrichtsszene zum Vergleich von Baum- und Tabellenstrukturen

Im Folgenden benannten dann Schüler (gemäß Aufgabe 2b und c, Arbeitsblatt 6) nach freier Wahl verschiedene Wahrscheinlichkeiten aus den beiden Darstellungen. Die Inhalte des Rückblicks 1, in dem die drei verschiedenen Wahrscheinlichkeitsbegriffe „Einzelwahrscheinlichkeit“, „UND-Wahrscheinlichkeit“ und „Bedingte Wahrscheinlichkeit“ eingeführt und unterschieden wurden, wurden nun wiederholt und vertieft. Durch Zuruf wurden entsprechende Wahrscheinlichkeiten bezeichnet (auch in Kurzschreibweise) und aus dem Baum bzw. der Tabelle berechnet (Angabe der Häufigkeiten und Berechnung). Der Lehrer strukturierte die Ergebnisse jeweils an der Tafel. Vor allem auf die genaue Unterscheidung von bedingter und konjunktiver Wahrscheinlichkeit wurde penibel Wert gelegt. Die Schüler wurden sichtlich sicherer im Umgang mit Begriffen und der Berechnung der Wahrscheinlichkeiten mit beiden Darstellungsformen.

3.5.6 Denkweise des „Lernen aus Erfahrung“

Das Ziel, Bayesianisches Denken als Neubewerten von Hypothesen bei Vorliegen von Indizien („Lernen aus Erfahrung“) zu verankern, sollte über einen neuen realistischen Kontext „Gerichtsurteil in einem Mordfall“ (vgl. Arbeitsblatt 7, Anhang B) erarbeitet werden. Von allen Schülern sollten, wie beim AIDS-Problem, Schätzungen der gesuchten

Wahrscheinlichkeit abgegeben werden. Die Problemstellung war offen formuliert: „Ist die Überlegung des Sachverständigen richtig?“. Bereits ohne zu rechnen wurden sehr detaillierte Grundüberlegungen angestellt und eigene Wahrscheinlichkeitsschätzungen von Schülern angegeben (vgl. Tab.IV.26a, 4:20 und 6:45). Die Meinung des Sachverständigen wurde durch vernünftige Argumentationen verworfen. Dennoch blieben Meinungsunterschiede bestehen, die im Laufe der Stunde zur Entstehung argumentativer Unterrichtssituationen führten.

Unterrichtsstunde 10 <i>Materialgrundlage: Arbeitsblatt 7</i> <i>Thema: Mordfall</i>	
Zeit	Unterrichtsdialog <i>(Abkürzungen: L: Lehrer, S: Schüler/in, LE: Lehrerecho, AB: Arbeitsblatt)</i>
	[Der Text zum Mordfall wurde als Hausaufgabe durchgelesen und wird nun von einem Schüler zusammengefasst. Es wird kurz offen diskutiert. Aufgabe 1 des AB 7 wird laut vorgelesen]
3:40	L: Ist die Überlegung des Sachverständigen richtig?
	S: Also, die 96% auf keinen Fall. Ja, weil die 96% beziehen sich auf die Menschen, die nicht die Blutgruppe wie der Schornsteinfeger haben.
	L: Und warum ist das dann nicht richtig?
	S [fährt fort]: Ja, das heißt ja nicht, dass er der Täter sein muss. Das ist ja was ganz anderes.
	L: Hm.
4:20	S [anderer]: Ja, ich habe die Vermutung gemacht, dass in Wuppertal 100000 Menschen leben. Dann hab ich ausgerechnet, dass 4000 Menschen dem Täter entsprechen [meint hinsichtlich der Blutgruppe]. Das sind ja viel zu viele. Und der Beweis, dass er manchmal durch das Waldstück ging, ist ja auch nicht so toll. Deshalb habe ich dann gesagt, dass die Wahrscheinlichkeit, dass der Schornsteinfeger der gesuchte Täter ist, sehr gering ist.
	L: Ja, was heißt sehr gering? Was sagt denn der Sachverständige?
	S [fährt fort]: Ja, auf jeden Fall viel weniger als 96%.
	L: Du sagtest 4000 mögliche Täter. Was wäre dann die Wahrscheinlichkeit, dass der Schornsteinfeger der Täter ist?
	S [fährt fort]: 1 durch 4000.
	L: Ja, ein Viertausendstel. Das ist ja deutlich weniger als 96%, oder?
	L: O.k., ihr solltet einen Wert schätzen! Was habt ihr geschätzt?
	[fast alle Schüler sagen ihren Schätzwert: 80%, 70%, 60%, 4%, 50%, 20%, 20%, 20%, 85%]
6:13	S: Ich meine, dass die Wahrscheinlichkeit viel geringer ist, weil was die da genommen haben, das kann ja gar nicht sein. Deshalb habe ich den Wert so auf 1 bis 0,5% geschätzt.
6:26	S [anderer]: Kann das sein, dass es darauf ankommt, wie viele Menschen es überhaupt gibt. Wenn es nur 100 Menschen gäbe [meint, die als Täter in Frage kommen], dann hätten ja nur 4 diese Übereinstimmung der Blutgruppe, aber es gibt doch immer nur einen Täter. Dann wäre das jetzt in diesem Fall eine Wahrscheinlichkeit von ein Viertel. Vorhin war das bei 100000 ein Viertausendstel. Wenn man ein paar Millionen Menschen hätte, wärs noch weniger.
6:45	L: Nicht schlecht, hört sich gut an, oder? [Schüler diskutieren kurz untereinander] Dann wollen wir den Gedanken von S. mal im Hinterkopf behalten, da werden wir dann noch dazu kommen. Aufgabe 2 habt ihr ja gelöst, schauen wir uns das mal an!

Tab.IV.26a: Unterrichtsszene zum Konzept „Lernen aus Erfahrung“ mit einer offenen Problemstellung .


Zunächst wurde gemäß Aufgabe 2 auch eine rechnerische Modellierung des Falles erstellt (10000 wird als Grundmenge vorgegeben) und eine Wahrscheinlichkeit für die Täterschaft, wenn Blutgruppeneigenschaft erfüllt ist, ermittelt (9:00 bis 15:45; vgl. Tab.IV.26b und Abb.IV.38). Gründe für ein geäußerte richtige „Hypothese“ eines Schülers, dass ein Zusammenhang mit der angenommenen Grundmenge möglicher Täter besteht (6:26), konnten bei Bearbeitung der Aufgabe 3 nun herausgearbeitet werden (16:21 bis 20:50).

Fortsetzung Unterrichtsstunde 10	
7:23	L [liest Aufgabentext vor]: Wer hat Baumdiagramm gemacht? [Fast alle melden sich] Wer hat Tabelle gemacht? [Ein Schüler meldet sich]
	S: Ich habe ne Frage, haben die 10000 jetzt schon diese Blutgruppe?
	L: Was sagt ihr?
	S [anderer]: Also nein, das sind einfach nur mögliche Täter.
	L: Also, über die 10000 weiß man noch gar nix, außer vielleicht, dass sie im Wald öfter rumlungern, oder so. Oder vielleicht welche, die früher schon mal eine Straftat begangen haben und die man dann auch wieder für verdächtig hält. Mögliche Täter heißt also, man weiß nichts über sie, außer dass sie das Potential hätten, der Mörder zu sein. Was habt ihr nun gemacht?

9:00	[Schülerin zeichnet ein Baumdiagramm an die Tafel zu Aufg.2, AB 7]
12:12	L: O.k., können wir das Baumdiagramm jetzt noch weiterführen?
	[Schüler ergänzt das Baumdiagramm weiter nach unten]
	Abb.IV.38
	L: O.k., Gib nun eine Wahrscheinlichkeit dafür an, dass der Schornsteinfeger der Täter ist.
	S: 0,25%
	L: Wie bist du auf 0,25% gekommen?
	S [fährt fort]: Die 400 mit der Blutgruppe x sind ja praktisch die Verdächtigen und einer ist der Täter, 1 durch 400 = 0,25%
	L: Lasst uns das nochmal vernünftig hinschreiben. Gesucht ist die Wahrscheinlichkeit, dass der Schornsteinfeger tatsächlich der Täter ist, wenn die Blutgruppe x vorliegt oder kurz?
	S: P (Täter Blutgruppe x)
	L: Und das ist?
	S [fährt fort]: 1 durch 400
15:45	L: Also, die Wahrscheinlichkeit, dass einer mit dieser Blutgruppe – wie der Schornsteinfeger - der Täter ist, beträgt bei 10000 Menschen, die in Frage kommen, gerade 0,25%. Ist das klar?
16:21	L: Aufgabe 3. [liest Text vor]
	S: Die Wahrscheinlichkeit steigt.
	L: Warum? Was ändert sich an diesem Bruch (zeigt auf 1 / 400), wenn es weniger mögliche Täter gibt.
	S [fährt fort]: Der Nenner.
	L: Die 1 bleibt auf jeden Fall gleich, warum?
	S: Weil es nur einen Täter gibt.
	L: Allerdings werden das hier weniger [zeigt auf die 10000] und dadurch werden automatisch auch die 400 weniger. Also 1 durch eine kleinere Zahl ergibt dann entsprechend eine...?
	S: größere Zahl.
20:50	L: Gut. Antwort: Die Wahrscheinlichkeit steigt, da der Zähler des Bruches P(Täter Blutgruppe x) konstant bleibt und der Nenner kleiner wird. [schreibt es an die Tafel]

Tab.IV.26b: Unterrichtsszene zur rechnerischen Modellierung der offenen Problemstellung „Mordfall“ (AB 7).

Im folgenden Verlauf wurde der hypothetische Charakter („Annahme“) der A-priori-Wahrscheinlichkeit erarbeitet (Tab.IV.26c). Der Text „Indizien vor Gericht“ führte zur Diskussion über eine vernünftige Abschätzung der Grundmenge der möglichen Täter, denn – wie bereits erkannt wurde – hängt die gesuchte Wahrscheinlichkeit davon ab. Nahezu eigenständig nannten die Schüler vernünftige Grenzen und begründeten diese (23:00 bis 25:50). Die Klasse einigte sich nach kritischer Diskussion auf die Grenzen 100 bis 100000 (andere wären auch vernünftig begründbar gewesen) und die Arbeitsaufträge 4b-d wurden nun in Partnerarbeit erledigt (25:50 bis 35:07). Anschließend wurden die Ergebnisse an der Tafel zusammengetragen und bestätigten nochmal die Abhängigkeit der Wahrscheinlichkeit für die Täterschaft des Angeklagten von der hypothetischen Annahme der Grundmenge potentieller Täter. Bei der Betrachtung der jeweiligen Informationsgewinne entdeckten Schüler selbst, dass das Blutgruppenindiz jeweils unabhängig von der Grundmenge die Wahrscheinlichkeit für die Täterschaft um das 25-fache erhöht (ab 37:14, Abb.IV.39).

Fortsetzung Unterrichtsstunde 10	
21:00	[Text „Indizien vor Gericht“ wird gelesen. Über den Inhalt wird kurz diskutiert]
23:12	L: Aufgabe 4a. Überlege ein vernünftige obere und untere Grenze für die Grundmenge der möglichen Täter mit Begründung. S: Ja, das sind alle die dafür in Frage kommen und die im Wald gesichtet worden sind. L: Wie viele sollte man mindestens berücksichtigen? S [fährt fort]: Ich würde sagen 10. S [anderer]: 10 ist schon ganz schön wenig, oder? [Viele Schüler stimmen zu] S: mindestens 100. L: Warum meinst du mindestens 100? S [fährt fort]: Ich denke, ein paar mehr als 10. 100 wurden schon mindestens im Wald gesehen. L: 100 ist gut, denke ich. Und höchstens. Wir sind in einer Kleinstadt. Nehmen wir mal an, es wäre in Bünde passiert. Wie viel Einwohner hat z.B. Bünde mit Randgebieten. S: 40000? S [anderer]: 46000 [Kurze scherzhafte Diskussion, welche Gebiete zu Bünde gehören] S [anderer]: Wuppertal ist aber größer. L: Dann sagen wir mal höchstens 100000, o.k? L: Also gut, unter dieser Voraussetzung minimal 100, maximal 100000 löst ihr nun bitte Aufgaben 4 b – d.
25:50	[Bearbeitung von 4b-d in Partnerarbeit. Die Schüler können mit den Partnern diskutieren]
35:07	L: So wir haben 100 möglich Täter, wie groß ist die Wahrscheinlichkeit, bevor die Blutgruppe bestimmt wurde, dass diese Person der Täter ist? [zeichnet eine tabellarische Übersicht an die Tafel mit Spalten „mögliche Täter“, „A-priori“ und „A-posteriori-Wahrscheinlichkeit“] S: Ein Hundertstel. L: Ja, a-priori einer von 100 [trägt das jeweils in der Tabelle ein]. Das sind in Prozent? S [fährt fort]: 1% L: Und bei 100000 möglichen Tätern S: 0,001%, das ist 1 durch 100000 [L trägt es in der Tabelle ein] L: Und a-posteriori? S: 25% L: Wie kommst du auf die 25%? S [fährt fort]: Die 100 mal 4%, das sind dann 4, also 4, die diese Blutgruppe auch haben und dann muss es ja einer von den 4 sein. Also 1 durch 4. L: Richtig, 1 durch 4 ist 25%. [L trägt es in der Tabelle ein] L: Und wie viel haben wir bei 100000 S: 1 durch 4000 sind 0,025%. [L trägt es in der Tabelle ein] L: O.K. Und jetzt d) Überlege, welchen Informationsgewinn das Indiz „Übereinstimmung der Blutgruppe“ jeweils bringt. S: Jeweils das 25fache.
37:14	[L. zeichnet farbige Pfeile mit Aufschrift „· 25“]  Abb.IV. L: Ja, interessanterweise beide Male mal 25. Was heißt das jetzt? S: Egal wie groß die Menge der möglichen Täter ist, das Indiz ist immer gleich. L: Ja, nicht das Indiz. S [fährt fort]: Der Informationsgewinn. L: Ja, es steigert die Wahrscheinlichkeit um das 25fache.

Tab.IV.26: Unterrichtsszene zum Konzept „Lernen aus Erfahrung“ bei verschiedenen A-priori-Annahmen.

Insgesamt wird aus meiner Sicht deutlich, dass in dem dargestellten Unterrichtbeispiel Bayesianisches Denken im Sinne eines „Neubewertens von Hypothesen“ gut verankert werden konnte. Schüler konnten argumentativ Hypothesen bewerten und eigenständig überprüfen, wie sich die Wahrscheinlichkeiten für diese Hypothesen durch das Vorliegen von Indizien verändern. Insbesondere konnte selbstständig erkannt werden, dass die Wahrscheinlichkeit des Auftretens des Indizes „Blutgruppe“ nicht zur Beurteilung ausreicht, sondern jeweils A-priori-Annahmen über die Wahrscheinlichkeit einer Täterschaft nötig sind. Zu erkennen war, dass der Hypothesenbegriff im Zusammenhang mit Wahrscheinlichkeiten in diesem Kontext recht intuitiv verwendet werden kann. Die Verbindung der Begriffe Hypothese und Wahrscheinlichkeit, die einer Weitervertiefung im Rahmen der beurteilenden Statistik in der Sekundarstufe II bedürfen, geschieht hier sehr natürlich.

4 Zusammenfassung und Ausblick

Im untersuchten Modellunterricht auf Basis eines realitätsbezogenen „Häufigkeitskonzeptes“ haben Schülerinnen und Schüler der Sekundarstufe I insgesamt erfolgreich gelernt, Bayesianische Situationen vernünftig zu beurteilen, ohne dass die Regel von Bayes formal benutzt wurde. Das Ziel, innerhalb eines spiralförmigen Stochastik-Curriculums einen vorläufigen und tragfähigen Abschluss im Themenkreis „Bedingte Wahrscheinlichkeit und Bayesianwendungen“ in der Sekundarstufe I zu schaffen, wurde nach Ansicht aller Beteiligten erreicht. Ich schlage deshalb vor, das Häufigkeitskonzept in der Form einer realitätsbezogenen Unterrichtsreihe als Basis für den ersten Zugang in der Sekundarstufe I zum Themengebiet „Bedingte Wahrscheinlichkeit und Anwendungen des Satzes von Bayes“ zu verwenden. Es ist anzunehmen, dass mit dem häufigkeitsbasierten Einstieg Bayesianisches Denken in entsprechender Form sogar mit noch jüngeren Schülern (frühe Sekundarstufe I oder sogar Primarstufe) erarbeitet werden kann. Erste Ergebnisse mit Zehn- bis Zwölfjährigen im Labor (Zhu & Gigerenzer, eingereicht) legen diese Vermutung nahe.

Auf welche Erkenntnisse aus dieser Unterrichtsstudie kann dieser didaktische Vorschlag gestützt werden? Ich möchte zusammenfassend diese Punkte herausstellen:

1. Bayesianisches Denken und damit zusammenhängende Themen stellen einen sehr *geeigneten Inhalt für den Stochastikunterricht der Sekundarstufe I* dar. Sowohl Leistungsergebnisse, als auch Schüler- und Lehrerbewertungen lassen diesen Schluss zu. Gemessen an den meisten deutschen Lehrplänen schien das bisher keine allgemeine Überzeugung zu sein.
2. Der realitätsbezogene Zugang über natürliche Häufigkeiten ermöglicht einen *hohen und noch nach Monaten stabilen Lernerfolg*, der mit dem entwickelten Kompetenztest zum Bayesianischen Denken nachgewiesen werden konnte. Dieser hohe Lernerfolg stellte sich in allen fünf neunten Klassen mit verschiedenen Lehrenden, die unterschiedliche Einstellungen zum erfolgten Unterricht äußerten, ein.
3. Das Ziel der Unterrichtsstudie war kein quasi-experimenteller Performanzvergleich mit anderen Klassen der Sekundarstufe I, die nach herkömmlichen Methoden unterrichtet wurden, sondern die Exploration des häufigkeitsbasierten Zugangs. Einen *Vergleichsmaßstab* bieten untersuchte Leistungskurschüler in Mathematik, die üblichen Unterricht und ein zusätzliches „Regeltraining“ im Labor am Computer erhalten hatten (vgl. III.3.3.2). Nach ca. 3 Monaten lagen die Leistungen im Kompetenztest nur noch etwa bei 1/3 der Leistungen der Schüler, die Modellunterricht mit dem Häufigkeitskonzept erhalten hatten.

4. Bei fast allen Schülern konnten weitreichende Modellierungskompetenzen auch anspruchsvollerer, realitätsbezogener Situationen mit Häufigkeitsdarstellungen verankert werden. Lehrer- und Schülerbewertungen und Unterrichtsszenen unterstrichen deutlich den hohen didaktischen Nutzen von Häufigkeitsmodellierungen. Deutlich sichtbar war, dass anstatt der Einübung von Lösungsroutinen vor allem vernünftiges und argumentatives Denken angeregt wurde.
5. Der Unterricht wurde aufgrund der realistischen Themen von vielen Schülern als interessanter und abwechslungsreicher als sonst empfunden, selbst von Schülern die eine eher negative Einstellung zu mathematischen Inhalten haben.
6. Die sehr häufig vertretene Meinung, dass Themen zur „Anwendung der Bayesregel“ selbst für leistungsstärkere Schüler schwer zu erarbeiten sind, ist durch Leistungsbefunde und Lehrer- bzw. Schülerbewertungen gegenteilig zu beurteilen: Die Inhalte sind auch für Schüler mit geringeren algebraischen Fähigkeiten sehr geeignet und es ist vergleichsweise wenig Vorwissen nötig.

Aus den differenzierteren *Analysen von Modellierungsprozessen* ging hervor, dass auftretende Schwierigkeiten weitgehend auf Strategien beruhten, die auch aus Untersuchungen intuitiven Bayesianischen Denkens (vgl. Kap.I) bekannt sind. Im Sinne einer Korrektur bzw. eines Abbaus fehlerhafter primärer Intuitionen und eines Aufbaus und einer Verankerung adäquater sekundärer Intuitionen (vgl. Fischbein, 1975) unterstützen die Befunde ausdrücklich die sorgfältigere Beachtung von Grundintuitionen beim stochastischen Denken. So wurde z.B. deutlich, dass viele Schwierigkeiten auf fehlerhafte Intuitionen zu Grundeigenschaften der Wahrscheinlichkeit oder Prozentrechnen zurückzuführen waren, die noch nicht adäquat verbessert werden konnten. Defizite in diesen Bereichen müssten generell erst aufgeholt werden, bevor Berechnungen von Wahrscheinlichkeiten erfolgen können. Für eine gezielte Förderung der Schüler kann der Vergleich und die Rekonstruktion „individueller“ Strategien und Bedeutungen im Vergleich zu „institutionellen“ Vorgehensweisen eine entscheidende Grundlage sein. Eine solche Gegenüberstellung persönlicher und institutioneller Bedeutungen mathematischer Objekte und Praktiken sollte nicht nur in der didaktischen Forschung, sondern auch im Unterricht verbreiteter angewandt werden (vgl. Godino & Batanero, 1998).

Mit den Ergebnissen dieser Studie werden auch weiterführende Fragestellungen aufgeworfen, die sich aus meiner Sicht vor allem auf folgende Punkte beziehen.

1. Wie lässt sich dieser Zugang in der Sekundarstufe II zu formalen, wahrscheinlichkeitstheoretischen Konzepten ausbauen, die zu Zielen des gymnasialen Abschlussniveaus gehören? Der Ausbau ist meines Erachtens nicht nur unproblematisch, er wird sogar unterstützt. Aus den Häufigkeitsdarstellungen können völlig konsistent alle nötigen Wahrscheinlichkeitsregeln hergeleitet und begründet werden (vgl. Kap.II.3). Eine Pilotstudie hierzu ergab (vgl. Wassner et al., 2002), dass Schüler ohne Vorwissen nach Training mit dem „Häufigkeitskonzept“ allgemeine Regeln für den Zusammenhang von Wahrscheinlichkeiten¹²³ genau so gut herleiten konnten wie Schüler, die bereits ein „Wahrscheinlichkeitstraining“ absolviert hatten. Diese ersten, noch detaillierter zu untersuchenden Befunde weisen darauf hin, dass ein in der Curriculumspirale folgender formaler Ausbau des Themas „Bedingte Wahrscheinlichkeit und Bayesregel“ durch den „Häufigkeitszugang“ bereits gefördert wird.

¹²³ z.B. einen Zusammenhang zwischen bedingter und konjunktiver Wahrscheinlichkeit $P(A \cap B) = P(A) \cdot P(B | A)$.

2. Kann das Häufigkeitskonzept für weitere Inhaltsbereiche des Stochastikunterrichts gewinnbringend genutzt werden? In Sedlmeier & Köhlers (2001) wurde das „Konzept der natürlichen Häufigkeiten“ konsequent für einen Unterricht in Stochastik umgesetzt. Häufigkeitsbasierte Zugänge ohne wahrscheinlichkeitstheoretische Formalisierung werden zu wichtigen Grundinhalten der Stochastik aufgezeigt (Relative Häufigkeit und statistischer Wahrscheinlichkeitsbegriff, empirisches Gesetz der großen Zahlen, konjunktive Wahrscheinlichkeit, bedingte Wahrscheinlichkeit, Revision von Wahrscheinlichkeit (Bayes), Stichprobenverteilungen, Konfidenzintervalle und Signifikanztest). Das Häufigkeitskonzept wurde auch bereits in weiteren Trainingsstudien erfolgreich eingesetzt, um z.B. den Einfluss der Stichprobengröße auf die Genauigkeit von Wahrscheinlichkeitsschätzungen (Themenbereich „empirisches Gesetz der großen Zahlen“ bzw. „Verteilungen“) zu lehren (vgl. Sedlmeier, 1998, 1999). Die These liegt nahe, dass die wichtigsten Inhaltsbereiche des gymnasialen Stochastikunterrichts auf dem Häufigkeitskonzept basierend unterrichtet werden können und dass resultierende Darbietungsformen ähnliche Bildungswirkungen hervorrufen wie bei einem Unterricht im Bereich „Bayesianisches Denken“. Entsprechende Untersuchungen ähnlich dieser Arbeit wären gegebenenfalls nötig.

KAPITEL V ZUSAMMENFASSENDES FAZIT UND IMPLIKATIONEN

Diese Arbeit setzt sich mit Bayesianischem Denken unter zwei wesentlichen Sichtweisen auseinander: Erstens werden kognitive Bedingungen theoretisch hergeleitet und untersucht, die diese Form des Denkens unter Unsicherheit erleichtern bzw. erschweren können. Zweitens werden diese Ergebnisse dafür benutzt, ein neues Konzept für das Lehren und Lernen Bayesianischen Denkens zu erstellen und zu untersuchen.

Forschungsbefunde zu menschlichem Denken unter Unsicherheit mit dem Fokus „Bayesianisches Denken“ werden zunächst in **Kapitel I** zusammengefasst und analysiert. Das Anliegen ist, Erkenntnisse der Kognitionspsychologie, die Verarbeitungsprozesse der Lernenden beim Bayesianischen Denken betreffen, gezielt für die mathematikdidaktische Forschung zu erschließen. Von besonderem Interesse ist dabei das auf einer speziellen Häufigkeitsmodellierung beruhende „*Konzept der natürlichen Häufigkeiten*“. Es geht aus der kognitionspsychologischen Diskussion und Befunden des Forschungsprogrammes „Ökologische Rationalität“ hervor (Kapitel I.3; z.B. Gigerenzer, 1991). Im Unterschied zu den viel diskutierten Befunden des kognitionspsychologischen „Heuristikansatzes“ (z.B. Kahneman, Slovic & Tversky, 1982), die sich überwiegend auf die Identifizierung grundlegender Fehlvorstellungen beschränken, wird nun die Qualität intuitiver Verarbeitungsmechanismen nach dem Repräsentationsformat differenziert: Tragfähige Primärintuitionen zur prozeduralen Verarbeitung von Wahrscheinlichkeiten bestehen kaum, wohl aber zur Verarbeitung von Häufigkeiten. Entsprechende Studien belegten, dass bereits ein „Anstoßen“ häufigkeitsbasierter Modellierungen (z.B. durch die Art der Aufgabenstellung) den Menschen zu einem deutlich vernünftigeren Umgang mit Bayesianischen Situationen befähigen kann. Theoretische Begründungen basieren auf evolutorischen Argumenten der Adaption. Es wird deutlich, dass Erklärungsansätze des „Heuristikansatzes“ zu eng gefasst sind.

Kapitel II gibt eine curriculare Übersicht und beschäftigt sich detailliert mit didaktischen Konzepten und Realisierungen von Inhalten zu „Bedingter Wahrscheinlichkeit und Bayesanwendungen“. Ein Resümee ist, dass die Inhalte in der Sekundarstufe I bislang zumeist eine nur marginale Rolle spielten, auch wenn einige innovative Vorschläge zu einer Veränderung beitragen. Meines Erachtens unterstützen die Schwierigkeiten der Lernenden mit der Bayesregel, die auch mit den empirischen Ergebnissen dieser Arbeit bestätigt werden, die Anschauung, dass eine *formale und regelbasierte Behandlung* nicht Ziel des Stochastikunterrichts in der Sekundarstufe I sein sollte.

Hinzu kommt, dass vielfach in der schulischen Ausbildung das Themengebiet als Randgebiet – eine Art Anhängsel – an das wahrscheinlichkeitstheoretische Gebäude, gesehen wird. Ich führe in dieser Arbeit mit einigen Beispielen aus (z.B. Einführung), warum diese Bedeutungsbewertung *nicht* der praktischen und theoretischen Relevanz des Bayesianischen Denkens gerecht wird. Gründe für diese Unterbewertung liegen auch in einer Institutionalisierung des klassischen Zugangs in der statistischen Ausbildung. Sowohl in der didaktischen, als auch in der psychologischen Diskussion wird auf diese unbegründete Vernachlässigung hingewiesen und Vorschläge zur Abhilfe werden aufgezeigt. Meiner Meinung nach ist *Bayesianisches Denken* beim „alltäglichen“ Umgang mit Unsicherheit sogar so relevant, dass es *als zentrales Konzept* so früh wie möglich, auf jeden Fall aber in der Sekundarstufe I eingeführt werden sollte.

In Kapitel II.3 wird deshalb ein *Zugang mit „natürlichen Häufigkeiten“* erläutert, der die Herleitung und Verwendung der schwierigen Bayesregel umgeht und trotzdem die mathematische Modellierung realitätsbezogener Probleme im Zusammenhang mit „Bayesianischem Denken“ ermöglicht. Der Zugang dient schließlich als Grundlage für die Ausarbeitung und Untersuchung einer Unterrichtsreihe in der Sekundarstufe I, mit der Inhalte des Themengebiets mit nachhaltigem Lernerfolg vermittelbar sind (Kapitel IV).

Bayesianisches Denken zu trainieren und den Erfolg systematisch zu testen, wurde vor allem durch *Trainingsstudien* von Sedlmeier und Gigerenzer (2001) auf Grundlage ihrer Befunde im Programm „Ökologische Rationalität“ forciert. In **Kapitel III** werden die Trainingsstudien genauer dargestellt und diskutiert.

Meine eigenen Trainingsuntersuchungen erfolgten auf Grundlage des computerbasierten Trainingsprogramms von Sedlmeier und Gigerenzer mit Leistungskursschülern und Zehntklässlern der gymnasialen Sekundarstufe I. Ein entwickeltes Testinstrument zum Bayesianischen Denken belegt den hohen Nutzen häufigkeitsbasierten Trainings gegenüber auf Wahrscheinlichkeitsformaten basierendem Training. Der didaktisch entscheidende Punkt ist die *Nachhaltigkeit* des Lernerfolges. Nach Training mit dem „Häufigkeitskonzept“ ist der Lernerfolg deutlich stabiler im Gegensatz zu Wahrscheinlichkeitstraining, nach dem ein weitgehender Verfall des Lernerfolgs einsetzt.

Darüber hinaus verdeutlichen meine Studien mit Schülerinnen und Schülern der Sekundarstufe I und II, dass dieser höhere Lernerfolg nicht nur auf rein rechnerisches Modellieren (*prozedurales Wissen*) von dichotomen Bayesianischen Problemen beschränkt ist. Die Unterschiedseffekte beziehen sich auch auf Fähigkeiten zur Modellierung von Aufgaben mit komplexerer Informationsstruktur und auf begriffliche Modellierungsaufgaben (*konzeptuelles Wissen*).

Die Analysen der Modellierungsprozesse zeigen, dass vor allem der Schritt der Mathematisierung, d.h. der „Übersetzung“ der realen Situation in ein Modell durch das Häufigkeitskonzept besonders unterstützt wird. Die Befunde der Prozessanalysen wurden im folgenden Entwurf einer Unterrichtsreihe für die Sekundarstufe I verwendet.

Die im Rahmen dieser Arbeit entworfene *Unterrichtsreihe „Authentisches Bewerten und Urteilen unter Unsicherheit“* (**Kapitel IV.1 bzw. Anhang B**) basiert auf folgenden Grundlagen:

Zum einen erfolgt eine didaktische Umsetzung des „Konzeptes der natürlichen Häufigkeiten“ und dazu gewonnener empirischer Ergebnisse für eine Behandlung des Themengebiets „Bedingte Wahrscheinlichkeit und Bayesanwendungen“ in der Sekundarstufe I. Diese „kognitive Komponente“ stellt den innovativen Kern dar.

Zum anderen wird ein Zugang mit hohem Realitätsbezug verwirklicht, in dem weitere „allgemeinere Bildungsaspekte“ wie Lebensvorbereitung, Problemlösefähigkeit, kritischer Vernunftgebrauch bei Modellierung und Interpretation, Sinnstiftung, motivationale Faktoren etc. Beachtung finden. An verschiedenen Stellen der Arbeit finden sich Beiträge zur Diskussion dieser Thematik in der Mathematikdidaktik. Die besondere praktische Relevanz des Bayesianischen Denkens fordert meines Erachtens den realitätsbezogenen Zugang.

Eine Implementation dieser Unterrichtsreihe in der Sekundarstufe I wird schließlich detailliert untersucht. Ich habe dabei eine *explorative Unterrichtsstudie* einer quasi-experimentellen Vergleichsuntersuchung vorgezogen. Das Ziel im Rahmen der Möglichkeiten dieser Arbeit besteht darin, eine Exploration und Optimierung des Unterrichts mit dem „Häufigkeitskonzept“ auf möglichst breiter Basis zu erreichen.

Im Rahmen der Unterrichtsstudie wurden fünf 9. Gymnasialklassen auf Grundlage des entworfenen Materials etwa 15 Schulstunden unterrichtet. In **Kapitel IV.2** ist das genaue methodische Vorgehen beschrieben und die Resultate der Erhebungen sind in IV.3 dargestellt und analysiert. Leistungsbezogene Ergebnisse zeigen die schon in Trainingsstudien festgestellte hohe Nachhaltigkeit von Lernerfolgen. Die angestrebten Kompetenzerwartungen können von den meisten Schülern auch 14 Wochen nach dem Unterricht noch erreicht werden. Die Lehrer- und Schüleraussagen bei den durchgeführten Befragungen machen deutlich, dass der hohe Lernerfolg maßgeblich auf die Verwendung des „Häufigkeitskonzeptes“ zurückzuführen ist. Beobachtungsergebnisse stützen diesen Befund.

Der Umgang mit realistischen Situationen führt an vielen Stellen des Unterrichts zur Förderung weiterer allgemeinerer „Bildungsaspekte“ (vgl. oben). Vor allem die Schüleraussagen belegen positive motivationale Wirkungen. In vielen Unterrichtsphasen (vgl. Kap.IV.3.4) werden Gelegenheiten zur Interpretation und Diskussion auch weiterer, spontaner Fragestellungen genutzt. An anderen Punkten ergeben sich auch Schwierigkeiten durch realistische Problemstellungen, die oft mit der höheren Komplexität der Erarbeitung zusammenhängen.

Das Gesamtergebnis meiner Arbeit bestätigt den hohen *Nutzen des häufigkeitsbasierten Unterrichtskonzeptes* für die *Schaffung eines vorläufigen und tragfähigen Abschlusses zu Themen um „Bedingte Wahrscheinlichkeit und Bayesianwendungen“ in der Sekundarstufe I*. Der Erfolg besteht offensichtlich in einer *Förderung grundlegender Intuitionen* und Vorstellungen des Menschen in diesem Bereich, die zweifellos häufigkeitsbasiert sind. Als Konsequenz sollte das didaktische „Häufigkeitskonzept“ als Zugang zur *Einführung „Bayesianischen Denkens“ in der Schule* verwendet werden (vgl. Wassner, Martignon & Biehler, 2004).

Eine frühe Verankerung und Zentrierung Bayesianischen Denkens im schulischen Stochastikunterricht bietet meines Erachtens neben der Überwindung des klassischen Verfahrensdeterminismus besondere didaktische Möglichkeiten: Es kann in ausgezeichneter Weise dafür geeignet sein, eine „Kultur“ im Mathematikunterricht zu schaffen, die Raum für kritische Reflexion und sinnvolle Auseinandersetzung mit stochastischen Problemen und allen dazugehörigen Aspekten des Lebens ermöglicht. Das „Häufigkeitskonzept“ bietet dafür eine Basis, die für nahezu alle Schüler zu erreichen ist.

LITERATURVERZEICHNIS

- [AK Stochastik] Arbeitskreis Stochastik in der Schule der Gesellschaft für Didaktik der Mathematik (2002). Empfehlungen zu Zielen und zur Gestaltung des Stochastikunterrichts. *Mitteilungen der Gesellschaft für Didaktik der Mathematik* 75, 75-83.
- Arbuthnot, J. (1710). An Argument for Divine Providence, taken from the Constant Regularity observ'd in the Births of Both Sexes. *Philosophical Transactions of the Royal Society* 27. 186-190.
- Althoff, H. (1985). *Wahrscheinlichkeitsrechnung und Statistik*. Stuttgart: Metzlersche Verlagsbuchhandlung.
- Athen, H., Griesel, H. & Postel, H. (1984, Hrsg.). *Mathematik heute: Leistungskurs Stochastik*. Hannover: Schroedel.
- Athen, H., Griesel, H. & Postel, H. (1985, Hrsg.). *Mathematik heute: Gymnasium 8.Schuljahr*. Hannover: Schroedel.
- Bar-Hillel, M. (1980). The Base-Rate Fallacy in Probability Judgments. *Acta Psychologica* 44, 211-233.
- Bar-Hillel, M. (1983). The Base-Rate Fallacy Controversy. In: Scholz, R.W. (Ed.). *Decision Making under Uncertainty*. Amsterdam: North Holland, 39-62.
- Barnett, V. (1999). *Comparative Statistical Inference*, 3rd ed. New York: Wiley.
- Barth, F. & Haller, R. (1983). *Stochastik Leistungskurs*. München: Ehrenwirth.
- Baumert, J., Gigerenzer, G. & Martignon, L. (2004). *Stochastisches Denken (Themenheft)*. *Unterrichtswissenschaft* 32 (1)
- Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr.Price, in a letter to John Canton, M.A. and F.R.S. *Philosophical Transactions of the Royal Society of London* 53. 370-418.
- Beck-Bornholdt, H.P. & Dubben, H.H. (2001). *Der Schein der Weisen – Irrtümer und Fehltritte im täglichen Denken*. Hamburg: Hoffmann & Campe.
- Bea, W. (1995). *Stochastisches Denken – Analysen aus kognitionspsychologischer und didaktischer Perspektive*. In: Crott, H.W. & Scholz, R.W. (Hrsg.). *Psychologie des Entscheidungsverhaltens und des Konflikts*, Bd. 6. Frankfurt a.M.: P. Lang.
- Behrens, R. et al. (1986). *Mathematik am Gymnasium: 8. Schuljahr*. Frankfurt a.M.: Diesterweg.
- Bentz, H.J. & Borovcnik, M. (1985). Probleme bei empirischen Untersuchungen zu Wahrscheinlichkeitsbegriff. *Journal für Mathematik-Didaktik* 6 (4), 241-264.
- Berger, J.O. (2003). Could Fisher, Jeffreys and Neyman have agreed on Testing? (with comments). *Statistical Science* 18 (1), 1-32.
- Betsch, T. & Pohl, D. (2002). Tversky and Kahneman's Availability Approach to Frequency Judgments: A critical analysis. In: Sedlmeier, P. & Betsch, T. (2002). *Etc. – Frequency Processing and Cognition*. Oxford: University Press, 109-119.
- Biehler, R. (1982). *Explorative Datenanalyse – eine Untersuchung aus der Perspektive einer deskriptiv-empirischen Wissenschaftstheorie*. IDM-Reihe Materialien und Studien, Band 24. Bielefeld: Universität Bielefeld.
- Biehler, R. (1985). Graphische Darstellungen. *Mathematica Didactica* 8, 57-81.
- Biehler, R. (1991). Computers in Probability Education. In: Kapadia, R. & Borovcnik, M. (Eds.). *Chance Encounters: Probability in Education*. Dordrecht: Kluwer, 169-212.
- Biehler, R. (1995). *Explorative Datenanalyse*. *Computer und Unterricht* 17 (Themenheft).

- Biehler, R., Heymann, H.W. & Winkelmann (1995, Hrsg.). *Mathematik allgemeinbildend unterrichten: Impulse für Lehrerbildung und Schule. Untersuchungen zum Mathematikunterricht 21 (IDM-Reihe)*. Köln: Aulis.
- Bigalke, A. & Köhler, N. (1997, Hrsg.). *Mathematik 13.2, Grund- und Leistungskurse*. Berlin: Cornelsen.
- Birkner-Ahsbahs, A., Jungwirth, H. & Kaiser, G. (2003). *Qualitative empirische Methoden in der Mathematikdidaktik*. In: Henn, H.W. (Hrsg.). *Beiträge zum Mathematikunterricht 2003*. Hildesheim: Franzbecker, 121-128.
- Birnbaum, I. (1982). *Die Interpretation statistischer Signifikanz (übersetzt von Günter Fillbrunn)*, *Stochastik in der Schule 2 (3)*, 42-45.
- Blum, W. (1991). *Applications and Modelling in Mathematics Teaching – A Review of Arguments and Instructional Aspects*. In: Niss, M., Blum, W. & Huntley, I. (Eds.). *Teaching of Mathematical Modelling and Applications*. Chichester: Horwood, 10-29.
- Blum, W. (1996). *Anwendungsbezüge im Mathematikunterricht – Trends und Perspektiven*. In: Kadunz, G. et al. (Hrsg.). *Trends und Perspektiven – Beiträge zum 7. internationalen Symposium zur Didaktik der Mathematik, Klagenfurt*. Wien: Hölder-Pichler-Tempsky, 15-38.
- Blum, W., König, G. & Schwehr, S. (1997, Hrsg.). *Materialien für einen realitätsbezogenen Mathematikunterricht, ISTRON Schriftenreihe, Band 4*. Hildesheim: Franzbecker.
- Böer, H. (1997). *AIDS – Was ist von einem positiven Test-Ergebnis zu halten?* In: Blum, W., König, G. & Schwehr, S. (Hrsg.). *Materialien für einen realitätsbezogenen Mathematikunterricht, Band 4*. Hildesheim: Franzbecker, 38-57.
- Borgida, E., & Brekke, N. (1982). *The base-rate fallacy in attribution and prediction*. In J. H. Harvey, W. J. Ickes & R. F. Kidd (Eds.). *New directions in attribution research*. Hillsdale, NJ: Lawrence Erlbaum, 66-95.
- Borromeo Ferri, R. (2003). *Mathematische Denkstile – visuell, analytisch, konzeptuell und ihre Präferenzen bei Jugendlichen am Ende der Sekundarstufe*. In: Henn, H.W. (Hrsg.). *Beiträge zum Mathematikunterricht 2003*. Hildesheim: Franzbecker. 141-144.
- Borovcnik, M. (1984). *Der Problemkreis Bayessche Formel*. *Mathematica didactica 7*, 207-224.
- Borovcnik, M. (1985). *Begünstigen – Eine stochastische Intuition*. *Praxis der Mathematik 27*, 327-333.
- Borovcnik, M. (1990). *Ein intuitiver Zugang zur bedingten Wahrscheinlichkeit und zur Bayes-Formel*. *Stochastik in der Schule 10 (3)*, 22-35.
- Borovcnik, M. (1992). *Stochastik im Wechselspiel von Intuitionen und Mathematik, Lehrbücher und Monographien zur Didaktik der Mathematik, Band 10*. Mannheim: BI-Wissenschaftsverlag.
- Borovcnik, M. & Bentz, H.J. (1991). *Empirical Research in Understanding Probability*. In: Kapadia, R. & Borovcnik, M. (Eds.). *Chance Encounters: Probability in Education*. Dordrecht: Kluwer, 73-106.
- Borovcnik, M., Engel, J. & Wickmann, D. (2001, Hrsg.). *Anregungen zum Stochastikunterricht*. Hildesheim: Franzbecker.
- Box, G.E.P. (1983). *An Apology for Ecumenism in Statistics*. In: Box, G.E.P. et al. (Ed.). *Scientific Inference, Data Analysis, and Robustness*. New York: Academic Press.
- Brunswik, E. (1943). *Organismic achievement and environmental probability*. *Psychological Review 50*, 255-272.
- Brunswik, E. (1955). *Representative design and probabilistic theory in a functional psychology*. *Psychological Review 62*, 193-217.
- Burrill, G. (1996). *Data-driven mathematics: A curriculum strand for high school mathematics*. *The Mathematics Teacher 89*, 460-465.

- Buth, M. (1991). Die Behinderung des gesunden Menschenverstandes durch Stochastik. *Stochastik in der Schule* 11 (3), 12-22.
- Buth, M. (2003). Methodische Anregungen zur Behandlung der bedingten Wahrscheinlichkeit. *Der Mathematisch-naturwissenschaftliche Unterricht* 56 (7), 391-394.
- Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine* 299, 999-1000.
- Chase, M.V., Hertwig, R. & Gigerenzer, G. (1998). Visions of Rationality. *Trends in Cognitive Sciences* 2 (6), 206-214.
- Christensen-Szalanski, J. J. & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior and Human Performance* 29, 270-278.
- Christensen-Szalanski, J.J. & Beach, L.R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist* 39, 75-78.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist* 49, 997-1003.
- Cohors-Fresenborg, E. (1996). Mathematik als Werkzeug zur Wissensrepräsentation. In: Kadunz, G. et al. (Hrsg.). *Trends und Perspektiven – Beiträge zum 7. internationalen Symposium zur Didaktik der Mathematik*, Klagenfurt. Wien: Hölder-Pichler-Tempsky, 85-90.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition* 31, 187-276.
- Cosmides, L. & Tooby, J. (1996). Are Humans Good Intuitive Statisticians after all? Rethinking Some Conclusions from the Literature on Judgment Under Uncertainty. *Cognition* 58, 1-73.
- Cukrowicz, J. & Zimmermann, B. (2001, Hrsg.). *MatheNetz 9*, Ausgabe N, Gymnasium. Braunschweig: Westermann.
- Davies, C.M. (1965). Development of Probability Concept in Children. *Child Development* 36, 779-788.
- Dawes, R.M. (1988). *Rational Choice in an uncertain world*. San Diego: Harcourt Brace Jovanovich.
- Daston, L.J. (1988). *Classical probability in the Enlightenment*. Princeton, NJ: University Press.
- Dougherty, M.R., Gettys, C.F. & Ogden, E. (1999). Minerva-DM: A memory model for judgments of likelihood. *Psychological Review* 106, 180-209.
- Ebbinghaus, (1913). *Memory: A contribution to experimental psychology* (H.A. Ruger & C.E. Bussenes, Transl.). New York: Teachers College, Columbia University. (Orig. Arbeit 1885 veröff.)
- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In: D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: University Press, 249-267.
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.). *Formal representation of human judgment*. New York: Wiley, 17-52.
- Eichler, A. (2003). Was ist Wahrscheinlichkeit? Individuelle Unterrichtskonzepte von Lehrerinnen und Lehrern. *Der Mathematikunterricht* 49 (3), 69-82.
- Einhorn, H.J. & Hogarth, R.M. (1981). Behavioral Decision Theory: Processes of Judgment and Choice. *Annual Review of Psychology*, 53-88.
- Engel, A. (1975). The probabilistic Abacus. *Educational Studies in Mathematics* 6, 1-22.
- Engel, A. (1976). Why does the probabilistic Abacus work?. *Educational Studies in Mathematics* 7, 59-69.

- Engel, J. (2001). Die NCTM-Standards zur Stochastik und das Quantitative Literacy Programm. In M. Borovcnik, J. Engel & D. Wickmann (Hrsg.). *Anregungen zum Stochastikunterricht*. Hildesheim: Franzbecker, 43-52.
- Eulner, M. (1993). *Didaktik der Wahrscheinlichkeitsrechnung – Anspruch und Wirklichkeit*. Diplomarbeit, Universität Karlsruhe, Fakultät f. Wirtschaftswissenschaften.
- Falk, R. (1979). Revision of probabilities and the time axis. *Proceedings of the third international conference for the psychology of mathematics education*. Warwick (GB), 64-66.
- Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning* 9, 83-96.
- Falk, R. (1988): Conditional Probabilities: Insights and Difficulties. In: Davidson, R., Swift, J. (Eds.). *Proceedings of the Second International Conference on Teaching Statistics*. Victoria: University of Victoria, 292-297.
- Falk, R. & Bar-Hillel, M. (1983). Probabilistic Dependence between Events. *Two-Year-College-Mathematics Journal* 14, 240-247.
- Feuerpfeil & Heigl (1997). *Wahrscheinlichkeitsrechnung und Statistik, Leistungskurs* (Neuaufgabe, orig. 1975). München: Bayerischer Schulbuch-Verlag.
- Fiedler, K. (1980). Kognitive Verarbeitung statistischer Information: Der „vergebliche Konsensus-Effekt“. *Zeitschrift für Sozialpsychologie* 11 (1), 25-37.
- Fiedler, K. (2000). Beware of Samples! A Cognitive-Ecological Sampling Approach to Judgment Biases. *Psychological Review* 107 (4), 659-676.
- Finetti, B. de (1989/1931). Probabilism. *Erkenntnis* 31, 169-223. (Orig. Arbeit 1931 veröff.)
- Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.
- Fischbein, E. (1987). *Intuition in Science and Mathematics: An Educational Approach*. Dordrecht: Reidel.
- Fischhoff, B., Slovic, P. & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance* 23, 339-359.
- Fischhoff, B. & Bar-Hillel (1984). Focusing techniques: A shortcut to improving probability judgments? *Organizational Behavior and Human Performance* 34, 175-194.
- Fisher, R.A. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.
- Fisher, R.A. (1956). *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd.
- Freudenthal, H. (1973). *Mathematik als pädagogische Aufgabe*, Band 2. Stuttgart: Klett.
- Freudenthal, H. (1983). *Didactical Phenomenology of Mathematical Structures*. Dordrecht: Reidel.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". In: W. Stroebe & M. Hewstone (Eds.). *European Review of Social Psychology* 2, 83-115.
- Gigerenzer, G. (1993a). The Superego, the Ego and the Id in Statistical Reasoning. In: Keren, G. & Lewis, C. (Eds.). *A handbook for data analysis in the behavioral sciences: Methodological issues*. Hillsdale, NJ: Lawrence Erlbaum, 311-339.
- Gigerenzer, G. (1993b). Über den mechanischen Umgang mit statistischen Methoden. In: Roth, E. & Heidenreich, K. (Hrsg.). *Sozialwissenschaftliche Methoden*, 3.Aufl. München: Oldenbourg.
- Gigerenzer, G. (1993c). Die Repräsentation von Information und ihre Auswirkung auf statistisches Denken. In: Hell, W., Fiedler, K. & Gigerenzer, G. (Hrsg.): *Kognitive Täuschungen - Fehl-Leistungen und Mechanismen des Urteilens, Denkens und Erinnerns*. Heidelberg: Spektrum, 99-127.
- Gigerenzer, G. (1996). On narrow norms and vague Heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review* 103, 592-596.
- Gigerenzer, G. (2002). *Das Einmaleins der Skepsis: Über den richtigen Umgang mit Zahlen und Risiken*. Berlin: Berlin Verlag.

- Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance* 14, 513-525.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 102, 684-704.
- Gigerenzer, G., Hoffrage, U. & Ebert, A. (1998). AIDS counselling for low risk clients. *AIDS Care* 10(2), 197-211.
- Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Lawrence Erlbaum.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1999). *Das Reich des Zufalls – Wissen zwischen Wahrscheinlichkeiten, Häufigkeiten und Unschärfen*. Heidelberg: Spektrum.
- Gigerenzer, G., Todd, P. & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Glaser, H., Scheid, H. & Wellstein, H. (1982). *Sigma – Grundkurs Stochastik*. Stuttgart: Klett.
- Godino, J.D. & Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in mathematics education. In: Sierpiska, A. & Kirpatrick, J. (Eds.). *Mathematics Education as a Research Domain: A Search for Identity*. Dordrecht: Kluwer.
- Götz, S. (1997). *Bayes-Statistik – ein alternativer Zugang zur beurteilenden Statistik in der siebenten und achten Klasse AHS*. Dissertation, Universität Wien.
- Götz, S. (2001). *Klassische und Bayessche Behandlung von Stochastik-Aufgaben in österreichischen Schulbüchern*. In: M. Borovcnik, J. Engel & D. Wickmann (Hrsg.). *Anregungen zum Stochastikunterricht*. Hildesheim: Franzbecker. 147-162
- Gould, S. J. (1992). *Bully for brontosaurus: Further reflections in natural history*. New York: Penguin Books.
- Griesel, H. & Postel, H. (1995, Hrsg.). *Elemente der Mathematik, Band 10*. Hannover: Schroedel.
- Griesel, H. & Postel, H. (2000, Hrsg.). *Elemente der Mathematik, Grundkurs NRW*. Hannover: Schroedel.
- Hacking, I. (1984). *The emergence of probability*. Cambridge University Press.
- Harten, G., von & Steinbring, H. (1984). *Stochastik in der Sekundarstufe I*. Köln: Aulis Deubner.
- Hasher, L. & Zacks, R.T. (1984). Automatic processing of fundamental information: the case of frequency of occurrence. *American Psychologist* 12, 1372-1388.
- Heitele, D. (1976). *Didaktische Ansätze zum Stochastikunterricht in Grundschule und Förderstufe*, Dissertation. Dortmund: Päd. Hochschule Ruhr.
- Hertwig, R. & Gigerenzer, G. (1999). The „conjunction fallacy“ revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making* 12, 275-305.
- Heymann, H.W. (1995). *Thesen zur Mathematiklehrerbildung aus der Perspektive eines Allgemeinbildungskonzeptes*. In: Biehler, Heymann & Winkelmann (Hrsg.). *Mathematik allgemeinbildend unterrichten: Impulse für Lehrerbildung und Schule. Untersuchungen zum Mathematikunterricht 21 (IDM-Reihe)*. Köln: Aulis.
- Hiebert, J. (1986, Ed.). *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum.
- Hoffrage, U. & Gigerenzer, G. (1998). Using Natural Frequencies to Improve Diagnostic Inferences. *Academic Medicine* 73, 538-540.
- Hoffrage, U., Kurzenhäuser, S. & Gigerenzer, G. (2000). Wie kann man die Bedeutung medizinischer Testbefunde besser verstehen und kommunizieren? *Zeitschrift für ärztliche Fortbildung und Qualitätssicherung* 94, 713-719.

- Hoffrage, U., Lindsey, S., Hertwig, R. & Gigerenzer, G. (2000). Communicating Statistical Information. *Science* 290, 2261-2262.
- Humphrey, G. (1951). *Thinking*. New York: Wiley.
- Huygens, C. (1657). *De ratiociniis in ludo aleae*. In: F. v. Schooten: *Exercitationes mathematicae*, Leyden.
- Inhelder, B. & Piaget, J. (1969). *The early growth of logic in the child*. New York: Norton Library.
- Israel, S. (2001). Was hat Aids mit Mathe zu tun? - Hinführung zum Satz von Bayes mit offenen Materialien. *Mathematik lehren* 104, S.62-66.
- Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford: University Press.
- Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: University Press.
- Kahneman, D. & Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology* 3, 430-454.
- Kahneman, D. & Tversky, A. (1973). On the Psychology of Prediction. *Psychological Review* 80, 237-251.
- Kapadia, R. (1984). *Children's intuitions and conceptions of probability*. London: Polytechnic of the South Bank.
- Konold, C. (1995). Issues in Assessing Conceptual Understanding in Probability and Statistics. *Journal of Statistics Education* 3 (1), online [Internet: <http://www.amstat.org/publications/jse/v3n1/konold.html>]
- Krämer, W. (1995). *So lügt man mit Statistik*. Frankfurt: Campus.
- Krauss, S. & Hertwig, R. (2000). Muss DNA-Evidenz schwer verständlich sein? Der Ausweg aus einem Kommunikationsproblem. *Monatsschrift für Kriminologie und Strafrechtsreform* 3, S. 155-162.
- Krauss, S., Martignon, L. & Hoffrage, U. (1999). Simplifying Bayesian Inference: The General Case. In: Magnani, L., Nersessian, N.J. & Thagard, P. (Eds.). *Model-Based Reasoning in Scientific Discovery*. New York: Kluwer Academic/Plenum.
- Krauss, S., Martignon, L., Hoffrage, U. & Gigerenzer, G. (eingereicht). Bayesian reasoning and natural frequencies: A generalization to complex situations.
- Krauss, S. & Wassner, C. (2001). Wie man das Testen von Hypothesen einführen sollte. *Stochastik in der Schule* 21 (1), 29-34.
- Kleiter, G. D. (1994). Natural sampling: Rationality without base rates. In: G. H. Fischer & D. Laming (Eds.). *Contributions to mathematical psychology, psychometrics, and methodology*. New York: Springer, 375-388.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Textkonzeption und Ergebnisse. In: Baumert, J. et al. (Hrsg.). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske & Budrich, 139-190.
- Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences* 19, 1-17.
- Koehler, J.J., Chia, A. & Lindsey, S. (1995). The random match probability (RMP) in DNA evidence: Irrelevant or prejudicial? *Jurimetrics* 35, 201-219.
- König, G. (1991). AIDS und Mathematikunterricht. *Zentralblatt für Didaktik der Mathematik* 23 (6), 207-220.
- [Kultus Berlin] Senatsverwaltung für Schule, Jugend und Sport (1987). *Rahmenplan für Unterricht und Erziehung in der Berliner Schule, Mathematik*. Berlin: H.Luchterhand.

- [Kultus Hamburg] Behörde für Bildung und Sport, Freie und Hansestadt Hamburg (2003). Rahmenplan Mathematik, Bildungsplan neunstufiges Gymnasium, Sekundarstufe I. Hamburg: Behörde für Bildung und Sport.
- [Kultusministerium Bayern] Bayerische Staatsministerien für Unterricht und Kultus und Wissenschaft, Forschung und Kunst (2003). Lehrplan für das Gymnasium in Bayern, Mathematik. Wolnzach: Kastner.
- [Kultusministerium Brandenburg] Ministerium für Bildung, Jugend und Sport des Landes Brandenburg (2002). Rahmenlehrplan Mathematik, Sekundarstufe I. Berlin: Wissenschaft und Technik.
- [Kultusministerium Mecklenburg-Vorpommern] Ministerium für Bildung, Wissenschaft und Kultur Mecklenburg-Vorpommern (2002). Rahmenplan Gymnasium, Mathematik, Jahrgangsstufen 7-10, Erprobungsfassung. Rogenzin: Adiant Druck.
- [Kultusministerium NRW] Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen (1993). Richtlinien und Lehrpläne für das Gymnasium, Sekundarstufe I, Mathematik. Frechen: Ritterbach.
- [Kultusministerium NRW] Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen (1999). Richtlinien und Lehrpläne für das Gymnasium, Sekundarstufe II, Mathematik. Frechen: Ritterbach.
- [Kultusministerium Sachsen] Sächsisches Staatsministerium für Kultus (2001). Lehrplan Gymnasium (gewichtete Fassung), Mathematik, Klassen- und Jahrgangsstufen 5-12. Dresden: Sächsisches Druck- und Verlagshaus.
- Kurzenhäuser, S. (2001). Risikokommunikation in der BSE-Krise – Illusorische Sicherheit und Transparenz. Bundesgesundheitsblatt für Gesundheitsforschung und Gesundheitsschutz 44, 336-340.
- Kurzenhäuser, S. & Hoffrage, U. (2002). Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher* 24, 516-521.
- Kuypers, W., Lauter, J. & Wutke, H. (1999, Hrsg.). *Mathematik 10. Schuljahr*. Berlin: Cornelsen
- Laplace, P.S. de (1814/1951). *A philosophical essay on probabilities* (F.W. Truscott & F.L. Emory, Transl.). New York: Dover. (Orig. Arbeit 1814 veröff.)
- Lergenmüller, A. & Schmidt, G. (2003, Hrsg.). *Mathematik – Neue Wege 9, Arbeitsbuch für Gymnasien*. Hannover: Schroedel
- Leibniz, G.W. (1703/1962). *G.W. Leibniz Mathematische Schriften* (Hrsg. von C.I. Gerhardt). Hildesheim: Olms. (Neudr. der Ausg. v. 1855).
- Lindeman, S. T., van den Brink, W. P., & Hoogstraten, J. (1988). Effect of feedback on base-rate utilization. *Perceptual and Motor Skills* 67, 343 - 350.
- Lindley, D.V. (1965). *Introduction to probability and statistics from a Bayesian viewpoint. Part 1: Probability*. Cambridge: University Press.
- Lindsey, S., Hertwig, R. & Gigerenzer, G. (im Druck). *Communicating statistical evidence. Jurimetrics*
- Lopes, L. L. (1991). The rhetoric of irrationality. *Theory and Psychology* 1, 65-82.
- Lücking, A. (2003). *The Development of Bayesian Reasoning in Children*, Diplomarbeit. Berlin: Freie Universität.
- Mandl, H. & Levin, J.R. (1989). *Knowledge acquisition from text and pictures*. Amsterdam: Elsevier.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Martignon, L. (2001). Comparing fast and frugal heuristics and optimal models. In: Gigerenzer, G. & Selten, R. (Eds.). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press, 147-171.

- Martignon, L. & Laskey, K. (1999). Bayesian Benchmarks for Fast and Frugal Heuristics. In: Gigerenzer, G., Todd, P. et al. (Eds.). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Martignon, L. & Hoffrage, U. (1999). Why Does One-Reason Decision Making Work?. In: Gigerenzer, G., Todd, P. et al. (Eds.). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Martignon, L., Vitouch, O., Masanori, T. & Foster, M.R. (2003). Naive and yet Enlightened: From Natural Frequencies to Fast and Frugal Decision Trees. In: Hardman, D. & Macchi, L. (Eds.). *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making*. New York: Wiley.
- Martignon, L. & Wassner, C. (2001). Repräsentation von Information in der Wahrscheinlichkeitstheorie. In: Borovcnik, M., Engel, J. & Wickmann, D. (Hrsg). *Anregungen zum Stochastikunterricht*. Hildesheim: Franzbecker, 163-170.
- Moore, D.S. (1997). Bayes for Beginners? – Some Reasons to Hesitate. *The American Statistician* 51(3), 254-261.
- [NCTM] National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, Va.: National Council of Teachers of Mathematics.
- [NCTM] National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, Va.: National Council of Teachers of Mathematics.
- Neubrand, J. (2002). Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen – Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie. Hildesheim.
- Neubrand, M., Klieme, E., Lüdtke O. & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung. *Unterrichtswissenschaft* 30 (2), 100-119.
- Neyman, J. and Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Ser. A* 231, 289-337.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Paivio, A. (1978). A dual coding approach to perception and cognition. In: Pick, H.L. & Saltzman, E. (eds.). *Modes of perceiving and processing information*. Hillsdale, NJ: Lawrence Erlbaum.
- Peterson, C.R. & Beach, L.R. (1967). Man as an Intuitive Statistician. *Psychological Bulletin* 68, 29-46.
- Peterson, C. R., DuCharme, W. M., & Edwards, W. (1968). Sampling distributions and probability revision. *Journal of Experimental Psychology* 76, 236-243.
- Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability model inference task. *Journal of Experimental Psychology* 72, 346-354.
- Piaget, J. & Inhelder, B. (1975). *The origin of the idea of chance in children* (L. Leake, Jr., P. Burrell & H.D. Fischbein, Trans.). New York: Norton (Franz. Originalausg. 1951).
- Piatelli-Palmarini, M. (1994): *Inevitable illusions: How mistakes of reason rule our minds*. New York: Wiley & Sons.
- Pollatsek, A., Well A., Konold, C., Hardiman, P. & Cobb, G. (1987): Understanding Conditional Probabilities. *Organizational Behaviour and Human Decision Processes* 40, 255-269.
- Rathgeber, C. (1999). Offener Stochastikunterricht. *Praxis der Mathematik* 41(1), 6-11.
- Rasfeld, P. (2004). Verbessert der Stochastikunterricht intuitives stochastisches Denken? Ergebnisse aus einer empirischen Studie. *Journal für Mathematik-Didaktik* 25 (1), 33-61.

- Redmayne, M. (1998). The DNA database: Civil liberty and evidentiary issues. *Criminal Law Review*, 437-454.
- Rheinberg, F. (2001). Bezugsnormen und schulische Leistungsbeurteilung. In: Weinert, F.E. (Hrsg.). *Leistungsmessungen in Schulen*. Weinheim: Beltz, 59-72.
- Riemer, W. (1985). *Neue Ideen zur Stochastik - Lehrbücher und Monographien zur Didaktik der Mathematik*, Band 3. Mannheim: Bibliographisches Institut.
- Riemer, W. (1986). Eine neue Sicht der Bayesschen Regel. *Stochastik in der Schule* 6 (3), 4-13.
- Riemer, W. (1991). Zur Rettung des gesunden Menschenverstandes in der Statistik: Skizze einer für den NRW-Lehrplan erprobten Lernsequenz. *Stochastik in der Schule* 11 (3), 24-36.
- Robert-Koch-Institut (2002). *Epidemiologisches Bulletin. HIV-Infektionen und AIDS-Erkrankungen in Deutschland, aktuelle epidemiologische Daten*. Berlin: P. Fieck. [online: <http://www.rki.de/infekt/epibull/epi.htm>]
- Ross, N., Jordan, A., Baumert, J., Blum, W., Brunner, M., Krauss, S., Löwen, K. & Neubrand, M. (im Druck). *Klassifikationsschema für Mathematikaufgaben*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Savage, L.J. (1954). *The foundations of statistics*. New York: Dover.
- Savage, L.J. (1976). On rereading R.A. Fisher (with discussion). *Annual Statistician* 4, 441-500.
- Schaefer, R. E. (1976). The evaluation of individual and aggregated subjective probability distributions. *Organizational Behavior and Human Performance* 17, 199-210.
- Schmid, A & Weidig, I. (1996, Hrsg.). *Lambacher Schweizer 9 – Mathematisches Unterrichtswerk für das Gymnasium, Ausgabe NRW*. Stuttgart: Klett
- Scholz, R.W. (1981). *Stochastische Problemaufgaben - Analysen aus didaktischer und psychologischer Perspektive*. IDM-Reihe Materialien und Studien, Band 23. Bielefeld: Universität Bielefeld.
- Scholz, R.W. (1987). *Cognitive Strategies in Stochastic Thinking*. Dordrecht: Reidel.
- Scholz, R.W. (1991). Psychological Research in Probabilistic Understanding. In: Kapadia, R. & Borovcnik, M. (Eds.). *Chance Encounters: Probability in Education*. Dordrecht: Kluwer, 213-254.
- Schrader, F.W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In: Weinert, F.E. (Hrsg.). *Leistungsmessungen in Schulen*. Weinheim: Beltz, 45-58.
- Schupp, H. (1984). Sinnvoller Stochastik-Unterricht in der Sekundarstufe I. *Mathematica didactica* 7, 233-243.
- Schupp, H. (im Druck). *Allgemeinbildender Stochastikunterricht (Hauptvortrag auf der Herbsttagung des AK Stochastik der GDM 2003, Kassel)*. *Stochastik in der Schule* 24 (2004).
- Sedlmeier, P. (1993). Training zum Denken unter Unsicherheit. In: W. Hell, K. Fiedler & G. Gigerenzer (Hrsg.). *Kognitive Täuschungen*. Heidelberg: Spektrum, 129-160.
- Sedlmeier, P. (1996). Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen. *Methods of Psychological Research* 1, online [Internet: <http://www.pabst-publishers.de/mpr/>]
- Sedlmeier, P. (1997). BasicBayes: A tutor system for simple Bayesian inference. *Behavior Research Methods, Instruments & Computers* 29, 328-336.
- Sedlmeier, P. (1998). The distribution matters: Two types of sample-size tasks. *Journal of Behavioral Decision Making* 11, 281-301.
- Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Lawrence Erlbaum.
- Sedlmeier, P. (2000). How to improve statistical thinking: Choose the task representation wisely and learn by doing. *Instructional Science* 28, 227-262.

- Sedlmeier, P. (2001). Statistik ohne Formeln. In: M. Borovcnik, J. Engel & D. Wickmann (Hrsg.). Anregungen zum Stochastikunterricht. Hildesheim: Franzbecker, 83-95.
- Sedlmeier, P. (2002). Associative Learning and Frequency Judgements: The PASS model. In: Sedlmeier, P. & Betsch, T. (Eds.). Etc. – Frequency Processing and Cognition. Oxford: University Press, 137-152.
- Sedlmeier, P. & Betsch, T. (2002). Etc. – Frequency Processing and Cognition. Oxford: University Press.
- Sedlmeier, P. & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General* 130 (3), 380-400.
- Sedlmeier, P. & Köhlers, D. (2001). Wahrscheinlichkeiten in der Praxis. Braunschweig: Westermann.
- Simon, H.A. (1956). Rational choice and the structure of the environment. *Psychological Review* 63, 129-138.
- Simon, H.A. (1990). Alternative visions of rationality. In: Moser, P.K. (Ed.). *Rationality in Action: Contemporary Approaches*. Cambridge: Univ. Press, 189-204.
- Shaughnessy, J.M. (1977). Misconceptions of Probability: An experiment with a small-group, activity-based, model-building approach to introductory probability at the college level. *Educational Studies in Mathematics* 8, 295-316.
- Shaughnessy, J.M. (1992). Research in Probability and Statistics: Reflections and Directions. D. Grouws (Ed.). *Handbook on Research in Mathematics Teaching and Learning*. New York: Macmillan, 465-494.
- Stern, E. & Hardy, I. (2001). Schulleistungen im Bereich der mathematischen Bildung. In: Weinert, F.E. (Hrsg.). *Leistungsmessungen in Schulen*. Weinheim: Beltz, 153-168.
- Sternberg, R. (1997). *Thinking Styles*. New York: Cambridge University Press.
- Strick, H. K. (1998). Einführung in die beurteilende Statistik. Hannover: Schroedel.
- Strick, H.K. (1999). Vierfeldertafeln im Stochastikunterricht der Sek. I und II. *Praxis der Mathematik* 41(2), 49-57.
- Strick, H.K. (2000). Über die Schwierigkeiten, verständlich über Vorsorgemaßnahmen zur Krebsfrüherkennung zu informieren. *Praxis der Mathematik* 42(6), 247-248.
- Tomlinson, S. & Quinn, R. (1997). Understanding Conditional Probability. *Teaching Statistics* 19 (1), 2-7. (Dt. Übersetzung in: *Stochastik in der Schule* 18 (1998, Nr.3), 27-40)
- Tooby, J. & Cosmides, L. (1992). Ecological Rationality and the Multimodula Mind. Rep.No.92-1. Santa Barbara: Univ. of California, Center for Evolutionary Psychology.
- Tukey, J.W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5, 207-232.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185, 1124-1131.
- Tversky, A. & Kahneman, D. (1982). Evidential impact of Base Rates. In: Kahneman, D., Slovic, P. & Tversky, A. (Eds.). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: University Press, 153-160.
- Wagenschein, M. (1975). *Verstehen lernen. Genetisch – Sokratisch – Exemplarisch*, 5.Aufl. Weinheim: Beltz.
- Wallsten, T.S. (1980). Processes and Models to describe Choice and Inference Behaviour. In: Wallsten, T.S. (Ed.). *Cognitive Processes in Choice and Decision Behaviour*. Hillsdale, NJ: Lawrence Erlbaum, 215-237.
- Wallsten, T. S. (1983). The theoretical status of judgmental heuristics. In Scholz, R. W. (Ed.). *Decision making under uncertainty*. Amsterdam: Elsevier, 21-39.

- Walter, H. (1983). Heuristische Strategien und Fehlvorstellungen in stochastischen Situationen. *Der Mathematikunterricht* 29, 11-23.
- Wassner, C. & Biehler, R. (im Druck). Authentisches Bewerten und Urteilen unter Unsicherheit in der Sekundarstufe I. In: Reiss, K. et al. (Hrsg.). *Beiträge zum Mathematikunterricht 2004*. Hildesheim: Franzbecker.
- Wassner, C., Krauss, S. & Martignon, L. (2002). Muss der Satz von Bayes schwer verständlich sein? *Praxis der Mathematik* 44 (1), 12-16.
- Wassner, C. & Martignon, L. (2002). Teaching decision making and statistical thinking with natural frequencies. In: Phillips, B. (Ed.). *Proceedings of the sixth international conference on teaching of statistics*, South Africa. Voorburg (NL): IASE/International Statistical Institute.
- Wassner, C. & Martignon, L. & Biehler, R. (2004). Bayesianisches Denken in der Schule. *Unterrichtswissenschaft* 32 (1). 58-96.
- Wassner C., Martignon L. & Sedlmeier, P. (2002). Die Bedeutung der Darbietungsform für das alltagsorientierte Lehren von Stochastik. In: Prenzel, M. & Doll, J. (Hrsg.). *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen, Teil I: Unterrichtsforschung in Mathematik*. 45. Beiheft der *Zeitschrift für Pädagogik*. 35-50.
- Weinert, F.E. (2001, Hrsg.). *Leistungsmessungen in Schulen*. Weinheim: Beltz.
- Wickmann, D. (1990). *Bayes-Statistik - Einsicht gewinnen und entscheiden bei Unsicherheit, Lehrbücher und Monographien zur Didaktik der Mathematik, Band 4*. Mannheim: BI-Wissenschaftsverlag.
- Wickmann, D. (1998). Zur Begriffsbildung im Stochastikunterricht. *Journal für Didaktik der Mathematik* 19 (1), 46-80.
- Wickmann, D. (2001). Der Theorieneintopf ist zu beseitigen – Versuch einer Klärung des Wahrscheinlichkeitsbegriffs zum Zwecke der Methodenbereinigung. In: M. Borovcnik, J. Engel & D. Wickmann (Hrsg.). *Anregungen zum Stochastikunterricht*. Hildesheim: Franzbecker. 123-132.
- Wittenberg, A.I. (1963). *Bildung und Mathematik*. Stuttgart: Klett.
- Wittmann, E. (1981). The Complementary Roles of Intuitive and Reflective Thinking in Mathematics Teaching. *Educational Studies in Mathematics* 12, 389-397.
- Wollring, B. (1992). Ein Beispiel zur Konzeption von Simulationen bei der Einführung des Wahrscheinlichkeitsbegriffs. *Stochastik in der Schule* 12 (3). 2-25.
- Wollring, B. (1993). Spielinterviews zur Erkundung stochastischer Vorstellungen bei Kindern im Vor- und Grundschulalter. In: Lorenz, J.H. (Hrsg.). *Mathematik und Anschauung, IDM-Reihe*. Köln: Aulis, 67-91.
- Wollring, B. (1994). Fallstudien zu frequentistischen Kompetenzen von Grundschulkindern in stochastischen Situationen - Kinder rekonstruieren verdeckte Glücksräder. In: Maier, H., Voigt, J. (Hrsg.): *Verstehen und Verständigung, IDM-Reihe*. Köln: Aulis, 144-181.
- Yost, P.A., Siegel, A.E. & Andrews, J.M. (1962). Nonverbal Probability Judgements by Young Children. *Child Development* 33, 769-780.
- Zacks, R.T. & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In: Sedlmeier, P. & Betsch, T. (2002). *Etc. – Frequency Processing and Cognition*. Oxford: University Press.
- Zhu, L. & Gigerenzer, G. (eingereicht). *Children and Bayesian Problems: The Role of Representation in Computation*.