Forschungsergebnisse aus dem
Revisionswesen und der betriebswirtschaftlichen Steuerlehre

Band 16

Risikoorientierte Datenprüfung bei unscharfen Informationen

Von

Stefan Göbel

Duncker & Humblot · Berlin

STEFAN GÖBEL

Risikoorientierte Datenprüfung bei unscharfen Informationen

Forschungsergebnisse aus dem Revisionswesen und der betriebswirtschaftlichen Steuerlehre

Herausgegeben von Prof. Dr. Erich Loitlsberger, Prof. Dr. Dieter Rückle und Prof. Dr. Jörg Baetge

Band 16

Risikoorientierte Datenprüfung bei unscharfen Informationen

Von

Stefan Göbel

Duncker & Humblot · Berlin

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Göbel, Stefan:
Risikoorientierte Datenprüfung bei unscharfen Informationen /
von Stefan Göbel. – Berlin : Duncker und Humblot, 1998
(Forschungsergebnisse aus dem Revisionswesen und
der betriebswirtschaftlichen Steuerlehre ; Bd. 16)
Zugl.: Regensburg, Univ., Habil.-Schr., 1996
ISBN 3-428-08909-X

Fotoprint: Berliner Buchdruckerei Union GmbH, Berlin
Printed in Germany

ISSN 0720-6909
ISBN 3-428-08909-X

Gedruckt auf alterungsbeständigem (säurefreiem) Papier
entsprechend ISO 9706 ∞

Vorwort

Der vorliegenden Arbeit liegt meine im Frühjahr 1996 von der Wirtschaftswissenschaftlichen Fakultät der Universität Regensburg unter dem gleichen Titel angenommene Habilitationsschrift zugrunde. Ziel der Arbeit ist es, anknüpfend an die Anfang der 80er Jahre unterbrochene Theoriediskussion im Bereich des Prüfungswesens die in neuerer Zeit aufkommenden Ansätze zur Orientierung von Prüfungen am Risiko theoretisch zu fundieren. Dazu muß neben einer operationalen Definition des Prüfungsrisikos auch ein geeignetes Maß für das Prüfungsrisiko gefunden werden, das der vorliegenden Prüfungssituation angemessen ist, da ohne die Möglichkeit, das Prüfungsrisiko zu messen, auch keine Orientierung der Prüfungshandlungen am Risiko möglich erscheint. Im Unterschied zu den bisher vorgeschlagenen risikoorientierten Prüfungsansätzen sollen nicht die einzelnen Risikokomponenten, sondern die unterschiedlichen Formen von Unschärfe der Prüfungsinformationen, die als ursächlich für die Entstehung von Risiko identifiziert werden, Ausgangspunkt der Untersuchung sein. Daher wird zunächst die Unschärfe, die bei den Prüfungsinformationen aus einzelnen Prüfungsmethoden auftreten kann, untersucht. Anschließend werden unterschiedliche Maße, namentlich Wahrscheinlichkeiten, Glaubwürdigkeiten sowie allgemeine Unschärfemaße aus der Fuzzy Set Theorie, im Hinblick auf ihre Eignung zur Messung des Prüfungsrisikos untersucht. Anschließend wird geprüft, inwieweit die Ergebnisse aus den bei Prüfungen eingesetzten Methoden zu einem Gesamtrisikomaß aggregiert werden können bzw. welche Weiterentwicklungen bei den einzelnen Prüfungsmethoden notwendig werden, um deren Integration in einen risikoorientierten Prüfungsansatz zu ermöglichen.

Die Arbeit wendet sich dabei sowohl an den Wissenschaftler als auch an den theoretisch interessierten Praktiker, der unterschiedliche risikoorientierte Prüfungsansätze kritisch hinterfragen möchte. Insbesondere der zuletzt genannte Personenkreis läßt sich, so hoffe ich, nicht von der in manchen Teilen notwendigen formalen Darstellung abschrecken, durch die eine aussagekräftige Analyse der vorliegenden Prüfungssituation und der sich für die Risikomessung ergebenden Konsequenzen erst möglich wird.

Mein besonderer Dank gilt Herrn Professor Dr. Gerhard Scherrer, der mir die für die Entstehung der Arbeit notwendigen Freiheiten einräumte und durch seine stete Diskussionsbereitschaft und vielfältige Anregungen die Entstehung und den Abschluß der vorliegenden Arbeit gefördert hat. Auch Herrn Professor Dr. Alfred Hamerle bin ich für seine Tätigkeit als Gutachter im Habilitationsverfahren und die vielfältigen Verbesserungsvorschläge dankbar. Für seine Unterstützung, eine konsistente formale Darstellung innerhalb der Arbeit zu erreichen, gilt mein Dank Herrn Dr. Reinhold Hübl. Meine Kollegen an der Wirtschaftswissenschaftlichen Fakultät haben die Arbeit durch ihre Bereitschaft, einzelne Probleme innerhalb des behandelten Themenkomplexes zu diskutieren, gefördert. Die noch verbleibenden Fehler gehen natürlich zu Lasten des Verfassers. Für die gute Verlagsbetreuung und ihre Geduld danke ich Frau Anja Papenfuß und last but not least den Herausgebern für die Aufnahme der Arbeit in die Reihe „Forschungsergebnisse aus dem Revisionswesen und der betriebswirtschaftlichen Steuerlehre".

Regensburg, im Oktober 1997

Stefan Göbel

Inhaltsverzeichnis

Problemstellung

Erster Teil

Prüfungsmodell bei unscharfer Information

Zweiter Teil

Risikomessung im betrachteten Prüfungsmodell

Dritter Teil

Ermittlung des Gesamtrisikos

Ergebnisse

DOI https://doi.org/10.3790/978-3-428-48909-1

Abbildungsverzeichnis

Abkürzungsverzeichnis

a. A.	anderer Auffassung
Abb.	Abbildung
Abs.	Absatz
AICPA	American Institute of Certified Public Accountants
Aufl.	Auflage
Bd.	Band
bearb.	bearbeitet
Beck Bil.-Komm	Beck'scher Bilanz-Kommentar, bearb. von Wolfgang Dieter Budde u. a.
BFuP	Betriebswirtschaftliche Forschung und Praxis (Zeitschrift)
BHR	Bonner Handbuch Rechnungslegung, hrsg. von Max A. Hofbauer u. a.
bzw.	beziehungsweise
c. p.	ceteris paribus
d. h.	das heißt
Diss.	Dissertation
ed.	edition
et al.	et alii, und andere
FG	Fachgutachten
Fn.	Fußnote(n)
h. M.	herrschende Meinung
HFA	Hauptfachausschuß des Instituts der Wirtschaftsprüfer in Deutschland
HGB	Handelsgesetzbuch
hrsg.	herausgegeben
HWB	Handwörterbuch der Betriebswirtschaft, hrsg. von Waldemar Wittmann u.a.
HWRev	Handwörterbuch der Revision, hrsg. von Adolf Gerhard Coenenberg und Klaus v. Wysocki
i. w. S.	im weiteren Sinn
IDW	Institut der Wirtschaftsprüfer in Deutschland e.V.
Jg.	Jahrgang
No.	Numero, Number
Rz.	Randziffer
SAS	Statement on Auditing Standards
Sp.	Spalte

u. a.	und andere
vgl.	vergleiche
Vol.	Volume
WPg	Die Wirtschaftsprüfung (Zeitschrift)
WPK	Wirtschaftsprüferkammer
WPO	Wirtschaftsprüferordnung
z. B.	zum Beispiel
ZfB	Zeitschrift für Betriebswirtschaft
ZfbF	Zeitschrift für betriebswirtschaftliche Forschung
zugl.	zugleich

Symbolverzeichnis

A	Menge der möglichen Handlungsalternativen im Entscheidungsmodell bzw. beliebige Menge
a_1	Element aus A: Beurteilung der Prüfungsgesamtheit als ordnungsmäßig
a_2	Element aus A: Beurteilung der Prüfungsgesamtheit als nicht ordnungsmäßig
$\tilde{A},\tilde{B},\tilde{E}$	beliebige unscharfe Mengen
B,C,E	beliebige gewöhnliche Mengen
bel	Glaubwürdigkeitsmaß
b_X, b_Y	Bewertungsfunktion für einzelne Elemente aus X bzw. Y
$\mathbb{C}$	Kontextmenge
D	Menge aller fehlerhaften Elemente aus X und den zugehörigen Sollobjekten aus Y
$\Delta(X,Y)$	wertmäßige Abweichung zwischen X und Y
Δ^*	maximal zulässige wertmäßige Abweichung zwischen X und Y
δ	Fehlergrenze
e	Stichprobenfehler bzw. Euler'sche Zahl
f	allgemeines Unschärfemaß
Γ	Kompatibilitätsrelation
I	Menge der möglichen Ausprägungen der Informationen im Entscheidungsmodell
$\mathbb{K}$	Ereigniskörper
m	grundlegende Wahrscheinlichkeitszuordnung
N	Anzahl der Elemente in der Prüfungsgesamtheit
n	Stichprobenumfang
nec	Notwendigkeitsmaß
O	Menge der in Bezug auf das betrachtete Unternehmen ordnungsmäßigen Abbildungen der Sachverhalte aus U'
o	logische Aussage: „Die Prüfungsgesamtheit ist ordnungsmäßig."
Ω	Möglichkeitsraum
pl	Plausibilitätsmaß
pos	Möglichkeitsmaß
pr	Wahrscheinlichkeitsmaß
π_D	Zugehörigkeitsfunktion für D, Fehlerfunktion
π_X, π_Y	Zugehörigkeitsfunktion für X bzw. Y
Θ	frame of discernment
θ	Fehleranteil in der Prüfungsgesamtheit
θ^*	maximal zulässiger Fehleranteil

$\mathbb{R}$	Menge der reellen Zahlen
S	Menge der möglichen Umweltzustände im Entscheidungsmodell
s_1	Umweltzustand im Entscheidungsmodell: tatsächliche Ordnungsmäßigkeit der Prüfungsgesamtheit
s_2	Element aus S: tatsächliche Nichtordnungsmäßigkeit der Prüfungsgesamtheit
U	Menge der realen Sachverhalte beschränkt auf die für die Abbildung relevanten Merkmale
U'	Menge der realen Sachverhalte
ω_X, ω_Y	Meßfunktion des Unternehmens bzw. des Prüfers zur Abbildung von U' auf den Merkmalsraum von U
X	Prüfungsgesamtheit
x_i	Element i der Prüfungsgesamtheit X
Y	Menge aller Sollobjekte der Prüfung
y_i	Sollobjekt zu x_i

Problemstellung

A. Probleme risikoorientierter Prüfung

Die Motivation insbesondere im Rahmen von betriebswirtschaftlichen Prüfungen[1], bestimmte Entscheidungen am Risiko zu orientieren, resultiert aus der für die Prüfung zu unterstellenden mehrfachen Zielsetzung im Hinblick auf den Grad der Sicherheit des abzugebenden Urteils einerseits und der angestrebten Gewinnerzielung andererseits[2]. Wird ein hoher Sicherheitsgrad des Urteils mit geringem Risiko gleichgesetzt, so kann unter risikoorientierter Prüfung eine bestimmte Vorgehensweise des Prüfers bei der Analyse und Beurteilung von Daten im Hinblick auf bestimmte Vorgaben verstanden werden. Risikoorientierung heißt in diesem Zusammenhang, daß der Prüfer zum einen die Auswahl von Prüfungshandlungen daran orientiert, welchen Beitrag einzelne Prüfungshandlungen in der jeweils vorliegenden Situation zur Verminderung seines Risikos leisten können[3], und zum anderen die Entscheidung für eine bestimmte Ausprägung seines Urteils am damit verbundenen Risiko ausrichtet. Risikoorientierte Prüfung erfordert somit, daß das Risiko meßbar ist, da ansonsten zumindest seine Verminderung nicht festgestellt werden kann.

Eine genaue Definition des Begriffs „Risiko" erfolgt im Verlauf der Arbeit[4], an dieser Stelle soll eine Abgrenzung und Charakterisierung ausschließlich in dem Umfang, der für die weiteren Ausführungen benötigt wird, erfolgen. Gegenstand der weiteren Betrachtungen ist ausschließlich das Risiko des Prüfers, ein falsches Urteil über die geprüften Daten abzugeben. Insbesondere das Geschäftsrisiko des Prüfers oder des Prüfungsunternehmens, für das der Prüfer tätig wird, sowie die Risiken, die bei dem geprüften Unternehmen auftreten können[5], werden nicht in die Untersuchung einbezogen. Darüber hinaus wird für das vom Prüfer abzugebende Urteil nur unterstellt, daß unterschiedliche Aus-

---

1 Eine genaue Definition des Begriffs erfolgt in Kapitel A. im Ersten Teil der Arbeit.

2 Vgl. *Leffson* et al., Sicherheit, 1969, S. 19; *v. Wysocki*, Wirtschaftlichkeit, 1992, Sp. 2176; *Thoennes*, Prüfungsansatz, 1994, S. 34.

3 Vgl. *Alderman/Tabor*, risk-driven audits, 1989, S. 56f.

4 Siehe unten Kapitel A. im Zweiten Teil der Arbeit.

5 Zu den unterschiedlichen Risiken vgl. *Dörner*, Audit Risk, 1992, Sp. 81f.

prägungen, zwischen denen der Prüfer wählen muß, möglich sind. Welche Ausprägung der Prüfer für sein Urteil wählt, hängt dabei von seinen Einschätzungen über die geprüften Daten ab. Dies setzt voraus, daß dem Prüfer Informationen vorliegen, aus denen er seine Einschätzungen ableiten kann. Zumindest in der Mehrzahl der Fälle werden dabei die einzelnen Ausprägungen des Urteils unterschiedlich stark durch die vorliegenden Informationen gestützt sein, so daß das Risiko ausschließlich im Hinblick auf eine bestimmte Ausprägung des Prüfungsurteils gemessen werden kann.

Darüber hinaus wird unterstellt, daß der Begriff des Risikos in einem theoretischen Sinn zu definieren ist, da bei der Ableitung des Risikomaßes ausschließlich der Umstand, wie gut die Überzeugung des Prüfers für eine bestimmte Ausprägung des Prüfungsurteils durch die vorliegenden Prüfungsergebnisse gestützt ist, berücksichtigt wird, und nicht die Konsequenzen, welche die Entscheidung für eine bestimmte Ausprägung des Prüfungsurteils hat, einbezogen werden sollen. Unterschiede zwischen den beiden Ansätzen ergeben sich dann, wenn die Konsequenzen der möglichen Fehlurteile unterschiedlich bewertet werden. Wie später deutlich wird, würde jedoch bei der Berücksichtigung der Konsequenzen des Prüfungsurteils die Risikomessung in einem nicht unerheblichen Umfang von der Bewertung dieser Konsequenzen beeinflußt. Demnach müßte zur Risikomessung auf die Untersuchung der Konsequenzen ebensoviel Wert gelegt werden wie auf die Vornahme von Prüfungshandlungen. Da die Konsequenzen des Prüfungsurteils jedoch regelmäßig unabhängig von den gewählten Prüfungshandlungen sind und insoweit auf die Auswahl der Prüfungshandlungen keinen Einfluß haben können, sollen deren Auswirkungen auf das Risiko nicht weiter betrachtet werden. Insoweit wird der Begriff des Risikos auf solche Sachverhalte, die durch die Ergebnisse der Prüfungshandlungen beeinflußt werden können, begrenzt.

Da für die Begründung des Prüfungsurteils ausschließlich die Überzeugung des Prüfers ausreicht, ergibt sich als Konsequenz, daß zur Abgabe eines negativen Prüfungsurteils nicht alle in den geprüften Daten enthaltenen Fehler tatsächlich gefunden werden müssen, soweit die vorliegenden Informationen aus der Sicht des Prüfers das Urteil ausreichend stützen. Insoweit wird die Aufgabe des Prüfers auch nicht in der Korrektur von eventuell in den Daten enthaltenen Fehlern gesehen, die es dann, wenn die Fehler beseitigt sind, erlauben würde, ein positives Urteil abzugeben[6]. Es wird in der vorliegenden Arbeit vielmehr

---

6 So zumindest für die Jahresabschlußprüfung gemäß §§ 316ff. HGB *Leffson*, Wirtschaftsprüfung, 1988, S. 326; *Quick*, Risiken, 1996, S. 2.

davon ausgegangen, daß die Erstellung der Daten und deren Prüfung als getrennte Prozesse zu verstehen sind[7]. Dabei soll nicht ausgeschlossen werden, daß ein negatives Prüfungsurteil zur Überarbeitung der Daten und ihrer erneuten Prüfung führen kann. Allerdings erfolgt eine positive Beurteilung dann nicht schon aus dem Grund, daß die zuvor aufgedeckten Fehler korrigiert sind, sondern erst, wenn die neu ermittelten Prüfungsinformationen dies zulassen. Zumindest für die handelsrechtliche Jahresabschlußprüfung läßt sich zur Rechtfertigung der zuletzt genannten Forderung der Wortlaut des § 319 Abs. 2 Nr. 5 HGB heranziehen. Dort werden explizit Abschlußprüfer, die über die Prüfungshandlungen hinaus bei der Erstellung des Jahresabschlusses mitgewirkt haben, von der Prüfung ausgeschlossen. Auch wenn der Begriff der „unerlaubten Mitwirkung" regelmäßig eng gefaßt wird[8], müßte zumindest für den Fall, daß alle vorgelegten Daten fehlerhaft sind, die Erstellung des Jahresabschlusses dem Prüfer zugerechnet werden, da er durch die Korrekturen tatsächlich den Jahresabschluß erstellt.

Für die im Rahmen einer risikoorientierten Prüfung notwendige Risikomessung sollen zwei weitere Annahmen getroffen werden. Zum einen wird die Messung des Risikos auf einer metrischen Skala angestrebt. Zum anderen soll als Ursache für das Entstehen von Risiko ausschließlich die Unschärfe der vorliegenden Prüfungsinformationen betrachtet werden. Es wird dazu insbesondere unterstellt, daß das Risiko durch ein noch näher zu bestimmendes Unschärfemaß gemessen werden kann. In diesem Zusammenhang wird sowohl der Begriff der Prüfungsinformation als auch der Begriff der Unschärfe weit gefaßt. Letztere soll unabhängig sowohl von den Gründen für ihre Entstehung, wie z. B. der Unvollständigkeit der vorliegenden Informationen oder der Komplexität der unterstellten Zusammenhänge, als auch von den Ansätzen, die zur Abbildung der Unschärfe zur Verfügung stehen, wie Wahrscheinlichkeiten oder unscharfe Mengen (fuzzy sets), alle Formen unscharfer Informationen umfassen. Zu den Prüfungsinformationen wird neben den im Rahmen der Prüfungshandlungen ermittelten Ergebnissen auch das Wissen des Prüfers über bestimmte Zusammenhänge und Gesetzmäßigkeiten im Hinblick auf die zu beurteilenden Daten gezählt.

Ziel der vorliegenden Arbeit ist es, zunächst eine theoretische Konzeption für die Risikomessung in der angeführten Prüfungssituation zu entwickeln. Dabei soll auch untersucht werden, inwieweit das Risiko intersubjektiv überprüf-

---

7 Vgl. *Wittmann*, Systemprüfung, 1980, S. 11f.

8 Vgl. *v. Wysocki*, Grundlagen, 1988, S. 67f.; *Adler/Düring/Schmaltz*, 1987, § 319 HGB Rz. 62.

bar gemessen oder zumindest das Gesamtrisiko einer Entscheidung aus der Einschätzung von Einzelrisiken nach bestimmten allgemeingültigen Regeln abgeleitet werden kann. Es handelt sich insoweit um eine normative Theorie der Risikomessung im Rahmen der Prüfung. Eine Übereinstimmung der für das vorliegende Modell beschriebenen Methode der Risikomessung mit den durch empirische Untersuchungen ermittelten Ergebnissen wird also nicht angestrebt[9]. Die hier zu entwickelnde Theorie der Risikomessung ist dabei ausschließlich als Teil einer umfassenden Prüfungstheorie zu sehen. Insbesondere die bei der Messung von Abweichungen für die Bildung von Einzelurteilen benötigten theoretischen Konzeptionen für die Ableitung von Sollobjekten in den einzelnen Bereichen sind nicht Gegenstand der vorliegenden Arbeit.

Erst wenn die Probleme der Risikomessung gelöst sind, kann geprüft werden, inwieweit eine risikoorientierte Auswahl von Prüfungshandlungen des Prüfers tatsächlich möglich ist. Die Frage, wie das Risiko im einzelnen bei den zu treffenden Entscheidungen zu berücksichtigen ist, spielt dabei in der vorliegenden Untersuchung ebensowenig eine Rolle wie die Frage, ob reale Entscheidungen im Rahmen von Prüfungen tatsächlich nach dem ihnen innewohnenden Risiko getroffen werden. Für die einzelnen Prüfungsverfahren wird ausschließlich untersucht, inwieweit die Auswirkungen ihres Einsatzes im Hinblick auf das Risiko prognostiziert werden können. Dabei soll, soweit dies möglich ist, keine Beschränkung des Modells auf bestimmte Prüfungsverfahren erfolgen. Es soll vielmehr auch geprüft werden, inwieweit es über die bei mathematisch-statistischen Verfahren erfolgende Risikomessung mit Hilfe objektiver Wahrscheinlichkeiten hinaus möglich ist, die in den Ergebnissen anderer Prüfungsverfahren enthaltene Unschärfe bei der Risikomessung zu berücksichtigen. Soweit bei einzelnen Verfahren keine Risikomessung möglich ist, werden ausschließlich Ansätze für deren Weiterentwicklung gezeigt, die Weiterentwicklung der Verfahren selbst ist nicht Gegenstand der Arbeit.

B. Aufbau der Arbeit

Ausgangspunkt der vorliegenden Untersuchung ist eine gegenüber der Jahresabschlußprüfung verallgemeinerte Prüfungssituation, in der vom Prüfer die Beurteilung einer vorliegenden Menge von Daten (Prüfungsgesamtheit), die bestimmte Aspekte einer realen Gegebenheit (Unternehmung) beschreiben, verlangt wird. Die Prüfungsgesamtheit bildet dabei bestimmte reale, in der

---

9 Zu einem Überblick über empirische Untersuchungen zur Messung des Prüfungsrisikos vgl. *Strawser*, Information Processing, 1990.

Vergangenheit liegende Ereignisse ab und ist in nicht näher definierter Form Entscheidungsgrundlage für den Empfänger der Daten. Die Beurteilung durch den Prüfer erfolgt daraufhin, ob bei der Abbildung der realen Sachverhalte bestimmte, vorgegebene Regeln beachtet wurden, d. h., ob die vorgefundenen Daten ordnungsmäßig im Sinn der angesprochenen Regeln sind.

Im Ersten Teil der Arbeit wird das der Untersuchung zugrundeliegende Prüfungsmodell ausführlich dargestellt. Dazu wird zunächst ein Grundmodell der Prüfung, das noch keinerlei Unschärfe bei den Prüfungsinformationen berücksichtigt, entworfen, um die der Beurteilung der Prüfungsgesamtheit zugrundeliegende Konzeption aufzuzeigen. Anschließend werden die Ursachen unscharfer Prüfungsinformationen, die im angeführten Modell auftreten können und entweder aus dem verwendeten Prüfungsverfahren oder den beurteilten Sachverhalten resultieren, analysiert, allerdings noch ohne auf die vorhandenen Ansätze zur Abbildung der Unschärfe im Modell einzugehen. Abschließend wird dann die Situation des Prüfers in ein Entscheidungsmodell, das die Grundlage für die weitere Vorgehensweise bei der Risikomessung bildet, übertragen.

Um zu einer Theorie der Risikomessung für das vorgestellte Modell zu gelangen, wird im Zweiten Teil der Arbeit zunächst der Gegenstand der Messung, also der für das Prüfungsmodell relevante Risikobegriff abgeleitet. Dabei wird eine Abgrenzung insbesondere gegenüber dem umgangssprachlichen Begriff des Risikos und dem in der Entscheidungslogik verwendeten Risikobegriff vorgenommen. Einen Schwerpunkt der Arbeit bildet die im Anschluß vorgenommene Untersuchung der Eignung bekannter Unschärfemaße, namentlich des Wahrscheinlichkeitsmaßes, des Glaubwürdigkeitsmaßes und eines allgemeinen Unschärfemaßes, für die Modellierung der unterschiedlichen unscharfen Prüfungsinformationen. Dazu wird für die einzelnen Maße nicht nur die ihnen zugrundeliegende Axiomatik dargestellt, sondern auch geprüft, inwieweit die mögliche Interpretation der Maße mit dem hier unterstellten Prüfungsmodell übereinstimmt. Da die bisher im Schrifttum zum Prüfungswesen diskutierten Ansätze zur risikoorientierten Prüfung, soweit sie eine Quantifizierung des Prüfungsrisikos anstreben, ausschließlich auf die Risikomessung mit Hilfe von Wahrscheinlichkeiten abstellen, wird am Ende des Zweiten Teils eine Abgrenzung zwischen den bestehenden Ansätzen und dem vorliegenden Ansatz vorgenommen.

Im Dritten Teil der Arbeit werden zunächst die Möglichkeiten zur Messung von Unschärfe im Rahmen der einzelnen Prüfungsverfahren dargestellt. Dazu werden die Verfahren der ergebnisorientierten Prüfung und die indirekten Messungen unterschieden. Ziel der Untersuchung ist es, zunächst die Voraussetzungen aufzuzeigen, die bei den einzelnen Verfahren erfüllt sein müssen, damit die in den Ergebnissen enthaltene Unschärfe gemessen werden kann. Da letztere im vorliegenden Prüfungsmodell die Ursache für das dem Prüfungsurteil beizumessende Risiko ist, muß in einem zweiten Schritt untersucht werden, wie sich die Unschärfe der Ergebnisse auf das Prüfungsrisiko auswirkt. Abschließend wird dann gezeigt, welche Ansätze zur Aggregation der im Rahmen der einzelnen Prüfungsverfahren erfolgten Risikomessungen, die nicht notwendigerweise auf einem einheitlichen Unschärfemaß beruhen, zu einem Gesamtrisikomaß verfügbar sind.

Abschließend werden im letzten Teil die wichtigsten Ergebnisse der Arbeit im Hinblick auf die oben angeführten Fragestellungen zusammengefaßt und bewertet. Dabei wird insbesondere auch untersucht, inwieweit Möglichkeiten zur Objektivierbarkeit der Risikomessungen tatsächlich bestehen und unter welchen Voraussetzungen die risikoorientierte Auswahl von Prüfungshandlungen durchführbar ist.

Erster Teil

Prüfungsmodell bei unscharfer Information

A. Grundmodell der Prüfung

Ein Grundmodell für die hier betrachtete Prüfungssituation soll ausgehend vom meßtheoretischen Prüfungsansatz zunächst ohne Berücksichtigung von einzelnen Formen der Unschärfe abgeleitet werden. Eine Analyse der für einzelne Bereiche des Modells relevanten Unschärfen erfolgt dann im folgenden Kapitel. Unter Prüfung soll hier, in Übereinstimmung mit der wohl inzwischen h. M., die Durchführung eines Soll-Ist-Vergleichs mit anschließender Beurteilung der festgestellten Abweichungen durch prozeßunabhängige, also nicht bei der Erstellung des Istobjektes der Prüfung beteiligte Personen verstanden werden[1]. Der meßtheoretische Prüfungsansatz interpretiert darüber hinaus den Prozeß des Soll-Ist-Vergleichs als Abweichungsmessung[2].

Die Abweichungsmessung erfolgt dabei zunächst für einzelne Prüfungsobjekte. Da die Prüfungsobjekte regelmäßig mehrere prüfungsrelevante Merkmale haben, ergibt sich bei Anordnung der Prüfungsobjekte in Zeilen und Anordnung der einzelnen Merkmale in Spalten eine Matrix[3], die im Rahmen der Prüfungshandlungen durch Feststellung von Soll-Ist-Abweichungen bei den Merkmalen einzelner Prüfungsobjekte in eine Fehlermatrix überführt werden kann. Die in der Fehlermatrix ausgewiesenen Abweichungen werden dabei in Abhängigkeit vom Skalenniveau der einzelnen Merkmale ebenfalls auf unterschiedlichem Skalenniveau gemessen. Die hierzu notwendige Theorie zur Ableitung des Sollobjektes für den Vergleich und die meßtheoretischen Grundlagen sind inzwischen zumindest abgegrenzt und definiert[4].

---

1 Vgl. *Sieben/Bretzke*, Typologie, 1973, S. 627-629, die in diesem Fall von Prüfungen des Typs II und III spricht; *Selchert*, Begriff, 1978, S. 132; *Leffson*, Wirtschaftsprüfung, 1988, S. 13; *v. Wysocki*, Grundlagen, 1988, S. 1-3; *Buchner*, Prüfungswesen, 1997, S. 225; *Hömberg*, Prüfung, 1993, Sp. 3571.

2 Vgl. hierzu *v. Wysocki*, meßtheoretischer Ansatz, 1978.

3 Vgl. *v. Wysocki*, Prüfungstheorie, 1992, Sp. 1550.

4 Vgl. *v. Wysocki*, Grundlagen, 1988, S. 128-137.

Weitgehend ungelöst sind die theoretischen Probleme zur Ableitung eines Gesamturteils über die Prüfungsgesamtheit aus der Fehlermatrix[5]. Generell wird davon auszugehen sein, daß die Beurteilung nicht trivial in dem Sinn ist, daß schon bei einem einzigen Fehler oder erst bei vollständig fehlerhaften Daten die Ordnungsmäßigkeit der geprüften Daten abgelehnt wird. Auch eine Beschränkung der Urteilsbildung auf eine Merkmalspalte wird regelmäßig nicht möglich sein. Im Schrifttum werden Ansätze zur Gewichtung der einzelnen gemessenen Abweichungen diskutiert. Soweit hierzu die einzelnen Abweichungen in Fehlerklassen (z. B. leicht, schwer, total) für die einzelnen Merkmale eingeteilt werden, für die dann jeweils die einzuhaltenden Obergrenzen festzulegen sind[6], fehlt es insbesondere an Lösungsansätzen für die Bestimmung der Klasseneinteilung und der Ableitung von Zusammenhängen zwischen den Fehlerklassen. Darüber hinaus dürfte eine isolierte Betrachtung der Abweichungen bei einzelnen Merkmalen nur dann zulässig sein, wenn der Zusammenhang zwischen den Abweichungen bei den einzelnen Merkmalen zutreffend bei der Fehlerklassenbildung berücksichtigt wurde.

Im hier zu entwickelnden Modell soll zunächst eine Beurteilung der einzelnen Prüfungsobjekte erfolgen. In einem zweiten Schritt sollen dann die Einzelurteile zu einem Gesamturteil aggregiert werden. Es sollen daher für das Grundmodell zunächst die drei folgenden Bereiche genauer definiert werden:

- der Prüfungsgegenstand, an dessen Elementen die Soll-Ist-Abweichungen gemessen werden,
- die Ableitung von Einzelurteilen aus den gemessenen Soll-Ist-Abweichungen und
- die Verdichtung der Einzelurteile zu einem Gesamtergebnis der Prüfung.

I. Prüfungsgegenstand

Unter Prüfungsgegenstand wird im folgenden die Menge der Elemente verstanden, über die im Rahmen der Prüfung eine Aussage zu treffen ist[7]. Welche

---

[5] Vgl. *v. Wysocki*, Grundlagen, 1988, S. 9. Auch in den Verlautbarungen des Berufsstandes ist ausschließlich festgelegt, daß ein Gesamturteil nicht als Summe von Einzelurteilen entsteht, sondern einer Gewichtung bedarf, ohne daß diese näher erläutert wird. Vgl. *IDW*, FG 3/1988, 1989, S. 28.

[6] Vgl. *Kolarik*, Buchprüfung, 1964, S. 35-39; *Loitlsberger*, Fehlergewichtung, 1985, S. 196-198.

[7] Im Unterschied zum Prüfungsgegenstand soll der Begriff des Prüfungsobjekts weiter gefaßt werden. Als Prüfungsobjekte werden alle Objekte bezeichnet, die im Rahmen von Prüfungshandlungen, insbesondere auch bei indirekten Messungen untersucht werden, unabhängig davon, ob über sie ein Gesamturteil abzugeben ist oder nicht.

Aussage getroffen werden soll und wie sie sich aus den Aussagen über einzelne Elemente der Prüfungsgesamtheit zusammensetzt, ist Gegenstand späterer Ausführungen[8]. An dieser Stelle sollen die Annahmen und Definitionen für die Prüfungsgesamtheit sowie die zu untersuchenden Merkmale der Elemente in der Prüfungsgesamtheit erläutert werden.

1. Prüfungsgesamtheit

Die Prüfungsgesamtheit sei zunächst als eine endliche Menge von N Elementen in der vorliegenden Form definiert[9]:

$$(1.1) \qquad X = \{x_i \mid i = 1 \ldots N\} \text{ mit } X \subseteq U.$$

Für die einzelnen Elemente der Gesamtheit wird angenommen, daß sie Messungen der Merkmale realer Sachverhalte bzw. Ereignisse[10] darstellen. Die Menge *X* stellt somit ein Modell der Realität dar, bei dem die einzelnen Elemente als Tupel aus den interessierenden Merkmalen eines realen Ereignisses definiert sind. Die Menge aller möglichen Ereignisse oder Sachverhalte[11], deren Anzahl zwar groß, aber endlich sein soll, wird mit *U* bezeichnet. Für die Abbildung auf *X* ist somit zunächst festzulegen, ob es sich bei einem betrachteten Ereignis um einen Sachverhalt handelt, der gemessen werden soll, oder nicht. Somit existiert also eine Funktion[12] π_X:

$$(1.2) \qquad \pi_X : U \to \{0;1\}.$$

Für diese Ansatz- oder Zugehörigkeitsfunktion soll gelten:

$$(1.3) \qquad \pi_X(x) = \begin{cases} 0, \text{ wenn } x \notin X \\ 1, \text{ wenn } x \in X \end{cases} \text{ mit } x \in U.$$

Insoweit ist durch die Definition der Zugehörigkeitsfunktion π_X die Menge *X* eindeutig bestimmt. Die Definitionen in (1.1) und (1.3) sind isomorph[13].

---

8 Siehe unten Kapitel A. II. und A. III. in diesem Teil der Arbeit.

9 In der folgenden und allen weiteren formalen Darstellungen werden Mengen durch kursiv gesetzte Großbuchstaben bezeichnet.

10 Die Begriffe „Ereignis" und „Sachverhalt" werden synonym verwendet.

11 Diese Definition für *U* soll zunächst ausreichen. Sie wird im folgenden Teilabschnitt weiter ausgebaut.

12 Funktionen werden im weiteren entweder mit griechischen Kleinbuchstaben oder mit kursiv gesetzten Kleinbuchstaben bezeichnet.

13 Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 10.

Überträgt man das Modell auf die Jahresabschlußprüfung, so handelt es sich beim Prüfungsgegenstand um die von einer Unternehmung im Jahresabschluß offenzulegenden Daten. Da im Jahresabschluß aggregierte Daten aus der Buchführung wiedergegeben werden, ergeben sich zwei mögliche Interpretationen in bezug auf die Prüfungsgesamtheit. Zum einen kann X als die Menge der in der Buchführung gespeicherten elementaren Daten (Buchungssätze) verstanden werden. Die Funktion π_X legt dann entsprechend den Grundsätzen ordnungsmäßiger Buchführung fest, welche Geschäftsvorfälle bzw. Ereignisse im Rechnungswesen zu erfassen sind. Zum anderen kann X als die Menge der in der Bilanz aggregiert ausgewiesenen Sachverhalte, also Vermögensgegenstände, Schulden, Rechnungsabgrenzungs- und Sonderposten betrachtet werden. Die Funktion π_X entspricht dann den handelsrechtlichen Ansatzvorschriften für Vermögensgegenstände und Schulden. Die Gewinn- und Verlustrechnung ergibt sich aus der Veränderung der Bilanzpositionen und soll daher zunächst vernachlässigt werden.

Probleme für die Bestimmung der Funktion π_X ergeben sich für den Fall, daß die einzelnen Buchungssätze die Elemente der Grundgesamtheit sind, da die Entscheidung, ob ein bestimmtes Ereignis zu einer Buchung führt, auch davon abhängen kann, inwieweit andere in der Buchführung gespeicherte Ereignisse betroffen sind. So sind z. B. Preisveränderungen ausschließlich dann relevante Ereignisse, wenn sie am Bilanzstichtag noch gelten und wenn sie Vorräte betreffen, die am Bilanzstichtag noch vorhanden sind. Diese Probleme treten beim Abstellen auf die Vermögensgegenstände und Schulden als Elemente der Prüfungsgesamtheit im allgemeinen nicht auf, da deren Eigenschaften festlegen, ob sie dem Kaufmann nach § 242 HGB zuzurechnen sind oder nicht. Darüber hinaus müssen in beiden Fällen auch Prognosen über zukünftige Sachverhalte, soweit diese zu antizipieren sind, in X enthalten sein.

2. *Prüfungsmerkmale*

Bisher wurde für U angenommen, daß es alle realen Sachverhalte oder Ereignisse enthält, ohne explizit darauf einzugehen, in welcher Form dieses erfolgt. Wie oben angeführt wurde, sollen durch die Prüfungsgesamtheit ausschließlich bestimmte Aspekte der realen Welt abgebildet werden. Da X eine Teilmenge von U darstellt, ist die Auswahl der Merkmale, die zur Abbildung herangezogen werden, schon für die Definition von U zu berücksichtigen. Es muß demnach eine Funktion ω_X existieren, die angibt, mit welchen Merkmalsausprägungen ein Sachverhalt berücksichtigt werden soll. Auch dieser Vorgang

DOI https://doi.org/10.3790/978-3-428-48909-1

stellt eine Messung dar, wobei die Ausprägungen der für die Abbildung relevanten Merkmale von Sachverhalten nach den bezogen auf das hier unterstellte Modell externen Meßvorschriften ω_X ermittelt werden. Dem hier unterstellten meßtheoretischen Ansatz liegen also neben den im Rahmen der Prüfung durchgeführten Messungen auch bei der Generierung von X Messungen zugrunde.

Im folgenden soll angenommen werden, daß die Abbildung der einzelnen realen Sachverhalte durch ein Tupel von Merkmalsausprägungen $\langle v_1,\ldots,v_M\rangle$ $v_j \in V_j$ $(j=1,\ldots,M)$ erfolgt. Wird die Menge, die alle realen Sachverhalte ohne die Beschränkung auf die zur Abbildung verwendeten Merkmale enthält, mit U' bezeichnet, so gilt für ω_X und für U:

$$\omega_X : U' \to V_1 \times V_2 \times \ldots \times V_M, \tag{1.4}$$

$$U \subseteq V_1 \times V_2 \times \ldots \times V_M. \tag{1.5}$$

Da nach (1.1) gilt, daß $X \subseteq U$ ist, sind somit die einzelnen Elemente der Prüfungsgesamtheit X ebenfalls Merkmalstupel. Dabei handelt es sich zum einen um die Ausprägungen der Merkmale, welche die einzelnen x_i eindeutig bestimmen. Zum anderen müssen auch alle Merkmale in das Modell aufgenommen werden, die zur Entscheidung über den Ansatz eines Sachverhaltes in der Prüfungsgesamtheit, d. h. zur Bestimmung des Wertes von π_X benötigt werden. Es muß demnach gelten:

$$x_i \in V_1 \times V_2 \times \ldots \times V_M \text{ für } i=1,\ldots,N. \tag{1.6}$$

Die einzelnen Merkmale V_j, deren Ausprägungen v_j veröffentlicht bzw. in X gespeichert werden, sind zumindest nominal skaliert. Darüber hinaus soll ein Teil der Merkmale metrisch skaliert sein, so daß eine Bewertungsfunktion der folgenden Form für die einzelnen Elemente in X existiert:

$$b_X : X \to \mathbb{R}. \tag{1.7}$$

Eine generelle Existenz der Bewertungsfunktion für alle Elemente in U wird nicht gefordert, da insbesondere in dem hier betrachteten Anwendungsfall der Jahresabschlußprüfung die Bewertbarkeit der Elemente in X eine der notwendigen Bedingungen für den Ansatz der Elemente ist. Neben den oben angesprochenen Merkmalen, die zur eindeutigen Identifikation der einzelnen Elemente in U notwendig sind, und den zur Bestimmung von π_X notwendigen Merkmalen sollen darüber hinaus auch alle Merkmale im Modell aufgenommen werden, die zur Berechnung der Bewertungsfunktion b_X notwendig sind. Die Bewertungsfunktion stellt somit ausschließlich eine Transformation der für die Bewertung relevanten Merkmale eines Elementes dar.

Übertragen auf die bisher für die Jahresabschlußprüfung angeführten Interpretationen ergibt sich folgendes: Wird X als die Menge der in der Buchführung gespeicherten Buchungssätze betrachtet, so ergeben sich die bei den einzelnen Elementen zu speichernden Merkmale aus den Grundsätzen ordnungsmäßiger Buchführung. Die Funktion b_X legt dann den Betrag, mit dem eine Buchung zu speichern ist, fest. Wird der Inhalt der einzelnen Bilanzpositionen als Prüfungsgesamtheit betrachtet, so sind die für die einzelnen Vermögensgegenstände und Schulden relevanten Merkmale zu erfassen. Die Funktion b_X entspricht dann den geltenden Bewertungsnormen für Vermögensgegenstände und Schulden. Da die Einzelbewertbarkeit notwendige Voraussetzung sowohl für das Vorliegen eines Vermögensgegenstandes als auch für das Bestehen einer Schuld ist, ergeben sich auch hier hinsichtlich der Existenz der Bewertungsfunktion keine Probleme. Da die bei den einzelnen Elementen zu erfassenden Merkmale alle für die Bewertung notwendigen Sachverhalte umfassen müssen, sind demnach auch zukünftige Ereignisse, soweit sie die Bewertung einzelner Vermögensgegenstände oder Schulden betreffen, zu berücksichtigen.

Beiden Interpretationen ist gemeinsam, daß sie reale Prüfungsprobleme vereinfachen, da sie die Prüfung des Anhangs und des Lageberichts, die zumindest von prüfungspflichtigen Unternehmen aufgestellt werden müssen und gemäß § 316 Abs. 1 Satz 1 HGB Gegenstand der Jahresabschlußprüfung sind, nicht beinhalten, soweit dort qualitative Angaben erforderlich sind. Diese Angaben können nicht wie gefordert auf einer metrischen Skala gemessen werden, so daß die Funktion b_X für diese Sachverhalte nicht existiert. Ansatz- und Bewertungsfunktion werden unabhängig von der verwendeten Interpretation durch das im Unternehmen installierte Rechnungswesen wahrgenommen. Sie erlangen insoweit noch bei den indirekten Messungen im Rahmen der Prüfung, wie sie z. B. bei der Prüfung des datenerzeugenden Systems vorgenommen werden, Bedeutung.

Im Prüfungsansatz entspricht insoweit die erste der beiden Interpretationen eher einer progressiven, die zweite eher einer retrograden Vorgehensweise bei der Prüfung. Da im ersten Fall ausgehend vom einzelnen durch Buchungen in der Prüfungsgesamtheit abzubildenden Ereignis die Aggregation der Daten bis in den Jahresabschluß verfolgt wird, während bei der zweiten Interpretation ausgehend von einem Vermögensgegenstand oder einer Schuld die jeweils betroffenen Buchungen geprüft werden[14]. Da die Bilanz und damit die einzelnen

[14] Vgl. *Minz*, Prüfungsmethoden, 1960, S. 91; *v. Wysocki*, Grundlagen, 1988, S. 155f.; *Buchner*, Prüfungswesen, 1997, S. 235f.

in der Bilanz ausgewiesenen Vermögensgegenstände und Schulden aus den in der Buchführung gespeicherten Daten abgeleitet werden, kommen beide Ansätze dann zum gleichen Ergebnis, wenn die in der Buchführung gespeicherten Daten vollständig sind. Vorteile des progressiven Prüfungsansatzes ergeben sich insbesondere im Hinblick auf die Beurteilung der Vollständigkeit der Prüfungsgesamtheit sowie für die Beurteilung der Qualität der gespeicherten Daten. Die erste Interpretation ist darüber hinaus elementarer als die zweite. Dies führt dazu, daß zur Beschreibung der einzelnen Elemente in der Prüfungsgesamtheit insgesamt weniger Merkmale benötigt werden. Für die zweite Interpretation spricht, daß der Nachweis des tatsächlichen Vorhandenseins der Vermögensgegenstände und der Schulden in vielen Fällen, insbesondere durch die Prüfung der Bestandsaufnahmen, die gemäß § 240 HGB für die Vermögensgegenstände und Schulden durchzuführen sind, einfacher zu führen ist. Ein weiterer Vorteil des Ansatzes, der die in der Bilanz ausgewiesenen Vermögensgegenstände und Schulden als Elemente der Prüfungsgesamtheit betrachtet, liegt darin, daß die Funktionen π_X und b_X direkt handelsrechtlichen Ansatz- und Bewertungsregeln zugeordnet werden können. Eine eindeutige Präferenz für eine der beiden Betrachtungsweisen ergibt sich somit aus den bisherigen Ausführungen nicht.

II. Ableitung von Einzelurteilen

Bei dem hier vorgeschlagenen Modell soll die Beurteilung der Prüfungsgesamtheit in zwei Schritten durchgeführt werden. In einem ersten Schritt erfolgt die Messung und Beurteilung der bei einzelnen Elementen festgestellten Soll-Ist-Abweichungen. Der zweite Schritt beinhaltet dann die Aggregation der Einzelurteile zu einem Gesamturteil[15]. Im folgenden soll zunächst die Messung einzelner Soll-Ist-Abweichungen und deren Beurteilung formal beschrieben werden. Die Messung einer Abweichung setzt die Möglichkeit voraus, zu einem gegebenen Istobjekt x_i ein Sollobjekt y_i abzuleiten. Analog der Definition der Prüfungsgesamtheit X existiert dann die Menge aller Sollobjekte Y. Hierzu müssen im vorliegenden Modell grundsätzlich die folgenden Funktionen existieren:

$$\omega_Y : U' \to V_1 \times V_2 \times \ldots \times V_M, \tag{1.8}$$

$$\pi_Y : U \to \{0,1\}, \tag{1.9}$$

$$b_Y : Y \to \mathbb{R}. \tag{1.10}$$

[15] Dieser Abschnitt der Urteilsbildung ist Gegenstand des folgenden Kapitels A. III.

Falls sich alle drei genannten Funktionen von den entsprechenden Funktionen, die zur Erzeugung von X herangezogen wurden, unterscheiden, heißt dies, daß der Prüfer eigene Vorstellungen sowohl von der Abbildung der einzelnen realen Sachverhalte als auch von deren Zugehörigkeit zur Prüfungsgesamtheit und über deren Bewertung hat. Es soll im weiteren gelten, daß gleich indizierte Elemente in X und Y die gleichen Sachverhalte aus U' betreffen. Außerdem soll gelten:

$$(1.11) \qquad \forall x_i \in X, y_i \in Y : x_i = y_i \Rightarrow b_X(x_i) = b_Y(y_i).$$

Es besteht demnach Einigkeit über die Bewertung von Sachverhalten, wenn diese identisch hinsichtlich ihrer Merkmalsausprägungen sind, so daß wertmäßige Differenzen nur aus der unterschiedlichen Messung bzw. Festlegung von Merkmalsausprägungen bei den einzelnen Elementen, nicht aber aus unterschiedlichen, der Bewertung zugrundegelegten Bewertungsfunktionen entstehen. Im folgenden wird daher die Indizierung der Bewertungsfunktionen weggelassen.

Abweichungen zwischen einzelnen Tupeln $\langle x_i, y_i \rangle$ können schon für nominal skalierte Merkmale der einzelnen Elemente festgestellt werden[16]. Die Beurteilung der einzelnen Elemente soll allerdings über die bloße Feststellung von Abweichungen bei den einzelnen Merkmalen der Elemente hinausgehen. Im Rahmen der Beurteilung müssen die bei den Merkmalen eines Elementes gefundenen Abweichungen jeweils zu einem Urteil darüber aggregiert werden, ob diese ausreichen, das Element als fehlerhaft einzustufen. Eine solche allgemeine Funktion π_D, die angibt, ob ein Fehler vorliegt oder nicht, müßte, da auch

$$(1.12) \qquad Y \subseteq V_1 \times V_2 \times \ldots \times V_M$$

gilt, demnach die folgende Bedingung erfüllen:

$$(1.13) \qquad \pi_D : (V_1 \times V_2 \times \ldots \times V_M)^2 \rightarrow \{0,1\}.$$

Es ist davon auszugehen, daß die Beurteilung nicht trivial in dem Sinn ist, daß schon eine einzelne Abweichung zur Fehlerhaftigkeit des Elementes führt, oder ein fehlerhaftes Element erst dann unterstellt werden kann, wenn bei allen Merkmalen Abweichungen auftreten. Es muß demnach ein Gewicht für die Abweichungen bei den einzelnen Merkmalen gefunden werden, das eine Gesamtbeurteilung zuläßt. Unabhängig vom verwendeten Verfahren zur Festle-

---

16 Vgl. hierzu v. *Wysocki*, Prüfungstheorie, 1992, Sp. 1546.

gung einer solchen Funktion[17] wächst der Aufwand mit zunehmender Anzahl der in der Prüfungsgesamtheit bei den betrachteten Sachverhalten zu erfassenden Merkmale exponentiell.

In der vorliegenden Arbeit soll daher ein anderer Ansatz für die Bildung von Einzelurteilen verwendet werden. Da Ansatz und Bewertung der einzelnen Elemente zusammen bereits eine mögliche Gewichtung der einzelnen Merkmale darstellen, kann die Größe der Abweichung durch die Länge des Intervalls $|b(y_i)-b(x_i)|$ gemessen werden[18]. Die Möglichkeit, daß neben der Länge auch die Richtung der Abweichung von Interesse für die Beurteilung der Menge X ist, soll zunächst nicht weiter untersucht werden, da sich dadurch keine grundsätzlich neuen Probleme ergeben. Eine Funktion π_D für das hier diskutierte Grundmodell ist demnach folgendermaßen bestimmt:

$$(1.14) \qquad \pi_D(\langle x_i, y_i \rangle) = \begin{cases} 0, \text{ wenn } \dfrac{|b(y_i)-b(x_i)|}{\delta} \leq 1 \\ 1, \text{ wenn } \dfrac{|b(y_i)-b(x_i)|}{\delta} > 1 \end{cases} \text{ mit } \delta > 0.$$

Da die Bewertungsfunktion metrisch skaliert ist, beschränkt sich das Problem der Beurteilung darauf, das Vergleichsintervall δ geeignet zu wählen. Die Wahl von δ soll hier zunächst nicht weiter problematisiert werden. Es wird zunächst nur angenommen, daß nicht jede Abweichung zu einem Fehler führt, also $\delta > 0$ tatsächlich gilt, und daß δ eindeutig bestimmt werden kann. Denkbare Ansätze zur Bestimmung des Vergleichsintervalls wären, neben der hier verwendeten einfachsten Form der absoluten Festlegung von δ, die Festlegung als Funktion von x_i, y_i oder als Funktion des Gesamtwerts der Menge X definiert als $\sum_{i=1}^{N} b(x_i)$ bzw. dem entsprechend definierten Gesamtwert der Menge Y.

Darüber hinaus müssen die Sonderfälle, daß Elemente in X aber nicht in Y enthalten sind bzw. der umgekehrte Fall, definiert werden. Es soll gelten, daß nicht enthaltene Elemente den Wert Null haben. In der oben gezeigten formalen Definition der Beurteilungsfunktion ist demnach die Funktion $b(x_i)$ bzw. $b(y_i)$ jeweils zu ersetzen durch die folgendermaßen definierte Funktion b':

$$(1.15) \qquad \forall u \in U: b'(u) = \begin{cases} b_X(u), & \text{für } u \in X \\ b_Y(u), & \text{für } u \in Y \\ 0 & \text{sonst} \end{cases}.$$

---

[17] Vgl. *Kolarik*, Buchprüfung, 1964, S. 33-41; *Schettler*, Planung, 1971, S. 85; *Loitlsberger*, Fehlergewichtung, 1985.

[18] Vgl. auch *Schettler*, Planung, 1971, S. 85.

Aus den bisherigen Ausführungen ergibt sich für die Ordnungsmäßigkeit der in der Prüfungsgesamtheit vorliegenden Daten, daß diese zum einen vollständig und zum anderen mit dem richtigen Wert in X angesetzt sein müssen. Dabei ist Vollständigkeit in zwei Richtungen definiert, es dürfen weder Elemente in X fehlen, noch dürfen Elemente, die nicht angesetzt werden dürfen, in X enthalten sein. Die richtige Bewertung ist durch die extern vorgegebene Bewertungsfunktion b bzw. b' für gegebene Merkmalsausprägungen definiert.

Werden die angeführten Voraussetzungen auf die Prüfungssituation bei der Jahresabschlußprüfung übertragen, so läßt sich zunächst feststellen, daß zwar die Ansatz- und die Bewertungsfunktion kodifiziert sind, allerdings lassen die Formulierungen im Gesetz einige Spielräume für unterschiedliche Interpretationen zu, wie z. B. in § 253 Abs. 1 Satz 2 HGB, wonach Rückstellungen mit dem Wert, der „vernünftiger kaufmännischer Beurteilung" entspricht, anzusetzen sind. Insoweit stellt die Annahme eindeutiger Ansatz- und Bewertungsfunktionen eine Vereinfachung gegenüber der Realität dar.

Einigkeit besteht im Schrifttum zur Jahresabschlußprüfung dahingehend, daß nicht jede Abweichung zwischen Buch- und Prüfwert zu einem Fehler führt, und somit generell davon auszugehen ist, daß eine kritische Grenze δ existiert, die wesentliche Abweichungen und damit Fehler von unwesentlichen Abweichungen trennt[19]. Die Bestimmung eines konkreten Wertes für die Grenze im Einzelfall ist allerdings umstritten. Im allgemeinen wird die Festlegung der Wesentlichkeitsgrenze nicht wie im Modell als absolute Größe vorgegeben, sondern wird relativ zu verschiedenen Bezugsgrößen zu ermitteln sein[20]. Auch wenn die genannten Prämissen im Rahmen der Jahresabschlußprüfung somit nicht generell erfüllt sind, sollen sie für das Grundmodell zunächst beibehalten werden.

Für das hier verwendete Grundmodell werden demnach nur die Abweichungen bei Merkmalen, die sich auf den Wert der einzelnen Prüfungsobjekte auswirken, in die Beurteilung einbezogen. Das Gewicht der Abweichungen bei einzelnen Merkmalen wird dabei direkt gemessen. Soweit Abweichungen bei Merkmalen auftreten, die keinen direkten Einfluß auf den Wert des Prüfungsobjektes haben, wird davon ausgegangen, daß dadurch indirekte Aussagen über die Prüfungsgesamtheit getroffen werden können[21]. Betrachtet man die Ord-

---

19 Vgl. *Sperl*, Prüfungsplanung, 1978, S. 45f.; *v. Wysocki*, Grundlagen, 1988, S. 140; *Würtele*, Operationalisierung, 1989, S. 272.

20 Vgl. z. B. *Leffson*, Wesentlich, 1986, S. 443-445, *Ossadnik*, Materiality, 1995, S. 38f.

21 Einen ähnlichen Vorschlag macht schon *Kolarik*, dort werden allerdings die den Wert der

nungsmäßigkeitskriterien, die in § 239 Abs. 2 HGB für die Buchführung gefordert werden, so zielen diese nur zum Teil direkt auf die Korrektheit, d. h. Vollständigkeit und Richtigkeit, der aufgezeichneten Daten ab. Soweit eine geordnete Dokumentation gefordert wird, ist diese Voraussetzung für die Prüfbarkeit der Buchführung. Die zeitgerechte Erfassung der Daten wirkt sowohl auf die Vollständigkeit als auch auf die Richtigkeit der Prüfungsgesamtheit[22]. Somit wird die Beurteilung der Vollständigkeit der Prüfungsgesamtheit zu einem bestimmten Zeitpunkt auch durch eine Aussage über die Zeitgerechtigkeit der Erfassung beeinflußt. Die Richtigkeit der erfaßten Daten wird ebenfalls indirekt durch die zeitgerechte Erfassung beeinflußt, da unterstellt werden kann, daß die Dokumentation einzelner Sachverhalte vertrauenswürdiger ist, je zeitnäher die Erfassung im Buchführungssystem erfolgt. Da die Zeitspanne zwischen der Beurteilung der einzelnen Sachverhalte x_i durch das Unternehmen und der Prüfung der Sachverhalte durch den Prüfer, also die Ermittlung der zugehörigen Elemente y_i, lang sein kann, kommt der Beurteilung der Vertrauenswürdigkeit der Aufzeichnungen auch Bedeutung für die Vertrauenswürdigkeit der Beurteilung der Korrektheit der Prüfungsgesamtheit zu. Diese Zusammenhänge werden insbesondere im Rahmen der Integration indirekter Meßverfahren eine Rolle spielen. Das im Schrifttum diskutierte Problem, daß im Rahmen der Jahresabschlußprüfung oftmals nur nominal skalierte Merkmale vorliegen[23], reduziert sich damit auf solche Merkmale, die den Ansatz oder den Wert eines Sachverhaltes nicht beeinflussen und im Rahmen indirekter Messungen keine Bedeutung erlangen. Diese Fälle sollen hier vernachlässigt werden.

III. Ergebnis der Prüfung

Im zweiten Schritt der Beurteilung müssen die Einzelurteile zu einem Gesamtergebnis über die Prüfungsgesamtheit aggregiert werden. Es soll im folgenden die Vereinbarung gelten, daß die Ordnungsmäßigkeit der Prüfungsgesamtheit X, soweit eine logische Aussage darüber in formalen Ausdrücken gebraucht wird, symbolisch durch o[24], die Nichtordnungsmäßigkeit durch ~o ausgedrückt wird.

Elemente nicht beeinflussenden Abweichungen zu einem Urteil über die formale Ordnungsmäßigkeit aggregiert. Vgl. *Kolarik*, Buchprüfung, 1964, S. 35-39.

22 Vgl. *Leffson*, Grundsätze, 1987, S. 165f.; *Lang*, Grundsätze, 1986, S. 242.

23 Vgl. z. B. *v. Wysocki*, Prüfungstheorie, 1992, Sp. 1547.

24 In den weiteren Ausführungen werden einfache logische Aussagen mit Kleinbuchstaben bezeichnet.

Für die Beurteilung der Prüfungsgesamtheit kommen insbesondere zwei Kriterien in Frage, die Anzahl bzw. der Anteil der Fehler in der Prüfungsgesamtheit und ein Maß über die wertmäßige Abweichung zwischen X und Y. Die Menge D wird folgendermaßen definiert:

$$D=\{\langle x_i, y_i\rangle | x_i \in X \wedge y_i \in Y \wedge \pi_D(\langle x_i, y_i\rangle)=1\}. \tag{1.16}$$

Die Anzahl der festgestellten Fehler in der Prüfungsgesamtheit wäre demnach $|D|$. Der wegen seiner Unabhängigkeit vom Umfang der Prüfungsgesamtheit (N) aussagefähigere Fehleranteil θ ergibt sich durch:

$$\theta=\frac{|D|}{N}. \tag{1.17}$$

Für die Gesamtbeurteilung einer Prüfungsgesamtheit ist demnach eine Obergrenze für den Fehleranteil (θ^*) festzulegen, deren Überschreiten zur Ablehnung der Ordnungsmäßigkeit der vorgelegten Menge X führt. Es gilt[25]:

$$\theta>\theta^* \rightarrow \sim o. \tag{1.18}$$

Darüber hinaus soll auch bei einem Fehleranteil θ, der geringer oder gleich θ^* ist, die Ordnungsmäßigkeit von X abgelehnt werden, wenn die wertmäßige Abweichung $\Delta(X,Y)$ zwischen X und Y eine zu bestimmende Grenze Δ^* überschreitet. Diese wertmäßige Nebenbedingung ist notwendig, da durch die in (1.18) angeführte Bedingung alle Fehler unabhängig von der Höhe der ihnen zugrundeliegenden Abweichung gleichgewichtet werden[26]. Durch die zusätzliche Bedingung soll also auch für den Fall weniger, aber betragsmäßig hoher Fehler eine Ablehnung der Ordnungsmäßigkeit möglich werden. Fraglich ist, wie diese wertmäßige Abweichung zu messen ist. Dabei sind zum einen die Vergleichsobjekte genauer abzugrenzen, zum anderen ist zu prüfen, ob und welche Metrik zu verwenden ist .

Als Vergleichsobjekte kommen entweder die kompletten Mengen X und Y oder die in D enthaltenen Tupel $\langle x_i, y_i\rangle$ in Betracht. Für die zuletzt genannte Alternative spricht, daß dabei ausschließlich die fehlerhaften Elemente zur Beurteilung von X herangezogen werden. Der Vergleich von X mit Y wäre dann relevant, wenn die Fehlerdefinition und die einzuhaltende wertmäßige Nebenbedingung unabhängig voneinander sind. Läge dieser Fall vor, so müßte der Fall zugelassen werden, daß, obwohl ausschließlich Abweichungen in der Prü-

[25] In diesem und den folgenden logischen Ausdrücken wird die Symbolik von *Scholz* verwendet. Vgl. dazu *Bochenski/Menne*, Grundriß, 1983, S. 34.

[26] Vgl. *Loitlsberger*, Treuhand, 1966, S. 81.

DOI https://doi.org/10.3790/978-3-428-48909-1

fungsgesamtheit auftreten, die keine Fehler darstellen, die kumulierten Abweichungen von knapp unter δ bei jedem Element aus X zu einer Ablehnung der Ordnungsmäßigkeit der Prüfungsgesamtheit wegen einer Überschreitung des zulässigen Gesamtfehlers führen könnten. Dies kann nicht gewollt sein, so daß die Festlegung von δ in einer Weise erfolgen muß, die sicherstellt, daß eine Prüfungsgesamtheit, die keine Einzelfehler enthält, in jedem Fall auch die genannte wertmäßige Nebenbedingung erfüllt.

Ein Unterschied zwischen den beiden Vorgehensweisen bleibt demnach nur für den Fall bestehen, daß die aus D ermittelte Abweichung so dicht an der Obergrenze Δ^* liegt, daß eine Abweichung in Höhe von δ bei den nicht in D enthaltenen Tupeln zum Überschreiten der Obergrenze führt. Dies könnte insbesondere dann von Bedeutung sein, wenn aus Verfahrensgründen die wertmäßige Abweichung aus den Mengen X und Y einfacher bestimmt werden kann als aus D[27]. Eine genauere Analyse der aus den unterschiedlichen Vergleichsobjekten resultierenden Differenz bei der Abweichungsmessung kann allerdings erst dann erfolgen, wenn die zu verwendende Metrik spezifiziert ist.

Unabhängig vom konkreten Anwendungsfall werden an Metriken zumindest die folgenden Anforderungen gestellt[28], wobei X, Y und C jeweils Mengen gleicher Kardinalität bezeichnen:

(1.19) $$\Delta(X,Y) > 0 \text{ für } X \neq Y \text{ und } \Delta(X,X) = 0,$$

(1.20) $$\Delta(X,Y) = \Delta(Y,X),$$

(1.21) $$\Delta(X,Y) \leq \Delta(X,C) + \Delta(C,Y).$$

Ein allgemeines Distanzmaß für die hier zu betrachtenden Vergleichsobjekte X und Y, das diese Anforderungen nicht nur in trivialer Weise erfüllt, ist beispielsweise wie folgt definiert:

(1.22) $$\Delta(X,Y) = \sqrt[p]{\sum_{x_i \in X, y_i \in Y} \left| b'(y_i) - b'(x_i) \right|^p}\,.$$

Es liefert für $p=1$ die Summe der betragsmäßigen Abweichungen und für $p=2$ die euklidische Distanz in einem N-dimensionalen Raum. Fraglich ist nun,

27 Dies gilt z. B., wenn vor der Durchführung von Prüfungshandlungen eine genau abgegrenzte Prüfungsgesamtheit vorliegen muß, wie dies bei der Anwendung mathematisch-statistischer Stichprobenverfahren der Fall ist.

28 Vgl. *Suppes/Krantz/Luce/Tversky*, Measurement, Vol. II, 1989, S. 46.

DOI https://doi.org/10.3790/978-3-428-48909-1

inwieweit ein in der genannten Form konstruiertes Maß für die hier betrachtete Prüfungssituation geeignet ist.

Unproblematisch erscheint zunächst die erste Anforderung (1.19), die festlegt, daß der Abstand zwischen zwei identischen Mengen Null und zwischen verschiedenen Mengen größer Null ist. Auch die dritte Bedingung (1.21), die festlegt, daß eine Metrik gewährleisten muß, daß der kürzeste mögliche Abstand zwischen zwei Vergleichsobjekten gemessen wird, kann uneingeschränkt für die Prüfungssituation übernommen werden. Fraglich ist jedoch, ob im Rahmen der Prüfung die Symmetriebedingung (1.20) gilt. Diese impliziert für das Abstandsmaß und auch für die einzelnen Differenzen zwischen den zu vergleichenden Elementen der betrachteten Mengen, daß nur die Beträge der Differenzen, nicht aber die Vorzeichen relevant sind. Soweit also für Abweichungen unterschiedlicher Richtungen, wobei der Begriff der Richtung zunächst nicht weiter definiert werden soll, unterschiedliche Obergrenzen gesetzt werden sollen, scheitert die Verwendung eines symmetrischen Distanzmaßes. Diese Einschränkung, daß ausschließlich eine symmetrische Distanz für die wertmäßige Nebenbedingung verwendet werden kann, soll allerdings für das hier verwendete Grundmodell beibehalten werden.

Soweit in der Diskussion über die Verwendung von Stichprobenverfahren im Rahmen der Prüfung auf die Schätzung des gesamten oder durchschnittlichen Prüfwertes abgestellt wird[29], wird implizit anscheinend folgendes Distanzmaß zur Abstandsmessung verwendet:

$$\Delta(X,Y)=\left|\sum_{\mathrm{x_i}\in X} b'(\mathrm{x_i})-\sum_{\mathrm{y_i}\in Y} b'(\mathrm{y_i})\right|. \tag{1.23}$$

Dabei wird der Term $\sum_{\mathrm{y_i}\in Y} b'(\mathrm{y_i})$ durch das Stichprobenverfahren geschätzt[30]. Ein in dieser Weise definiertes Maß widerspricht der in (1.19) geforderten Bedingung, da sich einzelne Abweichungen ausgleichen können, und somit trotz unterschiedlicher Mengen X und Y ein Wert von Null für das Abstandsmaß resultieren kann. Es handelt sich demnach nicht um eine Metrik. Eine Verwendung der in (1.23) definierten Größe als wertmäßige Nebenbedingung käme ausschließlich dann in Betracht, wenn nur der Gesamtwert von X als Information relevant ist[31]. Als Vorteil ergäbe sich dann, daß die Ableitung

[29] Vgl. *v. Wysocki*, Grundlagen, 1988, S. 224; *IDW*, HFA 1/1988, 1988, S. 243.

[30] Vgl. *Mandl*, Anwendungsvoraussetzungen, 1984, S. 22.

[31] Dieser Fall trifft bezogen auf die einzelnen Bilanzpositionen zumindest für externe Jahresabschlußadressaten zu. Vgl. *Leffson*, Wirtschaftsprüfung, 1988, S. 312.

der Obergrenze Δ^* aus Materiality-Grenzen erfolgen könnte. Eine derartige Einschränkung des Anwendungsbereichs für das hier verwendete Grundmodell soll allerdings nicht erfolgen.

Ausgehend von den Überlegungen zur Wahl der Vergleichsobjekte und den Anmerkungen zur Wahl der Metrik für das Grundmodell soll im weiteren folgendes Distanzmaß für die wertmäßige Nebenbedingung gelten:

$$\Delta(X,Y)=\sum_{x_i\in X,y_i\in Y,(x_i,y_i)\in D}\left|b'(y_i)-b'(x_i)\right|. \tag{1.24}$$

Zumindest im Vergleich zur Verwendung des geometrischen Abstandes dürfte bei der Verwendung dieses Maßes die Festlegung der Obergrenze Δ^* einfacher möglich sein, da die Bedeutung der betragsmäßigen Abweichung im Rahmen des Rechnungswesens eher einzuschätzen ist als die des geometrischen Abstandes. Die Verwendung des geometrischen Abstandes als Distanzmaß hätte allerdings den Vorteil, daß große Abweichungen implizit stärker gewichtet würden als kleine Abweichungen.

Soweit im Rahmen der Verwendung von Stichprobenverfahren die Verwendung der Mengen X und Y als Vergleichsobjekte präferiert wird, gilt für die Differenz zwischen den beiden Maßgrößen:

$$0\leq\sum_{x_i\in X,y_i\in Y}\left|b'(y_i)-b'(x_i)\right|-\Delta(X,Y)\leq\left(\mathrm{N}-|D|\right)\delta. \tag{1.25}$$

Darüber hinaus muß bei dieser Vorgehensweise die Festlegung von δ so erfolgen, daß für den Fall einer fehlerfreien Prüfungsgesamtheit sichergestellt ist, daß die wertmäßige Nebenbedingung eingehalten wird. Demnach muß gelten:

$$\delta\leq\frac{\sum_{x_i\in X,y_i\in Y}\left|b'(y_i)-b'(x_i)\right|}{\mathrm{N}}. \tag{1.26}$$

Im Ergebnis soll für das hier verwendete Grundmodell gelten, daß eine Beurteilung der Prüfungsgesamtheit X als ordnungsmäßig nur dann erfolgt, wenn die beiden genannten Grenzwerte eingehalten werden, so daß gilt:

$$\theta\leq\theta^*\wedge\Delta(X,Y)\leq\Delta^*\leftrightarrow o. \tag{1.27}$$

Soweit demnach die benötigten Prüfungsinformationen, also insbesondere die Menge D, vollständig bekannt sind, ist ein Urteil über die Prüfungsgesamtheit eindeutig bestimmbar. Unterstellt man darüber hinaus, daß keine Meßfehler auftreten, kann im vorliegenden Grundmodell das Urteil über die Prüfungsgesamtheit mit Sicherheit abgeleitet werden. Somit kann in der bis zu diesem Punkt betrachteten Modellbildung kein Risiko für den Prüfer bestehen. Allerdings sind die getroffenen Annahmen in weiten Teilen sehr realitätsfern. Daher soll im nächsten Kapitel untersucht werden, welcher Form von Unschärfe die im einzelnen benötigten Prüfungsinformationen unterliegen.

B. Ursachen unscharfer Prüfungsinformationen

Als Prüfungsinformationen sollen die zur Beurteilung der Prüfungsgesamtheit benötigten Informationen, also Y und D sowie die daraus abgeleiteten Kriterien für die Urteilsfindung selbst, nämlich θ und $\Delta(X,Y)$, bezeichnet werden. Unter unscharfen Informationen sollen zunächst alle Informationen, die nicht eindeutig sind, verstanden werden. Mehrwertigkeit von Informationen führt dann nicht zu Unschärfe, wenn einzelnen Werte eindeutig angegeben werden können. Dabei beschränken sich die Ausführungen in diesem Kapitel weitestgehend auf die Klassifizierung von unterschiedlichen Formen der Unschärfe, ohne auf deren Berücksichtigung im Modell einzugehen, da die in Frage stehenden Unschärfemaße erst später diskutiert werden[32].

Hierzu sollen in den folgenden Abschnitten insbesondere drei Ursachen für die Unschärfe von Prüfungsinformationen untersucht werden. Zum ersten soll die Unschärfe, die durch den Umstand entsteht, daß die zur Beurteilung der Prüfungsgesamtheit zur Verfügung stehenden Informationen unvollständig sein können, untersucht werden. Zum zweiten kann es im Rahmen einer indirekten Vorgehensweise bei Abweichungsmessungen[33] zu unscharfen Ergebnissen kommen, wenn entweder der als Gesetzmäßigkeit unterstellte Zusammenhang zwischen Ersatztatbestand und Prüfungsgesamtheit nicht eindeutig ist, oder die relevanten Merkmalsausprägungen des Ersatztatbestandes nicht eindeutig festgestellt werden können. Schließlich kann Unschärfe drittens dadurch entstehen, daß die getroffenen Einzelurteile nicht, wie in Formel (1.14) unterstellt wird, eindeutig sind.

---

[32] Siehe unten Kapitel B. im Zweiten Teil der Arbeit.

[33] Unter indirektem Messen wird folgende Vorgehensweise im Rahmen des meßtheoretischen Prüfungsansatzes verstanden: Eine Regel, die den Zusammenhang zwischen den Ausprägungen des interessierenden Merkmales eines Prüfungsgegenstandes und den Merkmalsausprägungen eines Ersatztatbestandes beschreibt, ist bekannt oder zumindest allgemein akzeptiert (Majorprämisse). Durch Messen wird die Merkmalsausprägung des Ersatztatbestandes festgestellt (Minorprämisse). Die Ausprägung des interessierenden Merkmales beim Prüfungsgegenstand ergibt sich dann durch Konklusion. Nähere Ausführungen hierzu finden sich in Kapitel B. II. in diesem Teil der Arbeit.

I. Unvollständige Informationen

Die in Formel (1.27) angegeben Kriterien für die Ableitung des Prüfungsurteils können schon aus grundsätzlichen Überlegungen heraus nicht genau ermittelt werden. Da, wie aus (1.15) zu erkennen ist, die Vollständigkeit von X auch Gegenstand der Beurteilung ist, müßten, um eindeutige Prüfungsinformationen zu erhalten, alle Elemente in U' im Hinblick auf ihre Abbildung und Zugehörigkeit zu X geprüft werden. Selbst wenn, wie hier unterstellt werden soll, die Anzahl der Elemente in U' bei realen Prüfungen endlich ist, besteht diese Möglichkeit regelmäßig wegen der großen Anzahl von Sachverhalten und der begrenzten Prüfungszeit[34] nicht. Insoweit sind die zur Beurteilung der Vollständigkeit von X herangezogenen Prüfungsinformationen in realen Prüfungen regelmäßig unscharf im oben angesprochenen Sinn und zwar insbesondere im Hinblick auf die Bestimmung von Y und damit auch von θ und $\Delta(X,Y)$.

Darüber hinaus wird aber auch die Anzahl der in X enthaltenen Elemente regelmäßig so hoch sein, daß eine vollständige Prüfung in der vorgegebenen Zeit nicht möglich ist. Es wird daher allgemein für zulässig gehalten und aus Gründen der Wirtschaftlichkeit der Prüfungsdurchführung auch für notwendig erachtet, daß nur eine stichprobenweise Prüfung der Prüfungsgesamtheit erfolgt[35]. Deren Beurteilung erfolgt demnach, obwohl Y nur teilweise bekannt ist, d. h. aufgrund einer Stichprobe, für die gilt:

(1.28) $$Y' = \{y_i | i = 1, \ldots, n\} \text{ mit } n < N.$$

Die für die stichprobenweise Prüfung diskutierten Ansätze unterscheiden sich insbesondere durch das Auswahlverfahren der zu prüfenden Elemente und damit in der Art der Schlußfolgerungen, die aus den für die Stichprobe ermittelten Ergebnissen im Hinblick auf die Prüfungsgesamtheit gezogen werden können bzw. dürfen. Als Verfahren sollen im folgenden die zufällige und die bewußte Auswahl der Stichprobenelemente untersucht werden.

1. Zufallsstichproben

Unter Zufallsstichproben werden Stichproben verstanden, deren Elemente durch einen Zufallsprozeß aus der Grundgesamtheit ausgewählt werden. Dabei

---

[34] Die dem Prüfer zur Verfügung stehende Zeitdauer für die Prüfungsdurchführung wird bei der Jahresabschlußprüfung indirekt durch die Regelungen in den §§ 325, 326 HGB, die Fristen für die Offenlegung der geprüften Jahresabschlüsse festlegen, und regelmäßig in weitaus stärkerem Maße direkt durch die Auftragsbedingungen beschränkt.

[35] Vgl. *Leffson*, Wirtschaftsprüfung, 1988, S. 166; *IDW*, HFA 1/1988, 1988, S. 240f.

muß für jedes Element der Grundgesamtheit die Wahrscheinlichkeit, in die Stichprobe zu gelangen, größer als Null sein[36]. Die Wahrscheinlichkeit muß dabei nicht notwendigerweise für alle Elemente der Grundgesamtheit gleich groß sein. Als Zufallsprozeß kann entweder ein echter Zufallsprozeß, wie z. B. das Werfen einer Münze, das Mischen und Ziehen von Kugeln analog der Lottoziehung, die Verwendung von Zufallszahlentabellen, oder ein Pseudozufallsprozeß, wie er auf den meisten Computersystemen implementiert ist, verwendet werden. Werden Pseudozufallszahlen verwendet, so bestimmt die Qualität des zugrundeliegenden Pseudozufallsprozesses auch die aus der Stichprobe gewonnen Ergebnisse. Unschärfen, die aus solchen technischen Unzulänglichkeiten entstehen können, sollen im folgenden nicht berücksichtigt werden[37]. Es wird für die weiteren Ausführungen die Verwendung eines echten Zufallsprozesses unterstellt.

Da die Ergebnisse, die aus einer Zufallsstichprobe gewonnen werden, ohne zusätzliche Annahmen nur für die der Zufallsauswahl zugrundegelegten Grundgesamtheit gelten, ist diese geeignet abzugrenzen bzw. zu wählen. Für das hier diskutierte Grundmodell kommt als Grundgesamtheit entweder U' oder X in Frage. Für die Wahl von U' als Grundgesamtheit spricht, daß dann auch solche Fehler aufgedeckt würden, bei denen der Ansatz eines Ereignisses in X fälschlicherweise unterlassen wurde. D. h., eine Aussage über die Vollständigkeit von X wäre in der durch das Stichprobenverfahren gewonnenen Aussage enthalten. Dem steht entgegen, daß es sich bei U' um eine zwar endliche aber sehr große Menge handelt, deren eindeutige Abgrenzung bzw. Numerierung aller Elemente zur Auswahl der Stichprobenelemente nicht möglich ist. Als Grundgesamtheit für mathematisch-statistische Verfahren kann im hier behandelten Modell somit ausschließlich X verwendet werden, die Begriffe Grundgesamtheit und Prüfungsgesamtheit werden daher bis auf weiteres synonym verwendet. Die auf der Auswertung von Zufallsstichproben beruhenden Prüfungsverfahren liefern demnach keine Aussage über die Vollständigkeit von X, da in X fehlende Elemente aus U' nicht gefunden werden können.

Da es Zielsetzung sein muß, Aussagen über die für die Beurteilung relevanten Kriterien θ und $\Delta(X,Y)$ zu erhalten, kommt insbesondere die Verwendung von Parameterschätzern oder Parametertests für die Auswertung der Stichprobe

[36] Vgl. z. B. *Cochran*, Sampling, 1977, S. 18.

[37] Eine exakte Definition, wann eine Auswahlmethode zufällig ist, findet sich beispielsweise bei *Fisz*, Wahrscheinlichkeitsrechnung, 1978, S. 394. Da hier die auf mathematisch-statistischen Verfahren beruhenden Prüfungsmethoden nur in ihren Grundzügen erläutert werden, sollte die im Text angegebene verbale Definition der Zufallsstichprobe ausreichend sein.

in Betracht. Da es im vorliegenden Kapitel um die Gründe für die Unschärfe der resultierenden Aussagen geht, sind, soweit Parameterschätzverfahren eingesetzt werden sollen, ausschließlich Verfahren der Intervallschätzung von Interesse. Da die hier relevanten, den Schätz- und Testverfahren zugrundeliegenden Überlegungen unabhängig davon sind, welcher der interessierende Parameter der Grundgesamtheit ist, also im vorliegenden Fall entweder der Fehleranteil θ oder der Abstand $\Delta(X,Y)$, sollen die Grundzüge für die Verfahren anhand eines beliebigen Parameters ξ der Grundgesamtheit, für den hier immer $(\xi \geq 0)$ gilt[38], dargestellt werden.

Da die einzelnen Stichprobenelemente zufällig ausgewählt werden, handelt es sich bei der Merkmalsausprägung des Stichprobenelementes Z_i mit $(i=1,\ldots,n)$ um eine Zufallsvariable. Alle möglichen Zufallsstichproben vom Umfang n können demnach durch einen Zufallsvektor $(Z_1,Z_2,\ldots,Z_n)$ beschrieben werden. Eine Stichprobenfunktion $Z(Z_1,Z_2,\ldots,Z_n)$, die zur Schätzung des wahren Parameters ξ aus der Stichprobenrealisation verwendet wird, ist demnach ebenfalls eine Zufallsvariable. Für diese kann, bei genügend großem Stichprobenumfang n, zumindest eine Grenzverteilung $F_Z(x,\xi,n)$ angegeben werden[39]. Diese gibt für die Stichprobenfunktion Z die Wahrscheinlichkeit an, daß bei tatsächlich geltendem ξ und einem Stichprobenumfang von n ein Stichprobenwert, der kleiner oder gleich x ist, auftritt.

Da für einen bestimmten Wert des interessierenden Parameters in der Grundgesamtheit nur die Verteilung der möglichen Stichprobenrealisationen bekannt ist, kann auch der umgekehrte Schluß von einem Stichprobenergebnis auf den wahren, tatsächlich unbekannten Parameter der Grundgesamtheit nicht eindeutig sein. Bei bekannter oder als bekannt vorausgesetzter Grenzverteilung bestehen grundsätzlich zwei Alternativen[40] zur Auswertung eines realisierten Stichprobenergebnisses $z(z_1,z_2,\ldots,z_n)$. Einmal kann aus dem Stichprobenergebnis eine Intervallobergrenze z_o abgeleitet werden, so daß folgende Bedingung erfüllt wird[41]:

---

[38] Aus dieser Beschränkung für den Wertebereich folgt, daß ausschließlich einseitige Schätzintervalle und Tests relevant sind.

[39] Vgl. *Fisz*, Wahrscheinlichkeitsrechnung, 1978, S. 395.

[40] Die Verwendung des Theorems von *Bayes* soll hier nicht diskutiert werden, da dieses die wahren Parameter der Grundgesamtheit als Zufallsvariablen betrachtet und insoweit von den hier unterstellten Annahmen abweicht. Eine allgemeinere Diskussion der Anwendung des Theorems von *Bayes* im Rahmen des hier vorgestellten Ansatzes erfolgt im Zweiten und Dritten Teil der Arbeit.

[41] Vgl. *Fisz*, Wahrscheinlichkeitsrechnung, 1978, S. 571-575. Die Funktion *pr* gibt die Wahrscheinlichkeit (probability) eines Ereignisses an.

DOI https://doi.org/10.3790/978-3-428-48909-1

(1.29) $$pr(\xi \leq z_0) = 1 - \alpha .$$

Ein solches Intervall wird Konfidenzintervall genannt. $1-\alpha$ bezeichnet die Sicherheit der Schätzung, α die Irrtumswahrscheinlichkeit. Darüber hinaus gilt für die Genauigkeit der Schätzung, die mit steigendem e abnimmt:

(1.30) $$e = z_0 - z(z_1, z_2, \ldots, z_n).$$

Liegen die Prüfungsinformationen als Zufallsstichprobe vor, so kann demnach aus dem Stichprobenergebnis nicht auf einen eindeutigen Wert für θ und $\Delta(X,Y)$ geschlossen werden. Die Unschärfe der Informationen besteht darin, daß im Fall der Parameterschätzung ausschließlich ein Intervall, in dem sich der wahre Wert des betrachteten Parameters befindet, angegeben werden kann. Für eine bestimmte Zufallsstichprobe besteht zwischen der Sicherheit und der Genauigkeit einer Schätzung insoweit eine Abhängigkeit, als die Sicherheit nur durch eine Verminderung der Genauigkeit erhöht werden kann und umgekehrt. Wird der Fehleranteil θ in der Grundgesamtheit betrachtet und wurden in einer Stichprobe, deren Umfang n beträgt, m Fehler gefunden, kann als sichere Aussage $(1-\alpha=1)$ nur angegeben werden, daß der wahre Fehleranteil im Intervall $\left[{}^{m}\!/_{N}; {}^{N-n+m}\!/_{N}\right]$ liegt. Dabei ist die Ungenauigkeit insoweit maximal, da dieser Schluß auch ohne Kenntnis der Grenzverteilung möglich ist.

Die zweite Alternative das Stichprobenergebnis $z(z_1, z_2, \ldots, z_n)$ auszuwerten besteht darin, Hypothesen über den Wert des wahren Parameters der Grundgesamtheit zu testen. Dazu sollen die Nullhypothese $\xi = \xi_0$ und eine dazu konkurrierende Alternativhypothese $\xi = \xi_1$ betrachtet werden[42]. Bei Kenntnis der Grenzverteilung der Stichprobenfunktion für einen bestimmten Stichprobenumfang n soll nun der Stichprobenraum in genau zwei disjunkte Bereiche aufgeteilt werden[43]. Für den ersten Bereich soll gelten, daß die Wahrschein-

---

[42] Im folgenden soll für alle mathematisch-statistische Testverfahren die Existenz einer Alternativhypothese unterstellt werden, da keine wahrscheinlichkeitstheoretisch abgesicherte Aussage über eine einzelne statistische Hypothese erfolgen kann. Vgl. *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 145-149. Die Frage, ob die Alternativhypothese konkretisiert wird, und die Ableitung konkreter Werte für die beiden Hypothesen werden im Dritten Teil der Arbeit diskutiert. An dieser Stelle geht es nur darum, die Form der Unschärfe bei mathematisch-statistischen Verfahren aufzuzeigen.

[43] Es soll zunächst davon ausgegangen werden, daß der Test so konstruiert ist, daß er für alle möglichen Stichprobenergebnisse zu einer Entscheidung kommt, d. h., die beiden angesprochenen Bereiche sind tatsächlich überschneidungsfrei und ergeben zusammen den gesamten Stichprobenraum. Diese Bedingungen können im allgemeinen durch die geeignete Wahl des Stichprobenumfangs eingehalten werden. Die Probleme, die daraus resultieren, daß ein statistischer Test nicht entscheidet, werden im Dritten Teil der Arbeit diskutiert. An dieser Stelle geht es ausschließlich um die den Tests zugrundeliegende Konzeption.

lichkeit, daß ein Stichprobenergebnis aus diesem Bereich auftritt, bei gültiger Nullhypothese kleiner als eine vorzugebende, genügend kleine Wahrscheinlichkeit α ist, und gleichzeitig die Wahrscheinlichkeit, ein Stichprobenergebnis aus diesem Bereich zu erhalten, mindestens $1-\beta$ beträgt, wenn die Alternativhypothese den wahren Wert des betrachteten Parameters der Grundgesamtheit repräsentiert. Es soll gelten:

(1.31) $$0<\alpha<1-\beta.$$

Für den zweiten Bereich gilt dann, daß die Wahrscheinlichkeit, ein Stichprobenergebnis aus diesem Bereich zu erhalten, bei wahrer Nullhypothese $1-\alpha$ und bei wahrer Alternativhypothese β beträgt. Es soll nun die Vereinbarung gelten, daß die Nullhypothese dann verworfen wird, wenn ein Stichprobenergebnis aus dem zuerst genannten Bereich des Stichprobenraumes auftritt, da ansonsten akzeptiert werden müßte, daß, vorausgesetzt α ist genügend klein, ein unwahrscheinliches Ereignis beobachtet worden ist. Dabei bedeutet „verwerfen" der Nullhypothese nicht, daß diese falsch bzw. widerlegt ist, da die Wahrscheinlichkeit für einen Irrtum bei dieser Entscheidung α ist. Entsprechend soll die Annahme der Nullhypothese erfolgen, wenn ein Stichprobenergebnis aus dem zweiten Bereich auftritt. D. h., die Alternativhypothese wird im Hinblick auf die Nullhypothese verworfen. Auch im Hinblick auf die Annahme der Nullhypothese sind zwei Einschränkungen zu beachten. Zum einen bedeutet die Annahme nicht, daß die Nullhypothese wahr ist, da die Wahrscheinlichkeit, einen Fehler zu begehen, β beträgt. Zum anderen gilt die Annahme nur im Hinblick auf die gleichzeitig getestete Alternativhypothese. Bei Wahl eines anderen Wertes ξ_1 für die Alternativhypothese muß die getroffene Aussage nicht gelten[44].

Die Unschärfe des Testergebnisses kommt demnach zum einen im Abstand der beiden Hypothesen zum Ausdruck. Dies gilt unabhängig davon, welche der beiden Hypothesen verworfen wird. Zum anderen wird die Unschärfe durch die jeweilige Irrtumswahrscheinlichkeit bestimmt, also durch α, falls die Nullhypothese verworfen wird, bzw. durch β, wenn die Alternativhypothese verworfen und damit die Nullhypothese angenommen wird. Es gilt auch, wie bei den Schätzverfahren, daß für einen bestimmten Stichprobenumfang eine Verringerung des Abstandes der beiden Hypothesen zu einer Erhöhung der Irrtumswahrscheinlichkeiten α bzw. β führt. Zur Vereinfachung der folgenden Ausführungen soll die Genauigkeit der Aussage bei der Verwendung von Testverfah-

---

44 Vgl. *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 167; *Obermeier*, Abschlussprüfung, 1983, S. 28-34.

DOI https://doi.org/10.3790/978-3-428-48909-1

ren durch den Abstand der beiden Hypothesen ausgedrückt werden, wobei auch hier gilt, daß kleinere Abstände genauere Aussagen repräsentieren. Die Sicherheit der Aussage ergibt sich dann aus der jeweiligen Irrtumswahrscheinlichkeit und beträgt bei der Annahme der Nullhypothese $1-\beta$ und bei der Verwerfung der Nullhypothese, also der Annahme der Alternativhypothese, $1-\alpha$.

Wird bei der Verwendung von Zufallsstichproben einer der Werte der verwendeten Maßgrößen zur Beschreibung der Unschärfe, namentlich die Sicherheit oder die Genauigkeit festgelegt, kann, sobald eine echte Zufallsauswahl vorliegt, und der Umfang der Stichprobe hinreichend groß gewählt wird, so daß die Grenzverteilung für die verwendete Stichprobenfunktion existiert, der Wert der anderen Maßgröße formal aus den geltenden Gesetzen der mathematischen Statistik abgeleitet werden. Es handelt sich insoweit um objektivierte Größen, die, sobald die Verteilungsannahmen und statistischen Gesetze akzeptiert werden, intersubjektiv überprüfbar und damit weitgehend frei von subjektiven Einschätzungen sind. Dies gilt unabhängig davon, welches der beiden mathematisch-statistischen Verfahren eingesetzt wird, da sich die Verfahren ausschließlich durch die Fragestellung, nicht durch die zugrundeliegende Theorie unterscheiden.

Allen mathematisch-statistischen Stichprobenverfahren ist gemeinsam, daß sie keine konkreten Aussagen über die Merkmalsausprägungen der einzelnen nicht in die Stichprobe gelangten Elemente zulassen[45]. Aus diesem Grund müssen im Rahmen der Jahresabschlußprüfung bei der Anwendung von Zufallsstichprobenverfahren neben der Kenntnis der Grenzverteilung für die Stichprobenfunktion weitere Voraussetzungen erfüllt sein. Es wird insbesondere gefordert, daß nicht ein einzelner wesentlicher Fehler im Prüffeld vermutet wird, der allein zur Ablehnung der Ordnungsmäßigkeit führen würde. D. h., die Prüfungsgesamtheit sollte insoweit homogen sein, daß die einzelnen Fehler das Prüfungsurteil annähernd gleich beeinflussen. Darüber hinaus muß ein Häufigkeitsfall vorliegen, so daß die Fehler in ihrer Summe, nicht als einzelne Fehler, zur Beurteilung herangezogen werden können[46].

2. Stichproben mit bewußter Auswahl

Im Gegensatz zu den mathematisch-statistischen Verfahren existiert für die Verwendung von Stichproben mit bewußter Auswahl der Stichprobenelemente

---

45 Vgl. *Buchner*, Diskussion, 1983, S. 494f.

46 Vgl. *IDW*, HFA 1/1988, 1988, S. 243; *Leffson*, Wirtschaftsprüfung, 1988, S. 171f.

keine umfassende Theorie, die eine formale Ableitung von Größen zur Beschreibung der Unschärfe der resultierenden Prüfungsinformationen erlaubt. Da keine Zufallsauswahl stattfindet und somit bei gleichen Auswahlkriterien und gleicher Grundgesamtheit, falls von Erhebungsfehlern abstrahiert wird, immer nur ein bestimmtes Stichprobenergebnis ermittelt wird, handelt es sich bei der verwendeten „Stichprobenfunktion" nicht um eine Zufallsvariable. Unschärfe kann demnach ausschließlich durch Unwissen über die Eignung des verwendeten Auswahlkriteriums zur Aufdeckung der in der Grundgesamtheit enthaltenen Fehler entstehen. Allerdings unterliegen die Stichprobenverfahren mit bewußter Auswahl nicht der Beschränkung, daß die Grundgesamtheit vor Beginn der Auswahl vollständig abgegrenzt sein muß, so daß auch Sachverhalte, die nicht in X enthalten sind, in die Stichprobe gelangen können.

Da die Auswahl nicht willkürlich, sondern im Hinblick auf die in der Grundgesamtheit vermuteten Fehler erfolgen soll, werden im folgenden die einzelnen zur Anwendung kommenden Auswahlverfahren zunächst dargestellt und anschließend daraufhin untersucht, inwieweit die zugrundeliegenden Annahmen Schlußfolgerungen auf die zur Beurteilung der Prüfungsgesamtheit herangezogenen Kriterien θ und $\Delta(X,Y)$ zulassen. Im Rahmen der Stichprobenverfahren mit bewußter Auswahl werden im Schrifttum insbesondere die drei folgenden Auswahlkriterien diskutiert:

(1) Bei der Konzentrationsauswahl werden Elemente nach ihrer Bedeutung für das Prüfungsurteil ausgewählt[47]. Im vorliegenden Modell kann die unterschiedliche Bedeutung der Elemente nur aus unterschiedlich großen Abweichungen, die bei den einzelnen Elementen vorliegen, resultieren. Eine eindeutige Aussage über die Prüfungsgesamtheit ist bei diesem Auswahlkriterium ausschließlich dann möglich, wenn in der betrachteten Prüfungsgesamtheit nur Abweichungen nach unten auftreten können. D. h., es gilt:

$$\forall x_i \in X, \forall y_i \in Y : b(y_i) \leq b(x_i). \qquad (1.32)$$

Werden nun alle Elemente aus X, deren Wert die vorgegebene Fehlergrenze überschreitet $(b(x_i) > \delta)$, geprüft, so kann die Beurteilung der Prüfungsgesamtheit im Hinblick auf beide oben genannten Kriterien erfolgen. Dies setzt allerdings zum einen voraus, daß δ als absolute Größe oder in Relation zu einer vor der Prüfung bekannten Größe definiert ist, zum anderen

47 Vgl. *Leffson*, Wirtschaftsprüfung, 1988, S. 168; *v. Wysocki*, Grundlagen, 1988, S. 176; *IDW*, WP-Handbuch, 1996, P Rz. 85; *IDW*, HFA 1/1988, 1988, S. 241.

dürfte in realen Prüfungssituationen der bei dieser Vorgehensweise resultierende Auswahlsatz für die Stichprobe zu hoch sein, um eine ausreichende Verringerung der Prüfungsdauer zu gewährleisten.

Soll mit der Stichprobe ausschließlich die Einhaltung der wertmäßigen Nebenbedingung geprüft werden, so genügt es, wenn jeweils beginnend bei dem Element mit dem größten Wert solange Elemente in die Stichprobe aufgenommen werden, bis der Wert der nicht in die Stichprobe gelangten Elemente kleiner als Δ^* ist, oder eine Gesamtabweichung von mehr als Δ^* gefunden wird[48]. Diese Vorgehensweise eignet sich insbesondere dann, wenn die Werte in der Prüfungsgesamtheit eine hohe Konzentration aufweisen, d. h., wenige Elemente einen Großteil des Gesamtwertes der Prüfungsgesamtheit repräsentieren[49].

Im Rahmen der Jahresabschlußprüfung kommt dieses Auswahlverfahren insbesondere in Betracht, wenn die einzelnen Elemente der Prüfungsgesamtheit Vermögensgegenstände darstellen, die auf Überbewertung zu prüfen sind, und bei denen aus anderen Gründen Unterbewertungen für nicht relevant gehalten werden, oder wenn die Unterbewertung von Schulden bei Vernachlässigung von Überbewertungen geprüft werden soll. Für die Prüfung der Überbewertungen von Schulden bzw. der Unterbewertung von Vermögensgegenständen müssen andere Auswahlkriterien herangezogen werden. Auch hier soll zunächst gelten, daß Unterbewertungen vernachlässigt werden können. D. h.:

(1.33) $$\forall x_i \in X, \forall y_i \in Y : b(y_i) \geq b(x_i).$$

Die maximale Abweichung, die jetzt bei den einzelnen Elementen auftreten kann, ist jetzt nicht wie im vorher diskutierten Fall auf den Wert des Elements beschränkt, so daß kein Zusammenhang zwischen Bedeutung des Elements für das Prüfungsergebnis und seinem Wert besteht. Der Wert der einzelnen Elemente in X kann demnach nicht relevant für die Auswahl der Stichprobenelemente sein. Als Kriterium für die Auswahl von einzelnen Verbindlichkeiten werden hier insbesondere der Umfang der Geschäftsbeziehungen mit einzelnen Geschäftspartnern sowie die Auswahl wertmäßig umfangreicher Eventualverbindlichkeiten oder schwebender Geschäfte genannt[50], soweit Informationen darüber vorliegen. Eine ein-

48 Vgl. *Loitlsberger*, Treuhand, 1966, S. 91; *Egner*, Prüfungslehre, 1980, S. 129f.

49 Vgl. *Egner*, Prüfungslehre, 1980, S. 129f.

50 Vgl. *v. Wysocki*, Grundlagen, 1988, S. 176.

deutige Aussage ist bei der Verwendung dieser Auswahlkriterien nur dann möglich, wenn die für die Beurteilung relevanten Fehler tatsächlich von den genannten Kriterien abhängen.

(2) Bei der Auswahl typischer Fälle werden solche Elemente in die Stichprobe aufgenommen, die typisch für bestimmte Arbeitsabläufe sind[51]. Dieses Vorgehen für eine Prüfungsgesamtheit entspricht von der Intention her der Bildung von Äquivalenzklassen für die Elemente aus der Prüfungsgesamtheit. Dabei wird unterstellt, daß einige Elemente aus jeder Klasse repräsentativ für die gesamte Äquivalenzklasse sind[52]. Rückschlüsse auf die Prüfungsgesamtheit sind demnach dann möglich, wenn die Hypothese zutrifft, daß für die einzelnen Äquivalenzklassen gleiche Verarbeitungsregeln und insbesondere die gleichen Fehlerraten gelten.

(3) Als dritte Alternative für die bewußte Auswahl wird die dedektive Auswahl oder Auswahl nach der Fehlerwahrscheinlichkeit genannt[53]. Dabei wird ein Zusammenhang zwischen bestimmten Merkmalen einzelner Elemente und der Fehlerhaftigkeit dieser Elemente unterstellt. Die Informationen über diesen Zusammenhang stammen dabei aus anderen Prüfungshandlungen, wie der Prüfung des datenerzeugenden Systems. Der Unterschied zur Auswahl nach dem Konzentrationsprinzip besteht darin, daß das Merkmal für die Auswahl nicht der Wert des einzelnen Elements sein muß, sondern jedes bei den einzelnen Elementen gespeicherte Merkmal, das einen Rückschluß auf die Fehlerhaftigkeit des Elements zuläßt, sein kann. Kann für die Elemente, die nicht in die Stichprobe gelangen, ausgeschlossen werden, daß sie fehlerhaft sind, da alle Fehlerquellen bekannt sind, und damit alle fehlerhaften Elemente aufgrund des Auswahlverfahrens in die Stichprobe gelangt sein müssen, so kann die Beurteilung der Prüfungsgesamtheit eindeutig anhand der aus der Stichprobe ermittelten Ergebnisse erfolgen.

---

51 Vgl. *Loitlsberger*, Treuhand, 1966, S. 91; *Leffson*, Wirtschaftsprüfung, 1988, S. 168; *v. Wysocki*, Grundlagen, 1988, S. 178f.; *IDW*, HFA 1/1988, 1988, S. 242; *IDW*, WP-Handbuch, 1996, P Rz. 87f.

52 Vgl. *Göbel*, Prüfung, 1990, S. 157. Zur Auswahl von Elementen innerhalb der Äquivalenzklassen können die gleichen Kriterien wie bei der Auswahl von Testfällen herangezogen werden. Vgl. *Göbel*, Prüfung, 1990, S. 185f.

53 Vgl. *Loitlsberger*, Treuhand, 1966, S. 92; *Leffson*, Wirtschaftsprüfung, 1988, S. 168; *v. Wysocki*, Grundlagen, 1988, S. 177f.; *IDW*, WP-Handbuch, 1996, P Rz. 86; *IDW*, HFA 1/1988, 1988, S. 241.

DOI https://doi.org/10.3790/978-3-428-48909-1

Allen Auswahlverfahren ist gemeinsam, daß ohne die Berücksichtigung der ihnen zugrundeliegenden Hypothesen nur dann ein eindeutiges Ergebnis ermittelt wird, wenn in der Stichprobe genügend Fehler für die Ablehnung der Ordnungsmäßigkeit gefunden werden. Ist dies nicht der Fall, so können Rückschlüsse vom Stichprobenergebnis auf die für die Beurteilung relevanten Kriterien ausschließlich beim Vorliegen der oben genannten Voraussetzungen gezogen werden. Im Unterschied zu den mathematisch-statistischen Verfahren, die unscharfe Werte für die Kriterien zur Beurteilung der Prüfungsgesamtheit liefern, resultiert bei bewußter Auswahl der Stichprobenelemente die Unschärfe der Prüfungsinformationen daraus, daß die Gültigkeit der Voraussetzungen, die für das verwendete Auswahlverfahren unterstellt werden, nicht eindeutig bekannt ist[54]. Die Unschärfe der Prüfungsinformationen hängt dann davon ab, wie weit im konkreten Fall ein Maß für die Gültigkeit der Voraussetzungen gefunden werden kann, und welcher Art von Unschärfemaß dieses Maß entspricht. Dabei ist auch zu beachten, daß die Ableitung des Maßes für die Gültigkeit der unterstellten Hypothesen oftmals aus ähnlichen Gesetzmäßigkeiten oder Zusammenhängen abgeleitet wird, wie sie im folgenden Kapitel besprochen werden. Als Beispiel sei unterstellt, daß in Unternehmen, deren wirtschaftliche Lage schlecht ist, eher mit Überbewertungen als mit Unterbewertungen zu rechnen ist (Majorprämisse). Wird bei dem zu prüfenden Unternehmen festgestellt, daß seine wirtschaftliche Lage tatsächlich schlecht ist (Minorprämisse), so ist der Schluß erlaubt, daß eher mit Überbewertungen zu rechnen ist. Die Auswahl nach dem Konzentrationsprinzip stellt demnach ein adäquates Auswahlkriterium für die Prüfung von Positionen auf der Aktivseite der Bilanz dar.

II. Indirekte Messungen

Indirekte Messungen sind dadurch gekennzeichnet, daß Abweichungen nicht durch Vergleich des Prüfungsgegenstandes mit dem zugehörigen Sollobjekt, im vorliegenden Modell repräsentiert durch die Mengen X und Y, ermittelt wird, sondern durch die Prüfung eines Ersatztatbestandes[55]. Dazu wird ein Zusammenhang zwischen bestimmten Ausprägungen des Ersatztatbestandes und der zugehörigen Ausprägung des Prüfungsobjektes benötigt (Majorprämisse). Für diesen Zusammenhang sei zunächst unterstellt, daß er in Form einer Gesetzmäßigkeit vorliegt, die etwa folgende allgemeine Formulierung zuläßt: Wenn der

54 Vgl. *v. Wysocki*, Auswahl, 1986, S. 392.

55 Vgl. *Hagest*, Logik, 1975, S. 67; *v. Wysocki*, Grundlagen, 1988, S. 162f.

DOI https://doi.org/10.3790/978-3-428-48909-1

Ersatztatbestand die Ausprägung z aufweist, dann folgt daraus eine bestimmte Ausprägung der für die Beurteilung der Prüfungsgesamtheit X relevanten Kriterien $\theta=\theta'$ und $\Delta(X,Y)=\Delta'$. Wird nun durch die Prüfungshandlungen die Ausprägung z bei dem Ersatztatbestand festgestellt (Minorprämisse), so ist der Schluß erlaubt, daß auch $\theta=\theta'$ bzw. $\Delta(X,Y)=\Delta'$ vorliegt (Konklusion)[56]. Nach den für die Aussagenlogik geltenden Regeln läßt sich der Sachverhalt für den Zusammenhang zwischen Ersatztatbestand und Fehleranteil bzw. Fehlerausmaß in der Prüfungsgesamtheit formal folgendermaßen darstellen[57]:

$$(1.34) \qquad ((z \rightarrow \theta=\theta' \wedge \Delta(X,Y)=\Delta') \wedge z) \rightarrow \theta=\theta' \wedge \Delta(X,Y)=\Delta'.$$

Diese Form einer indirekten Messung ist insoweit idealtypisch, als dabei keinerlei Unschärfe auftritt. Diese Situation wird allerdings in den wenigsten Fällen vorliegen. Im folgenden soll nun für die wichtigsten indirekten Vorgehensweisen bei der Prüfung untersucht werden, an welchen Stellen Unschärfe auftritt und wie diese zu charakterisieren ist. Dazu werden die Prüfung des datenerzeugenden Systems und analytische Prüfungshandlungen unterschieden. Die zuletzt genannten Prüfungshandlungen stellen auf die Zusammenhänge zwischen die durch X abgebildete Realität und die Prüfungsgesamtheit selbst ab und verwenden Informationen über reale Sachverhalte zur Beurteilung der Plausibilität der in X offengelegten Daten. Die oben angesprochenen Stichprobenverfahren sollen nicht als indirekte Messungen interpretiert werden, auch wenn den Verfahren formal die gleiche Schlußform wie in (1.34) zugrundeliegt, da sie bestimmte Gesetzmäßigkeiten voraussetzen, namentlich die der mathematischen Statistik oder bestimmte für die Auswahl unterstellte Gegebenheiten[58]. Allerdings werden dort keine Ersatztatbestände untersucht, sondern einzelne Bereiche der Prüfungsgesamtheit direkt geprüft, so daß im hier vorliegenden Modell eine Trennung der beiden Vorgehensweisen erfolgen soll.

1. Prüfung des datenerzeugenden Systems

Die der Prüfung des datenerzeugenden Systems (Systemprüfung) im Rahmen der indirekten Vorgehensweise bei der Prüfung zugrundeliegende Gesetzmäßigkeit kann durch die folgenden Überlegungen abgeleitet werden. Oben wurde angenommen, daß die Prüfungsgesamtheit X durch die Funktionen π_X

[56] Vgl. *Egner*, Prüfungstheorie, 1992, Sp. 1570.

[57] Diese Schlußform entspricht dem modus ponendo ponens der klassischen Logik. Betrachtet man die Aussage p: „Der Ersatztatbestand hat die Ausprägung z", und q: „$Y=Y''$", so erhält man als Schlußform: $((p \rightarrow q) \wedge p) \rightarrow q$. Vgl. z. B. *Bochenski/Menne*, Grundriß, 1983, S. 46.

[58] Vgl. *v. Wysocki*, Grundlagen, 1988, S. 174.

DOI https://doi.org/10.3790/978-3-428-48909-1

und ω_X aus realen Sachverhalten bzw. Ereignissen abgeleitet wird. Es muß demnach ein datenverarbeitendes System existieren, das diese Funktionen ausführt und die Prüfungsgesamtheit erzeugt. Dieses System soll im weiteren als Verarbeitungssystem bezeichnet werden. Wenn unterstellt werden kann, daß die Funktionen π_X und ω_X deterministisch[59] vom Verarbeitungssystem ausgeführt werden, so können die in der Prüfungsgesamtheit X existierenden Fehler ausschließlich durch die Anwendung fehlerhafter Funktionsdefinitionen entstehen.

Trotzdem liegen mit der Kenntnis der tatsächlich ausgeführten Funktionen noch keine Informationen über die für die Beurteilung der Prüfungsgesamtheit benötigten Kriterien θ und $\Delta(X,Y)$ vor. Diese Informationen können erst dann gewonnen werden, wenn zusätzliche Kenntnisse über die tatsächlich verarbeiteten Daten vorhanden sind, da der zu beurteilende Systemoutput abhängig ist vom Zustand des Systems vor Verarbeitung der Daten, den durch das System ausgeführten Funktionen und dem zugeführten Input. Die für das Prüfungsurteil benötigten Aussagen lassen sich demnach eindeutig auch aus einer Prüfung des Verarbeitungssystems ableiten, wenn die folgenden Voraussetzungen gelten:

- Im Rahmen der Prüfungshandlungen werden die tatsächlich vom Verarbeitungssystem ausgeführten Funktionen π_X und ω_X bekannt.
- Das Verhalten des Verarbeitungssystems ist tatsächlich deterministisch.
- Alle vom Verarbeitungssystem verarbeiteten Sachverhalte bzw. Ereignisse sind bekannt.

Da insbesondere das Verhalten personeller Aufgabenträger nicht deterministisch ist, besteht regelmäßig, soweit Aufgaben delegiert werden, von Seiten der für die Ordnungsmäßigkeit der Daten Verantwortlichen ein Interesse daran, die Fehler, die durch das Verarbeitungssystem erzeugt werden, gering zu halten. Dazu wird in realen Systemen neben dem Verarbeitungssystem auch ein System von Kontrollfunktionen integriert sein[60]. Auch bei den Kontrollfunktionen handelt es sich um datenverarbeitende Funktionen, wenn auch mit anderer Intention. Somit bestehen für die Beurteilung des Kontrollsystems die gleichen Einschränkungen im Hinblick auf die Unschärfe der durch eine Systemprüfung gewonnenen Informationen, die auch für das Verarbeitungssystem gelten. Die folgenden Ausführungen betreffen demnach die Prüfung der Kom-

---

[59] Deterministisches Systemverhalten hinsichtlich der erfolgten funktionalen Beschreibung liegt dann vor, wenn bei bekanntem Anfangszustand des Systems das Systemverhalten eindeutig aus dem Systeminput abgeleitet werden kann. Vgl. *Niemeyer*, Systemtheorie, 1977, S. 88.

[60] Vgl. *Pougin*, Kontrollsystem, 1959, S. 49; *Adenauer*, Kontrollsystem, 1989, S. 21.

bination aus Verarbeitungs- und Kontrollsystem. Diese wird im weiteren auch kurz „Datenverarbeitungssystem“ genannt.

Bevor auf mögliche Unschärfen in den aus einer Systemprüfung gewonnenen Aussagen über die Prüfungsgesamtheit eingegangen wird, soll die logische Schlußform aus (1.34) für die bisher gezeigten Ansätze bei der Prüfung des Datenverarbeitungssystems angegeben werden. Hierzu wird eine bestimmte Aussage über die Umsetzung der Funktionen π_X und ω_X im Datenverarbeitungssystem mit z, eine Aussage über den tatsächlichen Systeminput mit i bezeichnet. Der logische Schluß aus den Ergebnissen der Systemprüfung auf den Fehleranteil θ und das Fehlerausmaß $\Delta(X,Y)$ in der Prüfungsgesamtheit stellt sich dann folgendermaßen dar:

$$(1.35) \qquad ((z \wedge i \rightarrow \theta = \theta' \wedge \Delta(X,Y) = \Delta') \wedge (z \wedge i)) \rightarrow \theta = \theta' \wedge \Delta(X,Y) = \Delta'.$$

Die für die Systemprüfung unterstellte Gesetzmäßigkeit wird dabei in (1.35) durch die erste Prämisse (Majorprämisse) $(z \wedge i \rightarrow \theta = \theta' \wedge \Delta(X,Y) = \Delta')$ der äußeren Implikation repräsentiert. Sie muß hier in Form einer formalen Implikation[61] vorliegen. Da aus unterschiedlichen Systemzuständen verschiedene Fehleranteile bzw. ein unterschiedlich großes Fehlerausmaß resultieren kann, muß für jede mögliche wahre Aussage z über die Funktionsweise des Datenverarbeitungssystems, welche die Systemprüfung liefern kann, in Verbindung mit dem tatsächlichen Systeminput eine wahre Aussage über die konkreten Werte der für die Beurteilung der Prüfungsgesamtheit benötigten Kriterien getroffen werden. Die Ergebnisse der Systemprüfung gehen in Formel (1.35) durch die zweite Prämisse (Minorprämisse) $(z \wedge i)$ der äußeren Implikation in die Schlußfolgerung ein. Da diese Aussagen gleichzeitig die Voraussetzungen der Majorprämisse darstellen, müssen, soweit man davon ausgeht, daß die unterstellte Gesetzmäßigkeit schon als Regel vorliegt, die für die Systemprüfung eingesetzten Methoden geeignet sein, diese Aussagen in der benötigten Form zu liefern[62].

Bei der Prüfung des datenerzeugenden Systems handelt es sich, wie bei der Prüfung der in der Prüfungsgesamtheit zusammengefaßten Sachverhalte, um einen Soll-Ist-Vergleich mit anschließender Beurteilung der festgestellten Abweichungen. Das Istobjekt der Prüfung ist das tatsächlich im Unternehmen eingesetzte Datenverarbeitungssystem. Seine Erfassung erfolgt regelmäßig nicht

---

[61] Zur Unterscheidung der Begriffe „materielle Implikation“, die keinerlei Zusammenhang zwischen Vorder- und Hintersatz verlangt, und „formale Implikation“, die eben diesen Zusammenhang zwischen Vorder- und Hintersatz unterstellt, vgl. *Tarski*, Einführung, 1971, S. 39.

[62] Vgl. *Wittmann*, Systemprüfung, 1980, S. 254.

direkt, sondern durch Prüfung der Umsetzung des vom Unternehmen konzipierten Systems (Transformationsprüfung). Für das Sollobjekt wird angenommen, daß es aus dem vom Unternehmen konzipierten System durch Beseitigung der festgestellten Schwachstellen und Fehler abgeleitet wird. Die Abweichung zwischen Soll- und Istsystem setzt sich dann aus den bei der Transformationsprüfung festgestellten Abweichungen und den bei der Ableitung des Sollsystems festgestellten Fehlern und Schwachstellen zusammen[63]. Im Rahmen der Beurteilung ist dann eine Aussage über das Verarbeitungssystem zu treffen, die für die oben genannte Schlußfolgerung verwendet werden kann.

Im Rahmen der Jahresabschlußprüfung wird keine der oben genannten Voraussetzungen, die eine eindeutige Ableitung der Beurteilungskriterien aus einer Systemprüfung erlauben, uneingeschränkt gelten. Unterscheidet man personelle Aufgabenträger einerseits und maschinelle Aufgabenträger (EDV-Anlagen) andererseits, so gilt, daß im ersten Fall zwar die ausgeführten Funktionen bekannt sind, da die Anweisungen zur Ausführung einzelner Funktionen auf einem relativ abstrakten Niveau erfolgen. Allerdings muß unterstellt werden, daß die Ausführung der Funktionen nicht deterministisch erfolgen wird, da beim menschlichen Handeln sowohl unbewußte (zufällige) als auch bewußte Abweichungen von der vorgeschriebenen Aufgabenstellung vorkommen. Demgegenüber ist bei der Ausführung der Funktionen durch EDV-Anlagen zwar weitestgehend sichergestellt, daß diese deterministisch erfolgt. Hier besteht das Problem darin, die tatsächlich ausgeführten Funktionen zu ermitteln[64].

Auch die Voraussetzung, daß alle zu verarbeitenden Sachverhalte bekannt sind, ist regelmäßig nicht uneingeschränkt erfüllt. Eine Erfassung des gesamten Systeminputs käme darüber hinaus im Aufwand der Prüfung aller Einzelsachverhalte sehr nahe, so daß dann die indirekte Vorgehensweise bei der Prüfung wenig Vorteile gegenüber der direkten Messung von Abweichungen hätte. Demnach kommen für Aussagen über die Inputdaten insbesondere Annahmen über die Verteilung bestimmter Attribute der einzelnen Sachverhalte bzw. Ereignisse in Betracht. Darüber hinaus ist jedoch zu beachten, daß es sich bei den Sachverhalten bzw. Ereignissen, deren Abbildung durch die Menge X erfolgt, um Vorgänge handelt, die bei den hier betrachteten Prüfungen regelmäßig in der Vergangenheit liegen. Der Prüfer benötigt die Ergebnisse der Systemprüfung somit auch zur Beurteilung der Vertrauenswürdigkeit der Daten, aus denen er die Menge der Sollobjekte Y ableitet, soweit keine Möglichkeit besteht,

63 Vgl. *Wittmann*, Systemprüfung, 1980, S. 86; *Knop*, Planung, 1983, S. 74f.; *Leffson*, Wirtschaftsprüfung, 1988, S. 231; *Adenauer*, Kontrollsystem, 1989, S. 65f.

64 Vgl. *Göbel*, Prüfung, 1990, S. 16.

bei der Erfassung der Sachverhalte anwesend zu sein oder durch den Vergleich mit von Externen gespeicherten Daten, die erfaßten Daten zu verifizieren.

Bei den im Rahmen der Prüfung des datenerzeugenden Systems unterstellten Gesetzmäßigkeiten und bei den zur Ableitung von Aussagen über die Werte für die Beurteilungskriterien θ und $\Delta(X,Y)$ verwendeten logischen Schlußformen treten nach den bisherigen Feststellungen verschiedene Formen von Unschärfe auf. Zunächst soll der als Majorprämisse unterstellte Zusammenhang zwischen bestimmten wahren Aussagen über die Funktionsweise des Datenverarbeitungssystems und Aussagen über die Wirksamkeit der darin integrierten Kontrollen einerseits und den daraus abzuleitenden Aussagen über konkrete Werte der für die Prüfungsgesamtheit relevanten Beurteilungskriterien andererseits betrachtet werden. Die Aussagen im Vordersatz der als Majorprämisse formulierten Implikation müssen grundsätzlich durch die Systemprüfung gewonnen werden können, da diese Aussagen dann auch als Minorprämisse benötigt werden. Dies bedeutet, daß die im Rahmen der hier betrachteten indirekten Vorgehensweise bei der Prüfung verwendbaren Gesetzmäßigkeiten durch die Aussagefähigkeit der zur Verfügung stehenden Systemprüfungsmethoden begrenzt wird. Da es sich beim Datenverarbeitungssystem regelmäßig um ein sehr komplexes System handeln wird, kann seine Funktionsweise schon aus diesem Grund nicht exakt beschrieben werden. Insoweit muß diese bestehende informationelle Unschärfe[65], die bei realen Rechnungswesensystemen wohl auch nicht durch sehr hohen Erfassungsaufwand behoben werden kann, bei der Formulierung der zugrundezulegenden Gesetzmäßigkeit berücksichtigt werden.

Zur Ableitung der Gesetzmäßigkeit zwischen Funktionsweise des Datenverarbeitungssystems und Fehlerhaftigkeit des Systemoutputs stehen grundsätzlich zwei Ansätze zur Verfügung. Zum einen kann versucht werden, das eingesetzte Datenverarbeitungssystem in einem Modell abzubilden, und durch Untersuchung des Modells Wahrscheinlichkeiten für Fehler und deren Ausmaß im Systemoutput abzuleiten[66]. Probleme bei diesen Ansätzen ergeben sich insbesondere daraus, daß entweder einfache Modelle, die stark von dem tatsächlichen Untersuchungsobjekt abstrahieren, oder hohe Aufwendungen für die Beschaf-

65 Unschärfe wird als „informationelle" bezeichnet, wenn grundsätzlich exakte Aussagen möglich sind, die zum Treffen der Aussage benötigten Informationen jedoch so umfangreich und/oder komplex sind, daß nur unscharfe Aussagen getroffen werden können. Demgegenüber liegt „intrinsische" Unschärfe vor, wenn Aussagen über menschliche Empfindungen, wie z.B. hoher Fehleranteil, ausgedrückt werden sollen. Vgl. *Rommelfanger*, Decision, 1994, S. 4; *Zimmermann*, Fuzzy Set Theory, 1991, S. 4.

66 Zur Darstellung und Analyse der Ansätze vgl. z. B. *Wittmann*, Systemprüfung, 1980, S. 97-131; *Sperl*, Prüfungsplanung, 1978, S. 130-156.

fung der benötigten Informationen[67] entstehen. Zum anderen besteht die Möglichkeit, eine Aussage über die Funktionsweise des betrachteten datenerzeugenden Systems als ordinale Bewertung seiner Zuverlässigkeit zu beschreiben. Der Wert für die Zuverlässigkeit wird dabei aus den bei der Systemprüfung festgestellten Abweichungen vom verwendeten Sollobjekt abgeleitet. Dieser Ansatz verlagert die beim vorhergehenden Ansatz auftretenden Probleme, die im wesentlichen aus den vorhandenen komplexen Systemzusammenhängen resultieren, insoweit auf den Prüfer, als dieser die ihm bei der Prüfung des datenerzeugenden Systems bekanntgewordenen Informationen zu einem einfachen, ordinal skalierten Wert aggregieren muß. Eine echte Verbesserung gegenüber der Situation beim ersten Ansatz ergibt sich nur dann, wenn Verfahren zur Verfügung stehen, die es ermöglichen, die Gesamtbeurteilung des datenerzeugenden Systems aus der Beurteilung einfacher Systemkomponenten abzuleiten, ohne einen zu großen Informationsbedarf zu erzeugen[68].

Während beim Vordersatz der als Majorprämisse verwendeten Implikation die enthaltene Unschärfe wegen der Komplexität realer Datenverarbeitungssysteme nicht beseitigt werden kann, könnten im Hintersatz prinzipiell exakte Aussagen gemacht werden, da hier einfache Wertzuweisungen erfolgen. Allerdings werden wegen der im Vordersatz enthaltenen informationellen Unschärfe ebenfalls keine exakten Werte in den aus der Funktionsweise und den Inputdaten abzuleitenden Aussagen angegeben werden, sondern verbale Umschreibungen, wie „geringer Fehleranteil" oder „hohes Fehlerausmaß", verwendet[69]. Die dabei auftretende Unschärfe resultiert aus der nicht exakt möglichen Zuordnung von Werten auf der zugrundeliegenden metrischen Skala, hier den reellen Zahlen zwischen Null und Eins für den Fehleranteil zur Angabe „gering" oder den positiven reellen Zahlen zur Angabe „hohes Fehlerausmaß". Dabei ist zu beachten, daß die Unschärfe nicht nur aus dem Umstand resultiert, daß mehrere unterschiedliche Werte für den Fehleranteil als „gering" bezeichnet werden können. In diesem Fall könnte eine exakte Formulierung durch Angabe eines scharfen Intervalls von Werten, die mit dem Ausdruck „gering" übereinstimmen, angegeben werden. Würde man sich auf ein solches exakt abgegrenztes Intervall einigen, z. B. „gering" entspricht dem Intervall $(0;0{,}01]$ auf der relevanten Skala, so wäre fraglich, warum nicht auch 0,011 oder 0,0101 noch als „geringe" Fehleranteile gelten können. Insoweit wird man regelmäßig Intervalle mit unscharfen Grenzen für die Zuordnung der Werte auf der zugrunde-

67 Vgl. *Wittmann*, Systemprüfung, 1980, S. 113.

68 Ansätze hierzu finden sich bei *Cooley/Hicks*, Fuzzy Set Approach, 1983, S. 317-334.

69 Vgl. *Hagest*, Logik, 1975, S. 108; *v. Wysocki*, Grundlagen, 1988, S. 166.

liegenden Skala zulassen müssen. Diese Art von Unschärfe wird, da sie im verwendeten Begriff selbst liegt, auch als intrinsische Unschärfe bezeichnet[70].

Eine dritte Form der Unschärfe für die unterstellte Majorprämisse kann sich letztendlich dadurch ergeben, daß man die Wahrheit der Implikation nicht in allen Fällen behauptet[71]. Gründe hierfür können sein, daß die ausgeführten Funktionen nicht die einzige Ursache für Fehler in der Prüfungsgesamtheit sind, oder daß nichtdeterministisches Verhalten bei einzelnen Systemkomponenten vorliegt. Als Beispiel für eine Majorprämisse, die alle angesprochenen Formen der Unschärfe enthält, ergibt sich dann:

Wenn das Datenverarbeitungssystem „zuverlässig" ist, dann ist der Fehleranteil im Prüffeld und die Distanz zwischen X und Y wahrscheinlich „gering".

Die angegebene Majorprämisse ist insofern noch unvollständig, als eine Aussage über den verarbeiteten Systeminput fehlt. Bei ordinaler Beschreibung der Funktionsweise des Datenverarbeitungssystems genügt eine Aussage über den Grad der Vollständigkeit der verarbeiteten Sachverhalte. Wird das Datenverarbeitungssystem in einem Modell abgebildet, müssen, soweit die tatsächlichen Inputdaten nicht zur Verfügung stehen, Verteilungsannahmen über den Systeminput gemacht werden[72]. Dabei gelten hinsichtlich der Unschärfe der Aussage, falls verbale Umschreibungen verwendet werden, die oben gemachten Anmerkungen zur Abbildung verbaler Aussagen auf metrischen Skalen entsprechend.

Insgesamt bleibt festzustellen, daß bedingt durch die zur Verfügung stehenden Prüfungsmethoden, durch die in realen Datenverarbeitungssystemen herrschende Komplexität, die eine genaue Beschreibung der Funktionalität des Datenverarbeitungssystems verhindert, und wegen der fehlenden Ableitbarkeit eines streng funktionalen Zusammenhangs zwischen der Funktionsbeschreibung des Datenverarbeitungssystems und den interessierenden Ausprägungen der Kriterien für die Beurteilung der Prüfungsgesamtheit keine Ableitung der beabsichtigten Konklusion nach den Regeln der klassischen Logik erfolgen kann[73]. Soll ein Maß für die der getroffenen Schlußfolgerung innewohnende

---

[70] Zum Begriff siehe oben Fn. 65 in diesem Teil der Arbeit.

[71] Vgl. *Hagest*, Logik, 1975, S. 70f. Im Gegensatz zu Hagest soll jedoch für das hier verwendete Modell keine Festlegung auf ein Wahrscheinlichkeitsmaß zur Messung der Unschärfe erfolgen.

[72] Vgl. *Sperl*, Prüfungsplanung, 1978, S. 134f.

[73] Vgl. *Hagest*, Logik, 1975, S. 108. Dieser geht allerdings in seinen Schlußfolgerungen von einem Wahrscheinlichkeitsmaß aus, ohne dies näher zu begründen.

Unschärfe gefunden werden, so muß es aus der Unschärfe der verwendeten Prämissen und Implikationen ableitbar sein.

Bisher wurde unterstellt, daß mit der Prüfung des Datenverarbeitungssystems auf die Ordnungsmäßigkeit der Prüfungsgesamtheit geschlossen werden sollte. Entsprechend wurde die Kenntnis eines Zusammenhangs zwischen einer bestimmten Zustandsbeschreibung des datenerzeugenden Systems und den für die Beurteilung der Prüfungsgesamtheit relevanten Kriterien gefordert. Im Rahmen des hier unterstellten Prüfungsmodells können allerdings auch andere Schlußfolgerungen aus den bei der Prüfung des Datenverarbeitungssystems gewonnenen Ergebnissen benötigt werden, die erst in der Kombination mit weiteren Prüfungsinformationen zu einem Urteil über die Prüfungsgesamtheit führen. Dies ist dann der Fall, wenn der Prüfer zur Bestimmung der Sollobjekte bei der Ableitung von Einzelurteilen in bestimmten Fällen auf die beim geprüften Unternehmen vorliegende Dokumentation von Ereignissen bzw. Sachverhalten angewiesen ist[74]. Die Vertrauenswürdigkeit seiner Prüfungsergebnisse hängt in diesen Fällen demnach auch von der Vertrauenswürdigkeit der verwendeten Dokumente ab. Darüber hinaus kann bei der Verwendung von Zufallsstichprobenverfahren die Vollständigkeit der Prüfungsgesamtheit im Hinblick auf fehlende Elemente nicht geprüft werden. Somit ist z. B. die Kombination von Zufallsstichprobenverfahren zur Beurteilung der in X abgebildeten Sachverhalte mit einer Aussage über die Vollständigkeit und die Vertrauenswürdigkeit der verwendeten Daten aus den Ergebnissen der Systemprüfung denkbar. Da die logische Struktur der verwendeten Gesetzmäßigkeiten für die hierzu benötigten Aussagen den oben angeführten Gesetzmäßigkeiten entspricht, können auch die gleichen Formen von Unschärfe auftreten.

2. Analytische Prüfungshandlungen

Wie im ersten Kapitel dieses Teils der Arbeit angesprochen, handelt es sich bei der Prüfungsgesamtheit X um die Abbildung realer Gegebenheiten aus U' nach bestimmten Regeln. Soweit demnach bestimmte Entwicklungen in der Realität Relevanz für die vorgenommene Abbildung haben, müssen sie auch in X ihren Niederschlag gefunden haben. Darüber hinaus kann, soweit zwischen den in der Menge X enthaltenen Daten bestimmte Relationen zwingend bestehen müssen, auch die Einhaltung dieser Relationen geprüft werden. Analytische Prüfungshandlungen dienen somit der Plausibilitätsbeurteilung, d. h. der Überprüfung, ob die in X enthaltenen Information im Verhältnis zueinander

74 Siehe oben in Kapitel A. II. auf S. 33.

und im Verhältnis zu den realen Entwicklungen plausibel erscheinen[75]. Im folgenden sollen daher die der Plausibilitätsbeurteilung zugrundeliegenden Schlußformen und die dabei möglicherweise auftretenden Formen von Unschärfe untersucht werden.

Betrachtet man zunächst wieder den Fall, der eine eindeutige Aussage über die Prüfungsgesamtheit erlaubt, d. h. eine ideale Plausibilitätsbeurteilung, so ist Ausgangspunkt für die Plausibilitätsbeurteilung eine bestehende Gesetzmäßigkeit, die es erlaubt, aus einer Aussage über die abzubildende Realität U' oder über bestimmte Relationen hinsichtlich von Teilmengen in X eine Kenngröße für X zu prognostizieren, die für ordnungsmäßige Prüfungsgesamtheiten gelten muß[76]. Eine solche wahre Aussage über die Realität oder bestimmte Relationen soll in den folgenden formalen Darstellungen mit u abgekürzt werden. Die zu prognostizierende Kenngröße sei funktional aus der Menge X ableitbar und wird im folgenden mit $k(X)$ bezeichnet. Da die Ordnungsmäßigkeit der Prüfungsgesamtheit o und die über die Realität oder bestimmte Relationen getroffene Aussage u im allgemeinen zwar eine notwendige, aber keine hinreichende Bedingung für die Prognose $k(X)=\mathrm{k}'$ darstellen, gilt:

(1.36) $$(\mathrm{o} \wedge \mathrm{u}) \rightarrow k(X)=\mathrm{k}'.$$

Im Rahmen der analytischen Prüfungshandlungen wird nun der tatsächliche Wert der Kenngröße $k(X)$ bestimmt. Trifft die Prognose nicht zu, gilt also $k(X) \neq \mathrm{k}'$, so kann folgender logischer Schluß gezogen werden[77]:

(1.37) $$((\mathrm{o} \wedge \mathrm{u}) \rightarrow k(X)=\mathrm{k}') \wedge k(X) \neq \mathrm{k}' \rightarrow \sim \mathrm{o} \vee \sim \mathrm{u}.$$

Dies bedeutet, daß entweder die Ordnungsmäßigkeit der Prüfungsgesamtheit nicht gegeben ist (~o), oder auch die über die Entwicklung der Realität getroffene Aussage nicht zutrifft (~u). Wird, eventuell nach weiteren Analysen, die Aussage über die Realität für wahr gehalten, folgt aus den Prüfungshandlungen die Nichtordnungsmäßigkeit der Prüfungsgesamtheit. Allerdings werden regelmäßig keine Aussagen über die tatsächlichen Ausprägungen der für die Beurteilung relevanten Kriterien θ und $\Delta(X,Y)$ möglich sein[78].

---

75 Vgl. *IDW*, FG 1/1988, 1989, S. 14; *Lück*, Jahresabschlußprüfung, 1993, S. 62.

76 Vgl. *Lachnit*, Globalabstimmung, 1992, Sp. 721f.; *Gärtner*, Prüfungshandlungen, 1994, S. 37f.; *AICPA*, SAS No. 56, 1988, AU Sec. 329 § 5.

77 Die Schlußform entspricht dem „modus tollendo tollens" der klassischen Logik. Vgl. *Bochenski/Menne*, Grundriß, 1983, S. 46.

78 Vgl. *v. Wysocki*, Grundlagen, 1988, S. 164.

Wird demgegenüber im Rahmen der analytischen Prüfungshandlungen festgestellt, daß die Prognose zutrifft, ist keine Aussage über die Ordnungsmäßigkeit der Prüfungsgesamtheit möglich[79], da bei wahrem Hintersatz die behauptete Implikation unabhängig vom Wahrheitswert des Vordersatzes wahr ist. Die Aussage, daß aus dem Zutreffen der Prognose auch die Ordnungsmäßigkeit der Prüfungsgesamtheit folgt, wäre nur dann möglich, wenn in Formel (1.36) eine Äquivalenzrelation bestünde. Diese ist allerdings immer dann nicht gegeben, wenn sich bei der Bestimmung von $k(X)$ einzelne Fehler in X ausgleichen können. Dies ist bei allen in der Prüfungspraxis eingesetzten Methoden der Verprobung oder Globalabstimmung, unabhängig davon, welche Prognosemethode eingesetzt wird, der Fall.

Ursachen für die Unschärfe bezüglich der durch analytische Prüfungshandlungen gewonnenen Informationen können zunächst aus der zugrundegelegten Gesetzmäßigkeit, der verwendeten Prognosemethode oder den Daten aus U', die für die Prognose benötigt werden, resultieren. Diese Komponenten sind in Formel (1.36) enthalten. Darüber hinaus kann auch die für eine Aussage notwendige Feststellung, daß der prognostizierte Wert für $k(X)$ nicht zutrifft, unscharf in dem Sinn sein, daß nicht jede Abweichung vom Prognosewert k' zur „Ungleichheit“ von $k(X)$ und k' in dem Sinn führt, und auch keine scharfen Intervallgrenzen für zulässige Abweichungen angegeben werden können.

Betrachtet man die unterschiedlichen, den analytischen Prüfungshandlungen zugrundegelegten Gesetzmäßigkeiten, so lassen sich insbesondere die folgenden Modelle unterscheiden:

- Modelle mit streng funktionaler Abhängigkeit der Daten innerhalb der Prüfungsgesamtheit, wie sie im Rahmen von formalen Globalabstimmungen und bei der formalen Verprobung verwendet werden[80].
  In diesen Fällen besteht weder im Hinblick auf die verwendete Gesetzmäßigkeit noch im Hinblick auf den prognostizierten Wert für $k(X)$ Unschärfe.
- Trendmodelle für die in der Prüfungsgesamtheit ausgewiesenen Daten, wobei sowohl konstante Modelle, bei denen geprüft wird, ob Abweichungen gegenüber dem Vorjahr oder gegenüber der Entwicklung anderer vergleichbarer Unternehmen durch Aussagen über die Realität U' gestützt werden[81],

[79] Vgl. *Blocher/Willingham*, Review, 1985, S. 10f.; *Gärtner*, Prüfungshandlungen, 1994, S. 66. A. A. *v. Wysocki*, Grundlagen, 1988, S. 163f.

[80] Vgl. *Lachnit*, Globalabstimmung, 1992, Sp. 720-725.

[81] Vgl. *Lachnit*, Globalabstimmung, 1992, Sp. 728-734; *Lück*, Jahresabschlußprüfung, 1993, S. 67.

als auch mathematisch-statistische Verfahren der Trendanalyse verwendet werden [82].

Unschärfe der Ergebnisse entsteht zum einen dadurch, daß die verwendeten Trendmodelle die Realität regelmäßig nicht exakt abbilden können, zum anderen gelten die Prognosen aus solchen Modellen uneingeschränkt nur dann, wenn die Bedingungen für den zu prognostizierenden Zeitraum den Bedingungen entsprechen, die den zur Prognose verwendeten Beobachtungsdaten zugrundeliegen. Darüber hinaus treten unabhängig von der eingesetzten Prognosemethode Vorhersagefehler auf. Bei der Verwendung mathematisch-statistischer Verfahren ergeben sich Zufallsschwankungen um die tatsächlichen Werte[83], bei anderen Verfahren der Trendanalyse können oftmals nur subjektive Einschätzungen für den möglichen Vorhersagefehler angegeben werden[84].

- Erklärungsmodelle, die auf den Zusammenhang zwischen erklärenden Größen für die betrachtete Kenngröße, die in der Realität beobachtet werden können, und den Werten für $k(X)$ selbst abstellen. Dabei werden sowohl Verfahren der Regressionsanalyse[85] als auch Verfahren, die auf der Verwendung bestimmter Input-Output-Relationen[86] beruhen, angewandt.

  Unschärfe entsteht hier ebenfalls durch die nicht exakte Abbildung der realen Zusammenhänge in den verwendeten Modellen. Für die ermittelten Prognosewerte gilt, wie bei der Zeitreihenanalyse, daß Vorhersagefehler auftreten können. Allerdings lassen sich bei Verwendung von statistischen Verfahren der Regressionsanalyse Konfidenzintervalle für den wahren zu prognostizierenden Wert angeben. Bei der Verwendung von Input-Output-Relationen entsteht Unschärfe bei der Regelformulierung insbesondere durch die vorzugebenden Intervalle für einzelne Relationen. Auch hier wird eine unscharfe Abgrenzung der Intervalle der Verwendung scharfer Intervallgrenzen vorzuziehen sein[87].

Soll aus dem Ergebnis der analytischen Prüfungshandlung auch für den Fall, daß die getroffene Prognose $k(X) = k'$ als zutreffend ermittelt wird, eine Aussa-

---

[82] Vgl. *Gärtner*, Prüfungshandlungen, 1994, S. 113-122.

[83] Vgl. *Schröder*, Zeitreihenprognose, 1994, S. 8.

[84] Vgl. *Blocher/Willingham*, Review, 1985, S. 66.

[85] Vgl. *Sperl*, Prüfungsplanung, 1978, S. 220-229; *Gärtner*, Prüfungshandlungen, 1994, S. 137-155.

[86] Vgl. *Lachnit*, Globalabstimmung, 1992, Sp. 725-728.

[87] Siehe oben die Diskussion zur Interpretation verbaler Angaben auf metrischen Skalen in Kapitel B. II. 1. auf S. 54.

ge über die Ordnungsmäßigkeit der Prüfungsgesamtheit getroffen werden, so wäre eine Gesetzmäßigkeit der folgenden Form notwendig:

$$(k(X) = \mathrm{k}' \wedge \mathrm{u}) \rightarrow \mathrm{o}. \tag{1.38}$$

Diese wird in der notierten strengen Form regelmäßig nicht vorliegen, da hier die oben zur Formel (1.36) vorgebrachte Argumentation, daß globale Kennzahlen sich ausgleichende Fehler nicht aufdecken können, analog gilt. Fraglich ist, ob nicht trotz fehlender funktionaler Abhängigkeit zwischen der Größe $k(X)$ und der Ordnungsmäßigkeit der Prüfungsgesamtheit eine gewisse Indikatorfunktion der Ausprägung $k(X) = \mathrm{k}'$ für die Ordnungsmäßigkeit vorliegen kann[88]. Diese könnte aus den bisherigen Erfahrungen des Prüfers entweder durch subjektive Schätzung oder durch statistische Auswertung im Rahmen von Diskriminanzanalysen ermittelt werden. Je nach verwendetem Verfahren erhält man dann entweder ein subjektives Vertrauensmaß oder ein Wahrscheinlichkeitsmaß für die Gültigkeit der in (1.38) zugrundegelegten Gesetzmäßigkeit.

III. Unscharfe Einzelurteile

Nachdem bisher untersucht wurde, inwieweit durch die vom Prüfer eingesetzten Prüfungsmethoden Unschärfe bei den Prüfungsinformationen entsteht, soll nun geprüft werden, ob in realen Prüfungssituationen die in (1.14) getroffene Annahme, daß für jede festgestellte Abweichung zwischen Soll- und Istobjekt eindeutig die Entscheidung fehlerhaft oder nicht fehlerhaft getroffen werden kann, in realen Prüfungssituationen, insbesondere im Rahmen der Jahresabschlußprüfung, zutrifft. Wie die in (1.14) gegebene Formel erkennen läßt, ergeben sich zwei Ansatzpunkte für mögliche Unschärfen. Zum einen kann die jeweils im Zähler stehende Abweichungsmessung unscharf sein, zum anderen kann Unschärfe bei der Bestimmung von δ auftreten. Beide Alternativen sollen für die Jahresabschlußprüfung in den folgenden Abschnitten getrennt untersucht werden.

1. Messung von Abweichungen

Die gemessene Abweichung zwischen den einzelnen Elementen in X und Y geht durch den Term $|b(\mathrm{y_i}) - b(\mathrm{x_i})|$ in Formel (1.14) ein. Da die vom prüfungspflichtigen Unternehmen erstellten Daten in X feststehen und eindeutig sind,

88 Vgl. *Lenz*, Urteilsbegründung, 1989, S. 1358.

kann Unschärfe der Messung nur aus dem vom Prüfer entwickelten Sollobjekt entstehen. Bei der Ableitung des Sollobjektes hat der Prüfer zwei unterschiedliche Fragen zu beantworten. Er muß zunächst prüfen, ob er den in x_i abgebildeten Sachverhalt auch in Y aufnimmt, d. h. die Frage des Ansatzes klären. Erst danach erfolgt die Prüfung der Ausprägungen für die einzelnen Merkmale, die den Wert des Elements bestimmen. Im folgenden sollen daher diese beiden Bereiche getrennt untersucht werden.

Für das hier verwendete Modell einer Prüfung soll davon ausgegangen werden, daß die Regeln für den Ansatz einzelner Elemente als generelle Bedingungen für die Ausprägungen einzelner Merkmale der betrachteten Elemente in U, die zum Ansatz des Elementes führen, angegeben sind. Unschärfe kann also nur dann entstehen, wenn die Messung der Ausprägung einzelner Merkmale, die für die Ansatzentscheidung relevant sind, unscharf ist. Für den hier im wesentlichen betrachteten Fall der Jahresabschlußprüfung sind bei Vermögensgegenständen die relevanten Merkmale der Charakter des Sachverhalts, die Einzelverwertbarkeit, der wirtschaftliche Wert und das wirtschaftliche Eigentum[89]. Die Menge der möglichen Ausprägungen für den Charakter des Sachverhalts ist {Sachen; Rechte; immaterielle Güter; Sonstige}, für die übrigen genannten Merkmale entstammen die möglichen Ausprägungen der Menge {ja; nein}, im Sinne von ist gegeben, ist nicht gegeben. Da es sich bei den Skalen, auf denen die Merkmale gemessen werden, um Nominalskalen handelt, ist für Unschärfe bei der Messung kein Raum. Ähnliches gilt für den Ansatz von Schulden, auch hier werden mit einer Ausnahme die Ausprägungen der relevanten Merkmale auf Nominalskalen gemessen. Die Ausnahme betrifft den Ansatz von Rückstellungen. Hier wird eine ausreichende Wahrscheinlichkeit für das Eintreten der Verpflichtung gefordert[90], so daß, wenn Wahrscheinlichkeit im Sinne der Wahrscheinlichkeitstheorie und nicht umgangssprachlich gemeint ist, das entsprechende Merkmal auf einer metrischen Skala gemessen wird. Da die angegebene Regel verbal formuliert ist, kann demnach Unschärfe hinsichtlich des Ansatzes entstehen. Diese Form der Unschärfe soll daher nicht ausgeschlossen werden.

Unschärfe bei der Beurteilung des Ansatzes einzelner Elemente in der Prüfungsgesamtheit entsteht somit ausschließlich dann, wenn nicht eindeutig entschieden werden kann, ob der Ansatz des betrachteten Elementes zulässig ist.

---

89 Vgl. *Kupsch*, BHR, 1986, § 246 HGB Rz. 22-28; *Adler/Düring/Schmaltz*, 1987, § 246 HGB Rz. 23-27; *Förschle/Kofahl*, Beck Bil.-Komm., 1995, § 247 Rz. 11-16.

90 Vgl. *Kupsch*, BHR, 1986, § 249 HGB Rz. 13; *Adler/Düring/Schmaltz*, 1987, § 249 HGB Rz. 61f.; *Clemm/Nonnenmacher*, Beck Bil.-Komm., 1995, § 249 Rz. 42f.

Die im Handelsrecht kodifizierten Ansatzwahlrechte führen demnach nicht zu einer Unschärfe bei der Abweichungsmessung, da der Prüfer die vom Unternehmen getroffene Entscheidung, das Wahlrecht auszuüben, oder nicht auszuüben, akzeptieren muß.

Da die Bewertungsfunktion metrisch skalierte Werte liefert, müssen ein oder mehrere der für die Bewertung relevanten Merkmale ebenfalls metrisch skaliert sein. Dadurch kann es immer dann, wenn verbale Bewertungsregeln, die sich auf metrisch skalierte Merkmale beziehen, existieren, zu Unschärfen kommen. Beispiele für solche durch verbale Umschreibungen entstehende Unschärfen im Rahmen der Jahresabschlußprüfung sind die Bewertung von Rückstellungen nach vernünftiger kaufmännischer Beurteilung (§ 253 Abs. 1 Satz 2 HGB), woraus sich eben im Falle der Bewertung von Pensionsrückstellungen keine exakte Abgrenzung des zu verwendenden Kalkulationszinssatzes ableiten läßt, oder die Definition der Herstellungskosten in § 255 Abs. 2 HGB, die für einige Kostenbestandteile das Wahlrecht zuläßt, angemessene Teile einzubeziehen. Auch hier läßt sich im allgemeinen keine exakte Grenze für die Angemessenheit von Kostenbestandteilen auf einer metrischen Skala ziehen.

Insgesamt ist festzustellen, daß die vom Prüfer zu fällenden Einzelurteile in einigen Fällen unscharf sein können, da die der Beurteilung zugrundeliegende Abweichungsmessung nicht in allen Fällen exakt sein kann[91]. Dabei ist festzuhalten, daß Unschärfe immer dann entstehen kann, wenn sich verbale Regelungen auf metrisch skalierte Merkmale der Prüfungsgesamtheit beziehen. Das Bestehen von Wahlrechten beim Ansatz oder der Bewertung selbst erzeugt noch keine Unschärfe im hier verwendeten Sinn, wenn die Intervallgrenzen für die zulässigen Werte exakt abgegrenzt werden können[92]. Die bei der Bestimmung der Einzelurteile über die Messung der Abweichung hinaus auftretende Unschärfe, die durch den in Formel (1.14) verwendeten Vergleichsmaßstab entsteht, ist Gegenstand des nächsten Abschnitts.

2. Definition des Fehlers

Wie schon oben bei der Ableitung von Einzelurteilen aus den festgestellten Abweichungen angesprochen wurde, ist die Ableitung einer eindeutigen Gren-

---

[91] So wohl auch *Gans*, Prüfungen, 1986, S. 382f. Allerdings soll der dort vertretenen Auffassung, soweit die Einwertigkeit als notwendige Voraussetzung für die Bezeichnung als Sollobjekt verlangt wird, nicht gefolgt werden.

[92] Vgl. auch *v. Wysocki*, Grundlagen, 1988, S. 140f. Dieser verwendet in diesen Fällen den Begriff „Toleranz“ im Gegensatz zum Begriff „Unschärfebereich“.

DOI https://doi.org/10.3790/978-3-428-48909-1

ze für δ, d. h. der Abweichung zwischen Soll- und Istobjekt, deren Überschreiten zu einem Fehler führt, selbst für den Fall, daß quantitative Abweichungen gemessen werden, umstritten[93]. Unabhängig davon, ob die Menge X insgesamt oder zu Teilgesamtheiten aggregiert dem Informationsempfänger zur Verfügung gestellt wird, ist δ aus dem Informationsbedürfnis des Informationsempfängers so abzuleiten, daß Abweichungen, die geringer als δ sind, entweder nicht wahrgenommen werden, oder zumindest die auf X beruhenden Entscheidungen nicht verändern[94]. Diese Form der deduktiven Ableitung der gewünschten Materiality-Grenze aus dem Informationsverarbeitungssystem des Informationsempfängers scheint jedoch nach bisherigem Stand der Theorie für die Jahresabschlußprüfung nicht durchführbar zu sein[95]. Allerdings ist zu vermuten, daß selbst dann, wenn die Probleme bei der Formulierung eines relevanten Entscheidungsmodels gelöst würden, eine informationelle Unschärfe für die abgeleitete Materiality-Grenze bestehen bliebe.

Eine weitere Möglichkeit der Gewinnung einer Materiality-Grenze wird darin gesehen, diese zu standardisieren, um zumindest eine objektivierte Größe zur Verfügung zu haben. Dabei ist regelmäßig eine relative Größe vorzugeben, da absoluten Größen, die unabhängig von der Größe der betrachteten Prüfungsgesamtheit festgelegt werden, nur geringe Aussagekraft zugeschrieben wird[96]. Für die Ermittlung eines solchen Standards kommen prinzipiell zwei Vorgehensweisen in Betracht. Einmal kann versucht werden, empirisch Grenzwerte zu ermitteln, die tatsächlich in der Prüfungspraxis oder von Informationsempfängern verwendet werden, und diese als Standard zu definieren. Die zweite Vorgehensweise bestünde darin, generelle Vorgaben durch verbindliche Regelungen festzuschreiben. Beiden Vorgehensweisen steht entgegen, daß sehr wahrscheinlich eine eindeutige Materiality-Grenze der Problemstellung nicht angemessen ist[97].

Werden trotzdem aus Gründen der Objektivierung von Prüfungen oder durch den Prüfer für die Prüfungsgesamtheit eine oder mehrere Bezugsgrößen zusammen mit einem Intervall festgelegt, das die Bestimmung der Materiality-Grenze relativ zu der oder den Bezugsgrößen ermöglicht, so ist die hier interes-

---

93 Siehe oben Kapitel A. II. auf S. 32.

94 Vgl. *Sperl*, Prüfungsplanung, 1978, S. 50.

95 Vgl. *Sperl*, Prüfungsplanung, 1978, S. 53-60; *Würtele*, Operationalisierung, 1989, S. 16-36; *Ossadnik*, Materiality, 1995, S. 36.

96 Vgl. *Buchner*, Prüfungswesen, 1997, S. 245; *Ossadnik*, Materiality, 1995, S. 38.

97 Vgl. *Leffson/Bönkhoff*, Materiality-Entscheidungen, 1982, S. 395f.; *Ballwieser*, Informationsökonomie, 1985, S. 63f.

sierende Frage, ob dieses Intervall notwendigerweise scharfe Grenzen haben muß. Wird z. B. eine Abweichung zwischen 0% und 2% vom Gesamtwert der Prüfungsgesamtheit für nicht wesentlich gehalten, so stellt sich die Frage, ob wirklich schon jede geringfügige Überschreitung der Intervallgrenze, die unterhalb der Fühlbarkeitsschwelle für den Informationsempfänger liegt, zu einem Fehler führen soll. Würde z. B. der Unterschied zwischen einer Abweichung von 2% und einer von 2,01% vom Informationsempfänger nicht wahrgenommen, ist fraglich, warum die eine Abweichung zu einem Fehler führen soll, die andere nicht. Würde aufgrund der vorhergehenden Argumentation das betrachtete Intervall erweitert, läßt sich der gleiche Einwand auf die neue Intervallgrenze anwenden, so daß sich, wenn Fühlbarkeitsschwellen beim Informationsempfänger bestehen, keine eindeutige Intervallgrenze finden läßt, die dem vorgebrachten Einwand entgeht.

Da sich die Argumentation natürlich auch auf die in (1.27) verwendeten Grenzwerte anwenden läßt, ergäbe sich als Resultat, daß keine Ablehnung der Ordnungsmäßigkeit erfolgen könnte, wenn der genannten Argumentation generell gefolgt würde. Insoweit muß zumindest für die Gesamtbeurteilung der Prüfungsgesamtheit an einer zu definierenden Stelle ein Schnitt erfolgen, der ordnungsmäßige und nicht ordnungsmäßige Prüfungsgesamtheiten trennt. Im Hinblick auf die oben vorgenommene Klassifizierung von Unschärfe[98] handelt es sich dabei um intrinsische Unschärfe, da die Materiality-Grenze grundsätzlich auf einer metrisch skalierten Skala angegeben werden könnte, für den Ausdruck „nicht wesentliche Abweichung" jedoch nur eine unscharfe Zuordnung von Werten auf der relevanten Skala möglich ist. Fraglich ist allerdings, ob ein solcher scharfer Schnitt auch schon bei den Einzelurteilen erfolgen sollte, oder ob dadurch nicht Informationen, die für die Beurteilung des Risikos relevant sind, verloren gehen. Dennoch soll, da diese Frage nicht ohne Kenntnis der für die Modellierung von unscharfen Grenzen für δ zur Verfügung stehenden Ansätze beantwortet werden kann, im folgenden zunächst davon ausgegangen werden, daß der Materiality-Grenzwert für die Abweichungen bei einzelnen Elementen der Prüfungsgesamtheit exakt angegeben werden kann.

C. Entscheidungsmodell für die Urteilsbildung

Unter einem Entscheidungsmodell soll im folgenden eine aus Entscheidungsfeld und Zielfunktion des Entscheidenden bestehende Abbildung eines

98 Siehe oben Fn. 65 in diesem Teil der Arbeit.

DOI https://doi.org/10.3790/978-3-428-48909-1

Entscheidungsproblems verstanden werden[99]. Da, wie anfangs dargestellt, kein Entscheidungskalkül für die Urteilsbildung abgeleitet werden soll, interessiert für die hier unterstellte Prüfungssituation nur das Entscheidungsfeld des Prüfers, um zu untersuchen, an welchen Stellen im Entscheidungsmodell das bei der Entscheidung für eine der Alternativen resultierende Risiko des Prüfers gemessen werden kann. Darüber hinaus soll untersucht werden, wie die Prüfungsinformationen und deren Unschärfe im Entscheidungsmodell zu berücksichtigen sind. Die folgenden Ausführungen betreffen zunächst das unterstellte Entscheidungsfeld. Im zweiten Abschnitt wird dann die Möglichkeit der Integration von Prüfungsinformationen als Ergebnis von Prüfungshandlungen in das Entscheidungsmodell untersucht.

I. Entscheidungsfeld

Das Entscheidungsfeld, welches der Entscheidung des Prüfers zugrundeliegt, besteht aus dem Aktionenraum A, der die Menge der dem Prüfer zu Verfügung stehenden alternativen Handlungsmöglichkeiten enthält, dem Zustandsraum S, der die Menge der für die Entscheidungssituation relevanten zukünftigen Umweltzustände beinhaltet, und den jeweiligen aus der Kombination von Umweltzustand und gewählter Alternative resultierenden Ergebnissen im Hinblick auf einzelne Zielgrößen des Prüfers[100].

Der Aktionenraum A umfaßt im vorliegenden Grundmodell der Prüfung zunächst ausschließlich die Aktion a_1, die der Beurteilung der Prüfungsgesamtheit als ordnungsmäßig entsprechen soll, und die Aktion a_2, die einer Beurteilung als nichtordnungsmäßig entspricht. Der Aktionenraum ist demnach endlich und, sieht man von der später diskutierten Möglichkeit der Informationsbeschaffung durch weitere Prüfungshandlungen ab, auch vollständig, d. h. es existieren keine weiteren unbekannten Aktionen. Die möglichen Aktionen entsprechen bei der Jahresabschlußprüfung der Erteilung eines uneingeschränkten Bestätigungsvermerks (a_1) gemäß § 322 Abs. 1 HGB und der Versagung des Bestätigungsvermerks (a_2) gemäß § 322 Abs. 3 HGB. Es erfolgt demnach insoweit eine Vereinfachung gegenüber realen Prüfungssituationen, als dem Prüfer dort die Möglichkeit offensteht, den Bestätigungsvermerk einzuschränken.

Der Zustandsraum S soll alle für die Entscheidung relevanten, d. h., das Ergebnis der einzelnen Aktionen beeinflussenden, externen Einflüsse, die nicht

[99] Vgl. *Laux*, Entscheidungstheorie I, 1991, S. 21.
[100] Vgl. *Schneeweiß*, Entscheidungskriterien, 1967, S. 13.

vom Entscheidenden beeinflußt werden können, enthalten. Eine Alternative, die Elemente des Zustandsraumes im vorliegenden Modell zu definieren, bestünde darin, die im Hinblick auf die Ordnungsmäßigkeit möglichen tatsächlichen Ausprägungen der Prüfungsgesamtheit X als relevante Umweltzustände zu betrachten. Demnach ergeben sich als mögliche Umweltzustände s_1: „X ist ordnungsmäßig" und s_2: „X ist nicht ordnungsmäßig". Der Ergebnisraum E ergibt sich durch Bewertung aller Elemente der Menge $A \times S$. Die vorläufige Ergebnismatrix ist in Abb. 1.1 dargestellt.

	s_1	s_2
a_1	e_{11}	e_{12}
a_2	e_{21}	e_{22}

Abb. 1.1: Vorläufige Ergebnismatrix für die Entscheidungssituation des Prüfers

Dabei betreffen die Ergebnisse e_{12} und e_{21} nach den oben getroffenen Festlegungen die Fälle, in denen der Prüfer ein Fehlurteil abgibt, die Ergebnisse e_{11} und e_{22} stellen die Fälle zutreffender Beurteilung der Prüfungsgesamtheit dar. Fraglich ist noch, wie sich im einzelnen die Beiträge zu den Zielgrößen der unterschiedlichen Ergebnisse ermitteln lassen. Im folgenden soll davon ausgegangen werden, daß der Prüfer Nutzenmaximierung unter der Nebenbedingung einer ausreichenden Urteilsqualität verfolgt[101]. Darüber hinaus soll unterstellt werden, daß die für die Nutzenbeiträge möglichen Ergebnisse als reelle Zahlen angegeben werden können. Es existiert demnach eine Funktion e, die folgende Abbildung erlaubt[102]:

(1.39) $$e: A \times S \to \mathbb{R}.$$

Dabei ist, soweit man eine vom Prüfungsergebnis abhängige Vergütung des Prüfers ausschließt, der zu ermittelnde Nutzen von dem Auftrag, in dessen Rahmen die Entscheidung über das Prüfungsergebnis zu fällen ist, unabhängig.

---

[101] Soweit sich im Schrifttum Ausführungen zur Zielsetzung von Abschlußprüfern finden, unterstellen diese Gewinnmaximierung bei Berücksichtigung der angegebenen Nebenbedingung. Vgl. *Leffson* et al., Sicherheit, 1969, S. 19; *v. Wysocki*, Wirtschaftlichkeit, 1992, Sp. 2176. Die Zielsetzung soll hier auf die Nutzenmaximierung ausgedehnt werden, damit neben der Einkommenserzielung auch Sachverhalte, wie Ansehensverluste durch bekanntwerdende Fehlurteile, in das Kalkül grundsätzlich einbezogen werden können.

[102] Für das hier verwendete Entscheidungsmodell sollen die Probleme der Abbildung von unterschiedlichen Zielbeiträgen auf einen metrisch skalierten Nutzenwert nicht weiter erörtert werden, da die konkreten Werte genau wie das Risiko selbst erst beim Entscheidungskalkül Berücksichtigung finden.

Nutzenbeiträge können sich also nur durch Wirkungen des Prüfungsergebnisses auf zukünftige Prüfungsaufträge beim gleichen oder anderen Unternehmen ergeben. Die Nebenbedingung einer ausreichenden Urteilsqualität wird dabei im Schrifttum regelmäßig im Zusammenhang mit einer ausreichend hohen Urteilssicherheit, oder umgekehrt formuliert mit ausreichend kleinem Risiko, verstanden[103]. Diese ist also insbesondere mit der hier verfolgten Zielsetzung der Risikomessung gekoppelt und soll im Rahmen der Beschreibung des relevanten Entscheidungsfeldes zunächst nicht weiter diskutiert werden.

Um die zukünftigen Ergebnisse für die unterschiedlichen Kombinationen aus Aktionen und Umweltzustand zu prognostizieren, genügt die Unterscheidung der beiden oben genannten Umweltzustände nicht, da insbesondere in den Fällen, in denen der Prüfer Fehlurteile abgibt, diese erst dann Wirkungen auf zukünftige Aufträge haben können, wenn sie als Fehlurteile dem Unternehmen oder anderen potentiellen Mandanten bekannt werden. Es existieren somit zwei externe vom Prüfer nicht gewußte oder beeinflußbare Faktoren, die auf seinen zukünftigen Nutzen einwirken, namentlich die Unsicherheit über die tatsächliche Ausprägung von X im Hinblick auf die Ordnungsmäßigkeit und die Unsicherheit darüber, ob die tatsächliche Ausprägung von X in der Zukunft bekannt wird. Es ergeben sich somit durch Permutation der relevanten Faktoren vier zu betrachtende Umweltzustände. Im weiteren soll folgende Benennung gelten: s_1 bezeichnet die Kombination der tatsächlichen Ordnungsmäßigkeit von X (o) und Bekanntwerden dieses Umstandes (b), s_2 die Kombination von o und ~b, s_3 die Kombination von ~o und b sowie s_4 die Kombination von ~o und ~b.

Risiko wird somit in dieser Entscheidungssituation des Prüfers durch die Ungewißheit darüber, welcher der Umweltzustände tatsächlich eintreten wird, modelliert. Wenn Risiko gemessen werden soll, müssen für die einzelnen Umweltzustände Maßzahlen gefunden werden, die den Grad angeben, mit dem der Eintritt des Umweltzustandes erwartet wird. In den klassischen Entscheidungsmodellen unter Risiko werden hier Wahrscheinlichkeitsmaße verwendet[104]. Inwieweit auch im vorliegenden Modell ein Wahrscheinlichkeitsmaß verwendet werden kann, soll zunächst offen bleiben. Diese Frage wird im Zweiten Teil der Arbeit diskutiert. Für das hier zugrundegelegte Maß p soll zunächst nur gelten:

(1.40) $$p:\mathbb{S}\rightarrow[0;1] \text{ mit } S\subseteq\mathbb{S}.$$

---

[103] Vgl. *Leffson*, Wirtschaftsprüfung, 1988, S. 122.

[104] Vgl. *Schneeweiß*, Entscheidungskriterien, 1967, S. 12; *Laux*, Entscheidungstheorie I, 1991, S. 25.

Im vorliegenden Fall muß die dem einzelnen Umweltzustand $s \in S$ zugeordnete Größe $p(s)$ aus entsprechenden Maßgrößen der Einflußfaktoren gebildet werden, durch deren Kombination der Umweltzustand s entsteht. Das Risiko wird demnach auch durch die Einschätzung des Prüfers über das Bekanntwerden des tatsächlichen Zustandes von X beeinflußt, da ausschließlich unter dieser Voraussetzung Fehlurteile aufgedeckt werden können. Daraus ergeben sich insbesondere zwei Probleme. Zum einen werden im Rahmen der Prüfungshandlungen ausschließlich Informationen über den Zustand der Prüfungsgesamtheit gewonnen. Daher können die Ergebnisse der Prüfungshandlungen auch ausschließlich zu Einschätzungen über den tatsächlichen Zustand von X herangezogen werden, Informationen über das Bekanntwerden des tatsächlichen Zustandes von X liefern die Prüfungshandlungen nicht. Außerdem kann vermutet werden, daß die Maßgrößen für das Bekanntwerden des tatsächlichen Zustandes von X nicht unabhängig von der gewählten Aktion des Prüfers sind. So wird im Schrifttum zur Jahresabschlußprüfung vielfach die Auffassung vertreten, daß sich bei einer Beurteilung der Prüfungsgesamtheit als nichtordnungsmäßig das geprüfte Unternehmen gegen die Beurteilung wehren wird, und somit ein eventuelles Fehlurteil aufgedeckt wird[105]. Es kann also bei Wahl der Aktion a_2 regelmäßig eher erwartet werden, daß der tatsächliche Zustand von X bekannt wird, als bei der Wahl von Aktion a_1. Soweit diese Umstände Berücksichtigung finden sollen, müßte ein spieltheoretischer Ansatz für das vorliegende Problem gewählt werden[106].

	s_1 $p(s_1)$	s_2 $p(s_2)$
a_1	e_{11}	e_{12}
a_2	e_{21}	e_{22}

Abb. 1.2: Entscheidungsfeld des Prüfers

Zum zweiten muß zumindest, wenn im Fall der Jahresabschlußprüfung das gemessene Risiko als Maßstab für die Urteilsqualität verwendet werden soll, das Risiko unabhängig von Überlegungen bleiben, die sich darauf beziehen, ob Fehler tatsächlich aufgedeckt werden. Würden Einschätzungen des Prüfers über das Bekanntwerden der tatsächlichen Ausprägungen von X in das Ent-

[105] Vgl. *Schildbach*, Stichprobenprüfung, 1992, Sp. 1848; *Stibi*, Prüfungsrisikomodell, 1995, S. 49.

[106] Vgl. hierzu *Loitlsberger*, Prüfungstheorie, 1992.

scheidungsmodell einbezogen, so würde allein der Umstand, daß die tatsächliche Ausprägung von X regelmäßig nicht bekannt wird, zu einer Reduktion des Risikos führen, ohne daß Prüfungshandlungen erfolgt sind[107]. Im weiteren soll daher, um eine Objektivierung dahingehend zu erreichen, daß das Risiko ausschließlich von den Ergebnissen der Prüfungshandlungen abhängig ist, vereinfachend angenommen werden, daß der tatsächliche Zustand der Prüfungsgesamtheit immer bekannt wird. Dies ist zugleich der Fall des maximalen Risikos für den Prüfer, solange er keine Abschätzung für die Unsicherheit bezüglich des Bekanntwerdens seiner Fehlurteile kennt. Das vorläufige Entscheidungsfeld unter Einbeziehung der Maßgrößen p für das Eintreten der einzelnen Umweltzustände ist in Abb. 1.2 dargestellt.

II. Prüfungshandlungen als Informationsbeschaffung

Wie schon im vorigen Teil angedeutet, entsprechen die vom Prüfer vorgenommenen Prüfungshandlungen im Rahmen des Entscheidungsmodells der Beschaffung von Informationen über das Entscheidungsproblem. Die Integration der Informationsbeschaffung kann grundsätzlich bei sämtlichen Komponenten des Entscheidungsfeldes ansetzen. Informationsbeschaffung kann demnach der Suche neuer Aktionen, also der Erweiterung von A, der Verbesserung der Ergebnisprognose für einzelne Elemente aus $A \times S$ sowie der Verbesserung der Prognosen über das Eintreten einzelner Umweltzustände dienen[108]. Im vorliegenden Fall ist nur die letzte Alternative von Interesse, da alle möglichen Aktionen bekannt sind und die Ergebnisse weitgehend aus der Untersuchung des Risikos ausgeschlossen werden sollen[109].

Bisher sind nur die beiden möglichen Urteile über die Prüfungsgesamtheit als Aktion im Aktionenraum enthalten. Tatsächlich steht dem Prüfer aber, sieht man von Zeitrestriktionen ab, die Möglichkeit offen, weitere Prüfungshandlungen durchzuführen, d. h., weitere Informationen über die tatsächliche Beschaffenheit der Prüfungsgesamtheit zu erhalten. Als mögliche Alternative ist demnach die Beschaffung von Informationen im Modell zu berücksichtigen. Da eine Entscheidung des Prüfers, weitere Informationen zu beschaffen, nicht sein Problem, die Prüfungsgesamtheit zu beurteilen, löst, steht er nach der Informationsbeschaffung im Hinblick auf die zur Verfügung stehenden Alternativen wieder vor der gleichen Entscheidungssituation. Das bisher einstufige Ent-

---

107 Vgl. *Stibi*, Prüfungsrisikomodell, 1995, S. 50.

108 Vgl. *Laux*, Entscheidungstheorie I, 1991, S. 281.

109 Siehe Fn. 102 in diesem Teil der Arbeit.

scheidungsmodell der Beurteilung der Prüfungsgesamtheit wird durch die Berücksichtigung der Informationsbeschaffung in ein mehrstufiges Entscheidungsmodell transformiert. Dabei enthält der Aktionenraum A keine einzelnen Handlungsalternativen mehr sondern Strategien, d. h. bedingte Handlungsanweisungen für die jeweiligen Entscheidungssituationen in allen Stufen des Entscheidungsmodells.

Bevor die neue Entscheidungssituation in Form eines Entscheidungsbaumes dargestellt werden soll, müssen noch einige Annahmen über die zu beschaffenden Informationen getroffen werden. Für die möglichen Ausprägungen i der beschafften Information gilt $i \in I$. Dabei soll die Menge der möglichen Ausprägungen der Informationen I vollständig und endlich sein. Außerdem muß, da die Informationen vereinbarungsgemäß die Einschätzungen über das Eintreten bestimmter Umweltzustände beeinflussen sollen, gelten:

$$\forall s \in S, \forall i \in I: p(s) \neq p(s|i). \tag{1.41}$$

In Abb. 1.3 ist das jetzt mehrstufige Entscheidungsmodell in Form eines Entscheidungsbaumes[110] dargestellt. Dabei symbolisieren die Quadrate Entscheidungssituationen oder Endergebnisse, die Kreise Zufallsknoten, durch die die Unsicherheit über die zukünftige Umweltentwicklung veranschaulicht wird. Dabei zählen zu den zukünftigen Umweltentwicklungen jetzt auch die a priori nicht bekannten Ergebnisse von Prüfungshandlungen. Unterschiede hinsichtlich der Entscheidungssituation vor und nach der Informationsbeschaffung ergeben sich sowohl bei den Ergebnissen, als auch bei den Einschätzungen $p(s \in S)$ über die zukünftigen Umweltzustände. So unterscheiden sich beispielsweise die Ergebnisse e_{11}^1 und e_{11}^2 durch die Kosten der Informationsbeschaffung, die bei e_{11}^2 berücksichtigt werden müssen. Dafür sollte sich die Prognose über den Eintritt der Umweltzustände aus S in t_1 gegenüber der Situation in t_0 nach Erhalt der Informationen verbessert haben.

Falls es sich bei $p(s)$ um ein Wahrscheinlichkeitsmaß handelt, können die bedingten Wahrscheinlichkeiten $p(s|i)$, die nach der Informationsbeschaffung zur Einschätzung der möglichen Umweltzustände verwendet werden, aus $p(s)$ und den die Informationsqualität beschreibenden bedingten Wahrscheinlichkeiten $p(i|s)$ nach dem Theorem von *Bayes* bestimmt werden[111]. Im weiteren Verlauf der Arbeit muß demnach neben der Festlegung des zu verwendenden

---

[110] Vgl. *Laux*, Entscheidungstheorie I, 1991, S. 257f.

[111] Vgl. *Laux*, Entscheidungstheorie I, 1991, S. 285f. Auf die formale Darstellung des Theorems von *Bayes* soll an dieser Stelle nicht weiter eingegangen werden. Seine Eignung wird ausführlich im Zweiten Teil der Arbeit diskutiert.

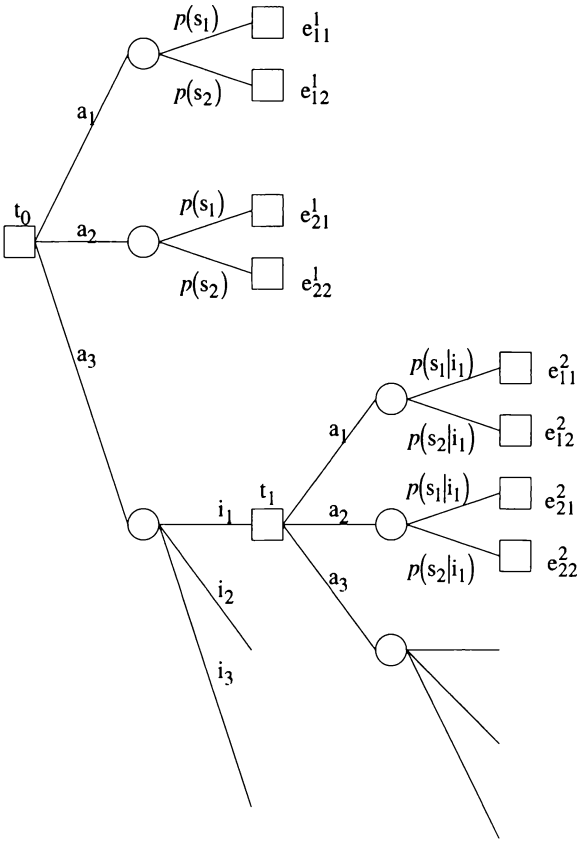

Abb. 1.3: Prüfung als mehrstufiges Entscheidungsproblem

Unsicherheitsmaßes auch geprüft werden, ob für dieses Maß eine dem Theorem von *Bayes* äquivalente Möglichkeit gefunden wird, die Prüfungsergebnisse zur Verbesserung der Prognose über die tatsächliche Beschaffenheit von *X* heranzuziehen.

Zweiter Teil

Risikomessung im betrachteten Prüfungsmodell

A. Risikobegriff

Bevor die Eignung unterschiedlicher Maße für die Risikomessung untersucht wird, soll eine ausführliche Diskussion des für das vorliegende Prüfungsmodell relevanten Risikobegriffs vorgenommen werden. Dieser muß operational in dem Sinn sein, daß der mit Risiko bezeichnete Sachverhalt zum einen meßbar ist und zum anderen durch die Prüfungshandlungen beeinflußt wird, damit sich Handlungsanweisungen für die Prüfung ableiten lassen. Dazu soll der hier relevante Risikobegriff zunächst vom allgemeinen bzw. umgangssprachlichen Risikobegriff sowie vom Begriff des Risikos in der Entscheidungstheorie abgegrenzt werden, bevor die Begriffsbestimmung selbst vorgenommen wird.

I. Umgangssprachlicher Risikobegriff

Umgangssprachlich bedeutet Risiko „Wagnis“ oder „Verlustgefahr“, aber auch „Wert des versicherten Gegenstandes“[1]. Da sich auch etymologisch keine weitere Präzisierung des Begriffsinhaltes erreichen läßt[2], ist demnach zunächst zu konstatieren, daß zumindest der umgangssprachliche Risikobegriff ausschließlich die negativen Erfolgsbeiträge von Entscheidungen bezeichnet, wenn unter Wert des versicherten Gegenstandes das Verlustrisiko der Versicherung verstanden wird. Soweit eine Messung des mit einer Entscheidung verbundenen Risikos in Betracht gezogen wird, kann diese sich sowohl auf die Höhe der negativen Erfolgsbeiträge als auch auf ein Maß für das Eintreten dieser „Risiken“ beziehen[3]. Darüber hinaus kann eine Verknüpfung dieser beiden Risikomaße zu einem kompensatorischen Maß dann notwendig sein, wenn, ohne dies hier näher zu quantifizieren, das Risiko bei hohen negativen Erfolgsbeiträ-

---

1 Vgl. z. B. *Mackensen*, Deutsches Wörterbuch, 1982.

2 Vgl. *Zimmer*, Risiko, 1994, S. 4.

3 So wohl auch die Ansätze zur Risikoforschung, die Risiko als ein Trippel aus Szenario, Eintrittswahrscheinlichkeit für das Szenario und Schaden betrachten. Vgl. *Kaplan/Garrick*, Bestimmung, 1993, S. 94f.

gen, denen ein geringes Maß für ihr tatsächliches Eintreten zugeordnet wird, gleich dem Risiko von niedrigeren negativen Erfolgsbeiträgen mit einem höheren Maß für das tatsächliche Eintreten empfunden wird. Wie diese Kompensation im einzelnen erfolgt, ist abhängig von den jeweiligen subjektiven Einstellungen.

Für die vorliegende Prüfungssituation ist demnach zu prüfen, welches „Wagnis" der Prüfer eingeht, oder besser, welche „Verlustgefahr" für den Prüfer in der beschriebenen Entscheidungssituation besteht. Dies erfordert eine genauere Analyse des Ergebnisraumes der vorliegenden Entscheidungssituation. Oben wurde die Annahme getroffen, daß es sich bei den Ergebnissen um Nutzenbeiträge handelt[4], so daß für die Abbildung in die reellen Zahlen einzelne Zielbeiträge aggregiert werden. Für die Frage, ob und welcher Verlust bei Wahl einer bestimmten Alternative aus A entsteht, muß zunächst festgelegt werden, auf welcher Bezugsebene ein Verlust gemessen werden soll. Dabei kann zum einen der Verlust im Hinblick auf einzelne Zielfunktionen, wie Einkommenserzielung, Ansehen oder Zufriedenheit mit der eigenen Leistung (Entscheidung) gemessen werden. Es werden somit ausschließlich die Verlustbeiträge zu einzelnen Zielfunktionen betrachtet. Zum anderen kann die Verlustmessung auf der Ebene der aggregierten Ergebnisse, also in Bezug auf den Nutzen erfolgen. Es müßte dann ein Nutzenniveau festgelegt werden, dessen Unterschreitung als Verlust betrachtet wird.

Zu dem Problem der Festlegung relevanter Maßgrößen für den Verlust kommt bei der Verwendung des umgangssprachlichen Risikobegriffs ein weiteres Problem hinzu. Die bei der Wahl einzelner Aktionen möglichen Verluste sind von den Ergebnissen der Prüfungshandlungen weitgehend unabhängig[5] und können demnach zumindest bezüglich ihrer Höhe nicht durch Prüfungshandlungen beeinflußt werden. Als Kriterium für die Auswahl von Prüfungshandlungen ist ein in dieser Form definiertes Risiko demnach nicht geeignet. Das hier betrachtete Risiko entspricht in seiner Bedeutung somit nicht dem umgangssprachlichen Risikobegriff.

---

[4] Siehe oben Kapitel C. I. im Ersten Teil der Arbeit.

[5] Abhängigkeiten zwischen Prüfungsergebnissen und auftretenden „Verlusten" liegen z. B. dann vor, wenn die getroffene Entscheidung des Abschlußprüfers bei den gegebenen Prüfungsergebnissen eine Pflichtverletzung im Sinne des § 323 Abs. 1 Satz 3 HGB darstellt, die zu Schadenersatzansprüchen führt.

II. Risikobegriff der Entscheidungslogik

In der Entscheidungslogik werden als Risikosituationen solche Entscheidungssituationen bezeichnet, in denen der relevante Zustandsraum S mehr als ein Element enthält (Ungewißheit i. w. S.), und der Entscheidende für das Eintreten der einzelnen Umweltzustände Wahrscheinlichkeiten angeben kann[6]. Bei den Wahrscheinlichkeiten muß es sich nicht zwingend um objektive Wahrscheinlichkeiten handeln. Es wird allgemein für ausreichend gehalten, wenn für das Eintreten der Umweltzustände subjektive Wahrscheinlichkeiten vorliegen[7]. Risiko im Sinne der Entscheidungslogik beschreibt demnach nur den Umstand, daß zum Entscheidungszeitpunkt der aus einer Aktion resultierende Nutzen nicht eindeutig bestimmt, sondern ausschließlich als Wahrscheinlichkeitsverteilung von möglichen Nutzenwerten bekannt ist. Dabei wird davon auszugehen sein, daß die Höhe des Risikos, das einer bestimmten Aktion $a \in A$ zugeordnet wird, von der Form der für die Ergebnisse von a vorliegenden Wahrscheinlichkeitsverteilung abhängt.

Ein Maß für die Höhe des Risikos wird im Rahmen der Entscheidungsfindung, wenn der Entscheidende ein für ihn geltendes Entscheidungsprinzip festgelegt hat, nicht benötigt, soweit er nicht das Risiko einer Aktion explizit neben den in seiner Nutzenmessung aggregierten Zielfunktionen für die Entscheidung berücksichtigen will[8]. Eine Risikomessung für die Entscheidungsfindung erfolgt demnach insbesondere dann nicht, wenn entweder eine Entscheidungsregel verwendet wird, die das Risiko der einzelnen Aktionen nicht explizit berücksichtigt, wie das z. B. bei der Erwartungswertregel der Fall ist, die alle Wahrscheinlichkeitsverteilungen mit gleichem Erwartungswert unabhängig von der Form der Verteilung in Bezug auf die Präferenzen des Entscheidenden als indifferent unterstellt[9], oder wenn die Risikoeinstellung nur implizit bei der Entscheidungsfindung berücksichtigt wird, wie dies beim *Bernoulli*-Prinzip der Fall ist. Darauf soll im folgenden kurz eingegangen werden.

Das Problem, unter den einzelnen risikobehafteten Aktionen $a \in A$ zu wählen, wird bei Anwendung des *Bernoulli*-Prinzips dadurch vereinfacht, daß unabhängig vom konkreten Entscheidungsproblem eine Risikonutzenfunktion des Entscheidenden bestimmt wird, die neben dem Nutzen eines Ergebnisses $e \in E$

[6] Vgl. *Schneeweiß*, Entscheidungskriterien, 1967, S. 27; *Laux*, Entscheidungstheorie I, 1991, S. 25.

[7] Vgl. *Müller*, Risiko, 1993, Sp. 3816. Zur Definition und Unterscheidung von objektiven und subjektiven Wahrscheinlichkeiten wird auf Kapitel B. I. in diesem Teil der Arbeit verwiesen.

[8] Vgl. *Schneider*, Informationstheorie, 1995, S. 115-119.

[9] Vgl. *Laux*, Entscheidungstheorie I, 1991, S. 167.

DOI https://doi.org/10.3790/978-3-428-48909-1

auch berücksichtigt, daß dieses Ergebnis nicht mit Sicherheit erreicht werden kann. Die Existenz der Risikonutzenfunktion ist gesichert, wenn die Präferenzen des Entscheidenden bestimmten Axiomen gehorchen[10]. Die Ermittlung der Risikonutzenfunktion erfolgt durch die Auswertung unterschiedlicher Indifferenzsituationen, in denen jeweils das zu einer einfachen Wahrscheinlichkeitsverteilung aus dem besten und schlechtesten Ergebnis aus E äquivalente sichere Ergebnis (Sicherheitsäquivalent) bestimmt wird. Ist die Risikonutzenfunktion bekannt, können alle Ergebnisse aus E in Risikonutzenwerte transformiert werden. Als optimale Handlungsalternative wird die Aktion mit dem höchsten Erwartungswert des Risikonutzens gewählt[11]. Wie angeführt erfolgt demnach bei der Verwendung des *Bernoulli*-Prinzips keine Risikomessung für einzelne Aktionen. Die Risikoeinstellung geht in die Entscheidungsfindung nur über die für den Entscheidenden zugrundegelegte Risikonutzenfunktion ein, so daß auch nach der Transformation der Ergebnisse in Risikonutzen keine Möglichkeit besteht, das Risiko einzelner Aktionen abzuleiten.

Es bleibt demnach zu untersuchen, welche Begriffsfassung für Risiko im Rahmen der Entscheidungsprinzipien, die eine explizite Berücksichtigung von Risiko als zusätzliche Zielfunktion bei der Entscheidung berücksichtigen, gewählt wird. Dabei sind insbesondere zwei Formen zu unterscheiden. Zum einen sind hier die Entscheidungsprinzipien zu nennen, bei denen Parameter der Ergebnisverteilung einzelner Aktionen als Risikomaße verwendet werden. Beispiele für diskutierte Parameter sind die Standardabweichung, höhere Momente der Verteilung oder der Verlusterwartungswert[12]. Die im Schrifttum diskutierten Risikomaße sollen hier nicht im einzelnen gezeigt werden, ihnen ist gemeinsam, daß ein in dieser Form gemessenes Risiko insbesondere auch von der Höhe des Nutzens, der den einzelnen Ergebnissen zugeordnet wird, abhängt. Insoweit sollen diese Maße aus den gleichen Gründen, die zur Ablehnung des umgangssprachlichen Risikobegriffs geführt haben, nicht zur Definition des Begriffs Risiko für das vorliegende Modell verwendet werden.

Eine zweite Form der Begriffsfassung definiert Risiko als die Wahrscheinlichkeit dafür, daß bestimmte Ziele durch die Wahl einer Aktion aus A nicht erreicht werden. Dazu zählen insbesondere Ansätze, die das Risiko durch die Ruinwahrscheinlichkeit messen. Diese drückt aus, mit welcher Wahrschein-

---

[10] Das bekannteste Axiomensystem, das die Existenz einer Risikonutzenfunktion sichert, stammt von *Luce/Raiffa*. Vgl. *Luce/Raiffa*, Games, 1957, S. 23-31.

[11] Vgl. *Laux*, Entscheidungstheorie I, 1991, S. 168.

[12] Vgl. *Schneeweiß*, Entscheidungskriterien, 1967, S. 52-57 und S 61; *Schneider*, Informationstheorie, 1995, S. 27.

lichkeit bei Wahl einer bestimmten Aktion ein Verlust eintritt, der nicht mehr hingenommen werden kann, da er den Ruin des Entscheidenden bedeutet[13]. Ebenfalls in diese Gruppe gehören Entscheidungsprinzipien, welche die Wahrscheinlichkeit dafür, daß ein bestimmter Mindestbetrag als Ergebnis nicht realisiert wird, als Risiko definieren[14]. Auch diese Entscheidungsprinzipien stellen auf bestimmte Ergebnisgrößen ab. Insoweit trifft der oben vorgebrachte Einwand hinsichtlich der Eignung von Risikomaßen, welche die Ergebnisse in die Risikomessung einbeziehen, auch hier zu. Den Ansätzen ist allerdings gemeinsam, daß sie Risiko als den Eintritt eines unerwünschten Ergebnisses definieren. Insoweit entsprechen sie dem umgangssprachlichen Risikobegriff. Darüber hinaus messen sie die Höhe des Risikos einzelner Aktionen durch die Eintrittswahrscheinlichkeit des unerwünschten Ergebnisses bei Wahl dieser Aktion. Dieser Gedanke soll im folgenden Abschnitt fortgeführt werden, um einen für das hier unterstellte Prüfungsmodell operationalen Risikobegriff zu erhalten.

III. Relevanter Risikobegriff für das verwendete Prüfungsmodell

Nachdem im vorigen Abschnitt der relevante Risikobegriff soweit konkretisiert wurde, daß damit ein Maß für den Eintritt unerwünschter Ergebnisse bei der Wahl einer bestimmten Aktion aus A gemeint ist, soll zunächst geprüft werden, welche Ergebnisse in E unerwünscht in dem Sinn sind, daß sie „Risiken" für den Prüfer darstellen könnten. In Frage kommen hierfür e_{12} und e_{21}, die beide Fehlurteile des Prüfers bedeuten, da die Einschätzung des Prüfers über die Prüfungsgesamtheit tatsächlich falsch ist. Dabei soll das Ergebnis e_{21} übereinstimmend mit den bei statistischen Hypothesentests möglichen Fehlern als Fehler 1. Art, das Ergebnis e_{12} als Fehler 2. Art bezeichnet werden. Im Schrifttum zur Jahresabschlußprüfung wird dabei ausschließlich ein Fehler 2. Art, also e_{12} als Prüferrisiko bezeichnet[15]. Die Situation in der das Ergebnis e_{21} eintritt, also der Fall, daß der Prüfer die Prüfungsgesamtheit als nicht ordnungsmäßig beurteilt, obwohl sie tatsächlich ordnungsmäßig ist, wird regelmäßig als Auftraggeberrisiko behandelt[16]. Diese unterschiedliche Behandlung der beiden möglichen Fehlurteile des Prüfers wird dem hier verwendeten Prü-

13 Vgl. *Schneeweiß*, Entscheidungskriterien, 1967, S. 57.

14 Vgl. *Koch*, Unsicherheit, 1989, Sp. 2062.

15 Vgl. *Schulte*, Methoden, 1970, S. 109; *Gans*, Prüfungen, 1986, S. 490; *v. Wysocki*, Grundlagen, 1988, S. 172; *Dörner*, Audit Risk, 1992, Sp. 82; *AICPA*, SAS No. 47, 1994, AU Sec. 312 § 2; *Ködel*, Abschlußprüfung, 1997, S. 66.

16 Vgl. *v. Wysocki*, Grundlagen, 1988, S. 172; *Wiedmann*, Prüfungsansatz, 1993, S. 15.

fungsmodell nicht gerecht, da der Prüfer dann bei der Wahl von Aktion a_2 kein Risiko eingehen würde.

Die Diskrepanz zwischen dem hier untersuchten Modell und den Auffassungen zum Risiko des Prüfers im Schrifttum resultiert insbesondere daraus, daß im vorliegenden Modell angenommen wird, daß der tatsächliche Zustand der Prüfungsgesamtheit immer bekannt wird, mithin jedes Fehlurteil des Prüfers aufgedeckt wird. Demgegenüber wird im Schrifttum zur Jahresabschlußprüfung regelmäßig davon ausgegangen, daß durch Reaktionen des geprüften Unternehmens auf das die Ordnungsmäßigkeit ablehnende Urteil des Prüfers Fehler 1. Art eher aufgedeckt werden als Fehler 2. Art[17]. Außerdem dürfte zumindest für die Prüfungspraxis der Fehler 2. Art bedeutender sein, da die überwiegende Zahl der Jahresabschlüsse einen uneingeschränkten Bestätigungsvermerk erhält[18], d.h., ein Fehler 1. Art kann in diesen Fällen nicht auftreten. Da diese Unterscheidung im vorliegenden Prüfungsmodell aus den oben genannten Gründen[19] nicht erfolgen soll, werden beide möglichen Fehler, die der Prüfer bei der Abgabe eines Urteils machen kann, als die für die Bestimmung des Risikobegriffs benötigten unerwünschten Ereignisse betrachtet. Die Ergebnisse sind für den Prüfer insbesondere durch den mit dem Bekanntwerden der Fehlurteile verbundenen Vertrauens- und Ansehensverlust und dem damit verbundenen Einfluß auf die zukünftige Einkommenserzielung als unerwünscht zu qualifizieren[20].

Neben der Definition, was als unerwünschtes Ergebnis zu betrachten ist, wird eine Vorschrift zur Messung des Risikos der in A enthaltenen Aktionen benötigt. Da die Höhe des Risikos von der Höhe des Nutzens, der durch die Ausprägungen der in E enthaltenen Ergebnisse ausgedrückt wird, unabhängig sein soll, verbleiben für die Risikomessung im Entscheidungsmodell nur die dem Eintreten der unterschiedlichen Umweltzustände zugeordneten Maßzahlen $p(s \in S)$ bzw., wenn die Einbeziehung der Ergebnisse von Prüfungshandlungen erfolgen soll, die sich nach Erhalt einer Prüfungsinformation ergebenden Einschätzungen für das Eintreten der einzelnen Umweltzustände $p(s \in S | i \in I)$. Fraglich ist jetzt nur noch, welche Art von Maß das in dieser Weise definierte Risikomaß ist. Im Schrifttum zur Jahresabschlußprüfung wird hauptsächlich unter-

---

17 Vgl. *v. Wysocki*, Grundlagen, 1988, S. 173; *AICPA*, SAS No. 47, 1994, AU Sec. 312 § 2; *Leffson*, Wirtschaftsprüfung, 1988, S. 190; *IDW*, HFA 1/1988, 1988, S. 244.

18 Vgl. hierzu *WPK*, Bericht, 1995, S. 2 und 8. Demnach wurde von 11225 Jahresabschlüssen einem der Bestätigungsvermerk versagt.

19 Siehe oben Kapitel C. I. im Ersten Teil der Arbeit.

20 Vgl. *Leffson*, Wirtschaftsprüfung, 1988, S. 313.

stellt, daß es sich, soweit überhaupt auf ein metrisch skaliertes Maß abgestellt wird, um ein Wahrscheinlichkeitsmaß handeln muß[21]. Falls die Ergebnisse aus stichprobenweiser Prüfung mit bewußter Auswahl der Stichprobenelemente oder aus den indirekten Messungen in die Beurteilung des Eintretens der Umweltzustände einbezogen werden sollen, kann die angeführte Unschärfe der Prüfungsinformationen nicht zwingend mit Hilfe eines Wahrscheinlichkeitsmaßes gemessen werden. Daher soll die Eignung unterschiedlicher Unschärfemaße zur Risikomessung im nächsten Kapitel diskutiert werden.

B. Eignung unterschiedlicher Unschärfemaße zur Risikomessung

Unter Messen eines Merkmales einer Menge von realen Sachverhalten soll die Zuordnung von Zahlen zu den einzelnen Elementen der betrachteten Menge in einer solchen Form verstanden werden, die eine adäquate Abbildung der Struktur der realen Sachverhalte in bezug auf das betrachtete Merkmal gewährleistet[22]. Im vorliegenden Kapitel wird demnach untersucht, mit welchen Unschärfemaßen die festgestellten, unterschiedlichen Formen von Unschärfe der vorhandenen bzw. zu gewinnenden Prüfungsinformationen[23] abgebildet werden können. Eine rein nominale Messung, wie sie in Aussagen der Form: „Bei Wahl der Aktion a_1 besteht das Risiko eines Fehlurteils" enthalten ist, oder komparative Messungen des Risikos, die ausschließlich Aussagen der Form: „Das Risiko eines Fehlurteils bei Wahl der Aktion a_1 ist größer als bei Wahl von a_2" ermöglichen, reichen für die Zielsetzung der Arbeit, eine am Risiko orientierte Auswahl von Prüfungshandlungen zu ermöglichen, nicht aus. Es werden daher ausschließlich Maße, die Abbildungen in die reellen Zahlen liefern, untersucht. Die weitere Gliederung richtet sich dabei nach der Strenge der Anforderungen, die an die Struktur der zu messenden Sachverhalte gestellt werden, nicht nach den Formen von Unschärfe, wie sie im Ersten Teil der Arbeit aufgezeigt wurden.

Als letzte Vorbemerkung soll noch auf die Isomorphie zwischen der Abbildung von Sachverhalten, oder genauer von Aussagen über Merkmale von Sachverhalten in der formalen Logik, wie sie oben für die Darstellung indirekter Messungen verwendet wurde, und der mengentheoretischen Darstellung, die für die folgenden Ausführungen gewählt wird, hingewiesen werden. Betrachtet

---

[21] Vgl. *Cushing/Loebbecke*, Audit Risk, 1983, S. 29; *v. Wysocki*, Grundlagen, 1988, S. 173; *Wiedmann*, Prüfungsansatz, 1993, S. 16.

[22] Vgl. *Krantz* et al., Measurement, 1971, S. 1.

[23] Siehe oben Kapitel B. im Ersten Teil der Arbeit.

wird die Menge aller im Hinblick auf die Untersuchung relevanten Sachverhalte U'. Mit A soll eine Teilmenge von U' bezeichnet werden, die dadurch gebildet wird, daß alle Sachverhalte in A eine bestimmte Eigenschaft A besitzen. Die mengentheoretische Festlegung, daß ein Element in A enthalten ist, formal mit $x \in A$ ausgedrückt, entspricht somit der logisch wahren Aussage a: „x hat die Eigenschaft A", wobei x keine Variable ist, sondern einen bestimmten Sachverhalt bezeichnet. Entsprechend gelten die folgenden Isomorphien: $\sim a$ ist isomorph $x \in \overline{A}$[24], $a \wedge b$ ist isomorph $x \in A \cap B$, $a \vee b$ ist isomorph $x \in A \cup B$, $a \rightarrow b$ ist isomorph $A \subseteq B$ und $a \leftrightarrow b$ ist isomorph $A = B$.

I. Wahrscheinlichkeitsmaß

Bevor eine Diskussion der Eignung von Wahrscheinlichkeitsmaßen für die Risikomessung im vorliegenden Modell erfolgt, soll zunächst eine Axiomatisierung des Wahrscheinlichkeitsmaßes und die Ableitung einiger Konsequenzen, die für die späteren Diskussionen benötigt werden, erfolgen.

1. Grundbegriffe

a) Axiome

Ausgangspunkt für die Definition eines Wahrscheinlichkeitsmaßes ist eine nicht leere, endliche Menge Ω elementarer Ereignisse (Möglichkeitsraum) und ein Mengenkörper $\mathbb{K}$ über Ω (Ereigniskörper). Somit gilt für $\mathbb{K}$ [25]:

(2.1) $$\Omega \in \mathbb{K};$$

(2.2) $$A \in \mathbb{K} \leftrightarrow \overline{A} \in \mathbb{K}^{26};$$

(2.3) $$\forall A, B \in \mathbb{K} : A \cup B \in \mathbb{K}.$$

Unter einem Wahrscheinlichkeitsmaß $pr : \mathbb{K} \rightarrow [0;1]$[27] wird dann jedes Maß verstanden, das den folgenden Axiomen gehorcht[28]:

(2.4) $$\forall A \in \mathbb{K} : pr(A) \geq 0;$$

---

24 Die Bildung des Komplements erfolgt dabei bezogen auf U', also $\overline{A} = U' - A$.

25 Vgl. z. B. *Krantz* et al., Measurement, 1971, S. 199.

26 Die Bildung des Komplements erfolgt dabei bezogen auf Ω, also $\overline{A} = \Omega - A$.

27 Die zu definierenden Unschärfemaße werden im weiteren mit kleingeschriebenen, kursiv gesetzten Abkürzungen bezeichnet, *pr* steht demnach für „probability".

28 Es werden hier die Axiome von *Kolmogoroff* verwendet, da die von ihm auf dem Mengenbegriff aufbauende Axiomatisierung des Wahrscheinlichkeitsmaßes für den weiteren Verlauf der Arbeit am zweckdienlichsten erscheint. Vgl. *Kolmogoroff*, Grundbegriffe, 1933, S. 2.

(2.5) $$pr(\Omega)=1;$$

(2.6) $$\forall A,B\in \mathbb{K}: pr(A\cup B)=pr(A)+pr(B)-pr(A\cap B).$$

Die Aussage $pr(A)=\mathrm{r}$ bedeutet demnach, daß die Wahrscheinlichkeit für das Eintreten eines der Elemente bzw. Ereignisse aus A durch die reelle Zahl r gemessen werden kann. Dieser Sachverhalt soll im weiteren kurz die Wahrscheinlichkeit von A genannt werden. Aus Axiom (2.5) und (2.6) folgt unmittelbar, daß für jede Menge $A\in\mathbb{K}$ die Wahrscheinlichkeit ihres Komplements mit $pr(\overline{A})=1-pr(A)$ bestimmt werden kann, insbesondere ist die Wahrscheinlichkeit des unmöglichen Ereignisses $pr(\varnothing)=0$[29].

b) Bedingte Wahrscheinlichkeiten

Für die weiteren Ausführungen werden neben den bisher betrachteten absoluten Wahrscheinlichkeiten noch bedingte Wahrscheinlichkeiten benötigt. Wenn $A,B\in\mathbb{K}$ gilt, so ist die bedingte Wahrscheinlichkeit für A, wenn B schon eingetreten ist, definiert durch[30]:

(2.7) $$pr(A|B)=\frac{pr(A\cap B)}{pr(B)} \text{ für } pr(B)>0.$$

Als speziellen Fall von (2.7) erhält man:

(2.8) $$pr(A|\Omega)=pr(A).$$

D. h., jede absolute Wahrscheinlichkeit läßt sich auch als bedingte Wahrscheinlichkeit ausdrücken.

Da Formel (2.7) symmetrisch ist, also auch für die bedingte Wahrscheinlichkeit des Eintretens von B, wenn A schon gegeben ist, gilt:

(2.9) $$pr(B|A)=\frac{pr(A\cap B)}{pr(A)} \text{ für } pr(A)>0.$$

Durch Kombination der beiden Formeln (2.7) und (2.9) erhält man das Theorem von *Bayes*[31]:

(2.10) $$pr(A|B)=\frac{pr(A)pr(B|A)}{pr(B)}.$$

---

[29] Vgl. *Kolmogoroff*, Grundbegriffe, 1933, S. 6.

[30] Vgl. *Kolmogoroff*, Grundbegriffe, 1933, S. 6.

[31] Vgl. *Kolmogoroff*, Grundbegriffe, 1933, S. 7.

Dieses läßt sich darüber hinaus auch für beliebige disjunkte Zerlegungen des Möglichkeitsraumes zeigen. Wird Ω in n disjunkte Teilmengen A_i zerlegt, gilt:

$$(2.11) \qquad A_i, A_j \subset \Omega \wedge \forall i \neq j: A_i \cap A_j = \emptyset \wedge \bigcup_{i=1}^{n} A_i = \Omega .$$

Eine beliebige Menge $E \in \mathbb{K}$ läßt sich dann ausdrücken als die Vereinigungsmenge der einzelnen Schnittmengen mit den Mengen A_i, d. h.:

$$(2.12) \qquad E = \bigcup_{i=1}^{n} E \cap A_i .$$

Die absolute Wahrscheinlichkeit von E (Satz von der totalen Wahrscheinlichkeit[32]) ergibt sich dann unter Verwendung von (2.5), (2.6), (2.7) und (2.12) mit:

$$(2.13) \qquad pr(E) = \sum_{i=1}^{n} pr(A_i)\, pr(E|A_i).$$

Damit läßt sich das Theorem von *Bayes* allgemein schreiben als:

$$(2.14) \qquad pr(A_i|E) = \frac{pr(A_i)\, pr(E|A_i)}{pr(E)} = \frac{pr(A_i)\, pr(E|A_i)}{\sum_{i=1}^{n} pr(A_i)\, pr(E|A_i)}, \text{ mit } pr(E) > 0.$$

Wird demnach unterstellt, daß sich die Ergebnisse der Prüfungshandlungen durch E darstellen lassen, und bezeichnet A_1 z. B. die tatsächliche Ordnungsmäßigkeit der Prüfungsgesamtheit X, so kann mit Hilfe des Theorems von *Bayes* die Wahrscheinlichkeit für die Ordnungsmäßigkeit der Prüfungsgesamtheit nach Erhebung der Prüfungsinformation dann ermittelt werden, wenn die totale Wahrscheinlichkeit $pr(E)$ sowie die likelihood $pr(E|A_1)$ von E bekannt sind[33].

Bei der Anwendung des Theorems von *Bayes* ist jedoch zu beachten, daß die resultierende a-posteriori-Wahrscheinlichkeiten in gleichem Maße von den a-priori-Wahrscheinlichkeiten $pr(A_i)$ und den likelihoods abhängen. Dies wird insbesondere dann deutlich, wenn ausschließlich das Verhältnis der Wahr-

---

32 Vgl. *Kolmogoroff*, Grundbegriffe, 1933, S. 7.

33 Zu einem solchen Ansatz am Beispiel der Beurteilung eines internen Kontrollsystems vgl. *Hömberg*, Prüfungen, 1981, S. 19-21.

scheinlichkeiten für A und $\overline{A}$ vor und nach Berücksichtigung der Prüfungsinformation E nach (2.14) betrachtet wird[34]:

$$\text{(2.15)} \qquad \frac{pr(A|E)}{pr(\overline{A}|E)} = \frac{pr(A)}{pr(\overline{A})} \frac{pr(E|A)}{pr(E|\overline{A})} = \frac{pr(A)}{pr(\overline{A})} \frac{pr(E|A)}{1 - pr(\overline{E}|\overline{A})}.$$

Wie in (2.15) zu erkennen ist, ist das Verhältnis der Eintrittswahrscheinlichkeiten, nachdem die Information E bekannt ist, in gleichem Maß vom Verhältnis der a-priori-Wahrscheinlichkeiten wie vom Verhältnis der beiden bedingten Wahrscheinlichkeiten abhängig. $pr(\overline{E}|\overline{A})$ gibt dabei an, wie spezifisch die Information E für A (Spezifität) ist, und $pr(E|A)$ beschreibt, wie stark der Zusammenhang (Sensitivität) zwischen A und E ist[35]. Zwischen der Sensitivität und Spezifität einer Information E besteht kein Zusammenhang, beide Größen müssen separat ermittelt werden. Eine Prüfungsinformation E stützt demnach dann A bzw. erschüttert $\overline{A}$, wenn gilt:

$$\text{(2.16)} \qquad \frac{pr(E|A)}{1 - pr(\overline{E}|\overline{A})} > 1.$$

Sollen a-posteriori-Wahrscheinlichkeiten zur Risikomessung verwendet werden, so kann aus dem Umstand, daß das Verhältnis der a-posteriori-Wahrscheinlichkeiten auch durch das Verhältnis der a-priori-Wahrscheinlichkeiten beeinflußt wird, für genügend kleine $pr(A)$ das Problem auftreten, daß die vorliegenden Informationen A zwar stark stützen, also A auch durch E gut erklärt wird, aber trotzdem bei Betrachtung der a-posteriori-Wahrscheinlichkeiten ein hohes Risiko angezeigt wird[36]. Dies ist immer dann der Fall, wenn gilt:

$$\text{(2.17)} \qquad \frac{pr(A)}{pr(\overline{A})} < \frac{1}{\dfrac{pr(E|A)}{1 - pr(\overline{E}|\overline{A})}}.$$

Soll auf a-posteriori-Wahrscheinlichkeiten zur Risikomessung abgestellt werden, muß demnach entweder sichergestellt werden, daß der Einfluß der a-priori-Wahrscheinlichkeit vernachläßigbar klein ist, oder es muß geprüft werden, inwieweit die Interpretation der a-priori-Wahrscheinlichkeiten deren Einbeziehung in die Risikomessung zuläßt. Diese Frage kann allerdings ausschließlich für die jeweilige Interpretation des Wahrscheinlichkeitsmaßes geklärt werden.

---

[34] Vgl. *Pearl*, Reasoning, 1988, S. 34f.

[35] Vgl. *Spies*, Unsicheres Wissen, 1993, S. 37.

[36] Vgl. hierzu auch *Popper*, Logik, 1994, S. 342f.

c) Unabhängigkeit von Ereignissen

Als letzer Begriff soll noch die statistische Unabhängigkeit zweier Ereignisse definiert werden. A ist statistisch unabhängig von B, wenn gilt[37]:

$$pr(A|B) = pr(A) \text{ für } pr(A) > 0. \tag{2.18}$$

D. h., die Information, daß ein Ereignis aus B eingetreten ist, ist für die Höhe der Wahrscheinlichkeit des Eintretens von A nicht von Bedeutung. Als Folgerung ergibt sich die Wahrscheinlichkeit des Eintretens von A und gleichzeitig B mit:

$$pr(A \cap B) = pr(A)pr(B). \tag{2.19}$$

Eine zweite Form der Unabhängigkeit von Ereignissen erlangt insbesondere dann Bedeutung, wenn bei der Anwendung des Theorems von *Bayes* aus Formel (2.14) die Prüfungsinformationen E sich aus einzelnen Prüfungsinformationen, z. B. E_1 und E_2, die nacheinander erworben werden, zusammensetzt. D. h., es ist folgende Wahrscheinlichkeit zu bestimmen:

$$\begin{aligned} pr(A_i|E_1 \cap E_2) &= \frac{pr(A_i)pr(E_1 \cap E_2|A_i)}{pr(E_1 \cap E_2)} \\ &= \frac{pr(E_2|A_i \cap E_1)pr(A_i|E_1)}{pr(E_2|E_1)}. \end{aligned} \tag{2.20}$$

Wie in Formel (2.20) zu erkennen ist, darf die nach Verwendung von Information E_1 ermittelte Wahrscheinlichkeit für A_i nicht ohne weiteres als neue a-priori-Wahrscheinlichkeit in Formel (2.14) eingesetzt werden. Dieses Vorgehen würde zum einen die Unabhängigkeit von E_1 und E_2 im Sinn von (2.18) erfordern, zum anderen müßte gelten:

$$pr(E_2|A_i \cap E_1) = pr(E_2|A_i). \tag{2.21}$$

Die in (2.21) angegebene Bedingung ist ausschließlich dann erfüllt, wenn gilt:

$$pr(E_1 \cap E_2|A_i) = pr(E_1|A_i)pr(E_2|A_i). \tag{2.22}$$

Diese Form der Unabhängigkeit von E_1 und E_2 in bezug auf einen Kontext A_i soll als bedingte Unabhängigkeit bezeichnet werden[38]. Dabei existiert kein Zu-

---

[37] Vgl. *Kolmogoroff*, Grundbegriffe, 1933, S. 11; *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 71.

[38] Vgl. *Pearl*, Reasoning, 1988, S. 31.

DOI https://doi.org/10.3790/978-3-428-48909-1

sammenhang zwischen der Unabhängigkeit zweier Ereignisse nach (2.18) und der bedingten Unabhängigkeit dieser Ereignisse bezogen auf einen Kontext A_i gemäß (2.22), insbesondere folgt nicht die bedingte Unabhängigkeit aus der Unabhängigkeit zweier Ereignisse nach (2.18), wie sich anhand des folgenden Gegenbeispiels leicht zeigen läßt. Es soll gelten, daß $\Omega = \{1;2;3;4;5;6\}$, $E_1 = \{2;4;6\}$, $E_2 = \{1;2\}$ und $A = \{1;2;3\}$ ist. Wird für die Elementarereignisse in Ω Gleichwahrscheinlichkeit angenommen, dann gilt:

(2.23) $$pr(E_2|E_1) = \tfrac{1}{3} = pr(E_2).$$

Die beiden Ereignisse E_1 und E_2 sind gemäß (2.18) statistisch unabhängig, ohne daß sie bedingt unabhängig im Sinne von (2.21) bezogen auf den Kontext A sind, wie ein Vergleich der bedingten Wahrscheinlichkeiten zeigt:

(2.24) $$pr(E_2|A \cap E_1) = 1 \neq pr(E_2|A) = \tfrac{2}{3} \text{ und}$$

(2.25) $$pr(E_1|A \cap E_2) = \tfrac{1}{2} \neq pr(E_1|A) = \tfrac{1}{3}.$$

2. *Interpretationen*

Die im vorherigen Abschnitt erfolgte Axiomatisierung eines Wahrscheinlichkeitsmaßes legt ausschließlich Anforderungen fest, die erfüllt sein müssen, um das Rechnen mit Wahrscheinlichkeiten zu erlauben. Sie enthalten keinerlei Festlegungen hinsichtlich des Gegenstandes und davon abhängend der Methode zur Messung von Wahrscheinlichkeiten. Für den hier vorliegenden Anwendungsfall der Risikomessung bei Prüfungen muß demnach festgelegt werden, welche Ereignisse im Möglichkeitsraum Ω zusammengefaßt werden, und wie die Messung der Wahrscheinlichkeiten für Ereignisse aus Ω erfolgen soll. Diese Festlegungen hängen eng mit der Interpretation der Wahrscheinlichkeit als objektive bzw. subjektive Maßgröße zusammen. Sie sollen daher getrennt für die beiden unterschiedenen Formen diskutiert werden.

a) Objektive Wahrscheinlichkeiten

Als objektiv sollen solche Interpretationen des Wahrscheinlichkeitsbegriffs bezeichnet werden, die zumindest prinzipiell eine empirische Überprüfung der Zuweisung von Wahrscheinlichkeiten zu Ereignismengen nicht ausschließen. Dabei kommen insbesondere Ansätze in Frage, die Wahrscheinlichkeiten als Grenzwerte relativer Häufigkeiten in Folgen von Ereignissen betrachten[39].

Nachdem gezeigt wurde, daß logische Widersprüche dann entstehen, wenn Wahrscheinlichkeiten für tatsächlich beobachtete Folgen von Ereignissen definiert werden sollen[40], verbleibt als objektive Interpretation des Wahrscheinlichkeitsbegriffs diejenige, welche die Wahrscheinlichkeit als die theoretische Eigenschaft einer bestimmten Versuchsanordnung, also den für die Erzeugung der Folge von Ereignissen geltenden Bedingungen zuschreibt[41]. Für die Versuchsanordnung wird dabei insbesondere gefordert, daß zum einen die einzelnen Versuche statistisch unabhängig im Sinn von (2.18) sind. Dabei folgt die statistische Unabhängigkeit nach h. M. aus der physikalischen Unabhängigkeit[42], allerdings nicht umgekehrt. Darüber hinaus muß die Chance, welche ein Ereignis hat, unabhängig von seiner Position innerhalb einer Folge von Versuchen sein[43]. Aus dieser Interpretation ergeben sich für die Bestimmung von Wahrscheinlichkeiten insbesondere zwei Konsequenzen. Zum einen kann die Wahrscheinlichkeit für ein Ereignis in Form einer Hypothese aus den Versuchsbedingungen abgeleitet werden, zum anderen können tatsächlich beobachtete Folgen von Ereignissen ausschließlich zur Überprüfung der aus den Versuchsbedingungen abgeleiteten Hypothesen verwendet werden[44].

Als nicht zu unterschätzender Vorteil für den vorliegenden Anwendungsfall ergibt sich, daß mit der zuletzt genannten Interpretation auch die Zuweisung von Wahrscheinlichkeiten zu einzelnen Ereignissen möglich ist, da sich diese als die Tendenz, ein bestimmtes Ereignis zu verwirklichen, aus der vorgegebenen Versuchsanordnung ergeben[45]. Würde man demgegenüber auf Wahrscheinlichkeiten als Grenzwert von Ereignisfolgen abstellen, so könnte einem einzelnen Ereignis ohne weiteres keine Wahrscheinlichkeit zugeordnet werden[46]. Erst wenn eine zusätzliche Vereinbarung der Form getroffen würde, daß die Wahrscheinlichkeit eines einzelnen Ereignisses der relativen Häufigkeit des Ereignisses in einer zukünftigen Folge von Ereignissen des gleichen Möglich-

---

39 Vgl. *v. Mises*, Wahrscheinlichkeit, 1972, S. 33f.; *Popper*, Logik, 1994, S. 109; *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 27.

40 Vgl. *Popper*, Propensity, 1959, S. 31-34; *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 32-41.

41 Vgl. *Popper*, Propensity, 1959, S. 34; *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 58f.

42 Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 90.

43 Vgl. *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 72.

44 Vgl. *Popper*, Propensity, 1959, S. 38.

45 Vgl. *Popper*, Propensity, 1959, S. 34.

46 Vgl. *v. Mises*, Wahrscheinlichkeit, 1972, S. 23; *Popper*, Logik, 1994, S. 162.

DOI https://doi.org/10.3790/978-3-428-48909-1

keitsraumes gleichgesetzt wird, wäre eine Zuordnung von Wahrscheinlichkeiten zu singulären Ereignissen bei diesem Ansatz möglich[47]. Dabei bleibt aber eine Begründung für die Zulässigkeit dieser Vereinbarung, soweit die Wahrscheinlichkeit für ein einzelnes Ereignis bestimmt werden soll, und nicht die Wahrscheinlichkeiten unterschiedlicher Ereignisse verglichen werde sollen, fraglich[48].

Neben der Festlegung des Möglichkeitsraumes Ω und dem Ereigniskörper $\mathbb{K}$ tritt beim objektiven Wahrscheinlichkeitsbegriff demnach auch die Festlegung der zum Eintritt der Ereignisse aus Ω führenden Versuchsanordnung. Zur Prüfung, welche der im Ersten Teil der Arbeit beschriebenen Formen von Unschärfe der Prüfungsinformationen durch objektive Wahrscheinlichkeiten abgebildet werden können, muß also untersucht werden, ob die Bedingungen, die zum Eintritt der noch zu definierenden Ereignisse aus Ω führen, tatsächlich eine Versuchsanordnung im oben angesprochenen Sinn darstellen. Dazu soll zunächst untersucht werden, inwieweit ein objektives Wahrscheinlichkeitsmaß grundsätzlich zur Messung der insgesamt entstehenden Unschärfe, also des gesamten Risikos in Betracht kommt.

In der vorliegenden Prüfungssituation interessiert, ob die tatsächliche Ausprägung von X, die eine Abbildung realer Sachverhalte aus U' in bezug auf das geprüfte Unternehmen darstellt, ordnungsmäßig im Sinn der genannten Kriterien ist. Die Gesamtheit der hier zu betrachtenden Menge Ω enthält demnach alle möglichen Ausprägungen, die eine solche Abbildung der Sachverhalte aus U' annehmen kann. Die Kriterien für die Ordnungsmäßigkeit der Prüfungsgesamtheit, die oben abgeleitet wurden[49], lassen mehrere unterschiedliche ordnungsmäßige Abbildungen zu. Die Menge aller ordnungsmäßigen Abbildungen soll im weiteren als O bezeichnet werden, es muß also gelten $O \subseteq \Omega$. Der oben im Entscheidungsfeld angeführte Umweltzustand s_1 entspricht dem Fall, daß das realisierte X in O enthalten ist, entsprechend gilt für s_2, daß X nicht in O, also in $\overline{O}$ bezogen auf Ω enthalten ist. Eine Zuweisung der objektiven Wahrscheinlichkeiten für O und $\overline{O}$ als Maße für das Eintreten der Umweltzustände s_1 und s_2, formal somit $p(s_1) = pr(O)$ und $p(s_2) = pr(\overline{O})$, ist nur dann zulässig, wenn die möglichen Abbildungen von U', von denen X eine Möglichkeit darstellt, durch

47 Vgl. *Popper*, Logik, 1994, S. 162f.

48 Zu einer ausführlichen Diskussion des Zusammenhangs zwischen Einzelereignissen und der Wahrscheinlichkeit als relative Häufigkeit in langen Ereignisfolgen vgl. *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 96-102.

49 Siehe oben Kapitel A. III. im Ersten Teil der Arbeit.

DOI https://doi.org/10.3790/978-3-428-48909-1

eine Versuchsanordnung zustande kommen, die zufällige Folgen von Abbildungen erzeugt.

Ein erster Ansatz könnte sein, über die Versuchsanordnung keine weiteren Aussagen zu treffen, außer der Annahme, daß sie alle möglichen Abbildungen von U' mit gleicher Wahrscheinlichkeit erzeugt. Die Elementarereignisse in Ω entsprechen dann allen Potenzmengen beliebiger Abbildungen von U' auf U. Aus dieser Vorgehensweise resultiert folgendes Problem: Wird mit U_i eine der möglichen Abbildungen von U', die durch die Funktion ω_i erzeugt wird, verstanden, so ergibt sich Ω bei beliebiger Ansatzfunktion π_i formal wie folgt[50]:

$$\Omega = \bigcup_i \mathbb{P}(U_i). \tag{2.26}$$

Bei diesem Ansatz ergibt sich als Problem, daß für beliebig viele mögliche Abbildungen ω_i $(i=1,\ldots,\infty)$ bei einer begrenzten Anzahl ordnungsmäßiger Abbildungen die Wahrscheinlichkeit für O Null ist. Damit ist, da dann gemäß (2.7) und (2.9) keine bedingten Wahrscheinlichkeiten definiert sind, die Anwendung des Theorems von *Bayes* nach (2.14) ausgeschlossen, um Informationen zur Veränderung der Wahrscheinlichkeitseinschätzung zu verwenden. Daher soll dieser Ansatz nicht weiter verfolgt werden.

Aber auch wenn die Zahl der möglichen Abbildungen beschränkt wird, damit die Anzahl der Elemente in Ω endlich ist, wird die Wahrscheinlichkeit für O sehr nahe bei Null sein. In diesen Fall ist zwar die Anwendung des Theorems von *Bayes* möglich, allerdings ergibt sich ein Problem daraus, daß die ermittelten a-posteriori-Wahrscheinlichkeiten sowohl von den absoluten Wahrscheinlichkeiten, als auch von den likelihoods abhängen. Wie die Diskussion oben gezeigt hat[51], erfordert dies Prüfungsmethoden, deren Ergebnisse sowohl besonders sensitiv als auch sehr spezifisch für die Ordnungsmäßigkeit der Prüfungsgesamtheit sind. Fraglich ist, inwieweit die bestehenden Prüfungsmethoden dieses leisten können. Das Problem soll durch das folgende Zahlenbeispiel verdeutlicht werden. Angenommen a priori wird $pr(O)=0{,}001$ gesetzt, eine sicherlich sehr große Wahrscheinlichkeit angesichts der oben angesprochenen Fülle von Elementarereignissen in Ω, und der Zusammenhang zwischen O und E soll mit $pr(E|O)=0{,}95$ bestimmt werden, dann muß, damit O a posteriori wahrscheinlicher wird als $\overline{O}$, $pr(\overline{E}|\overline{O})>0{,}999\overline{049}$ gelten. Umgekehrt heißt dies, daß das Ergebnis E für den Fall, daß die Prüfungsgesamtheit nicht ordnungsmäßig ist,

---

50 Mit $\mathbb{P}(U)$ wird die Potenzmenge von U, daß heißt die Menge aller Teilmengen von U, bezeichnet.

51 Siehe Kapitel B. I. 1. b) in diesem Teil der Arbeit.

nur mit einer Wahrscheinlichkeit von weniger als $0{,}000\overline{950}$ auftreten darf[52]. Insoweit sind Ansätze, die zu einer höheren absoluten Wahrscheinlichkeit für O führen, diesem Ansatz vorzuziehen.

Die Abbildung von U' in bezug auf die betrachtete Unternehmung im vorliegenden Modell erfolgt durch die im datenerzeugenden System des betrachteten Unternehmens implementierten Funktionen π_X und ω_X. Deren Verhalten ist, wie oben gezeigt wurde[53], zumindest für reale Prüfungssituationen nicht deterministisch hinsichtlich des produzierten Outputs. Eine Interpretation im vorliegenden Prüfungsmodell könnte demnach dahin gehen, in der vom Unternehmen gewählten Implementierung der Funktionen π_X und ω_X die Versuchsanordnung zu sehen, deren theoretische Eigenschaft die den Elementarereignissen in Ω zuzuordnenden Wahrscheinlichkeiten festlegt. Dies ist gleichbedeutend mit der in Teilen des Schrifttums zur Prüfungslehre vertretenen Ansicht, daß aus der Prüfung des datenverarbeitenden Systems die a-priori-Wahrscheinlichkeiten für das dargestellte Entscheidungsmodell abzuleiten seien[54].

Gegen diese Auffassung, zumindest im Rahmen der Verwendung von Wahrscheinlichkeiten im Sinne einer objektiven Interpretation, spricht, daß sie keine empirische Überprüfung zuläßt[55], da für das vorgefundene datenerzeugende System im Grenzfall der Erstprüfung nur eine einzige Abbildung, nämlich die Prüfungsgesamtheit selbst existiert. Auch bei Wiederholungsprüfungen liegt im allgemeinen keine genügend lange Folge von Abbildungen (Ereignissen) vor, die eine Überprüfung objektiver Wahrscheinlichkeiten rechtfertigt, zumal sich die einzelnen vorliegenden Abbildungen auf unterschiedliche Mengen realer Sachverhalte beziehen, so daß die beobachtete Folge von Abbildungen schon aus prinzipiellen Überlegungen nicht zur Überprüfung der Hypothesen über objektive Wahrscheinlichkeiten geeignet ist. Hinzu kommt als weiteres Problem, daß die tatsächliche Ausprägung der Prüfungsgesamtheit im Hinblick auf ihre Ordnungsmäßigkeit regelmäßig nicht bekannt werden wird, so daß auch dieser Umstand eine empirische Überprüfung verhindert.

Der letzte Einwand schließlich, der gegen die Interpretation des datenerzeugenden Systems im Unternehmen als Versuchsanordnung spricht, welche die Ableitung objektiver Wahrscheinlichkeiten erlaubt, richtet sich auf die notwen-

---

52 Dieser Wert entspräche der maximal zulässigen Wahrscheinlichkeit für einen Fehler 2. Art, wenn die eingesetzte Prüfungsmethode ein Alternativtest mit konkretisierter Gegenhypothese ist.

53 Siehe oben Kapitel B. II. 1. im Ersten Teil der Arbeit.

54 Vgl. *Adenauer*, Kontrollsystem, 1989, S. 204f.

55 Vgl. *Wittmann*, Systemprüfung, 1980, S. 229.

dige Unabhängigkeit der einzelnen Ereignisse aus unterschiedlichen Versuchen. Es müßte demnach für den Fall, daß eine wiederholte Bearbeitung der in U' enthaltenen Sachverhalte durch das datenerzeugende System erfolgen könnte, geprüft werden, ob die resultierenden Abbildungen X voneinander statistisch unabhängig sind. Eine physikalische Unabhängigkeit der Versuche, welche die statistische Unabhängigkeit implizieren würde, wäre ausschließlich dann gewährleistet, wenn es möglich wäre, vor jedem Versuch einen identischen Systemzustand herzustellen. Dies gelingt für Systeme mit personellen Aufgabenträgern regelmäßig nicht, da ansonsten die Aufgabenträger die schon verarbeiteten Sachverhalte aus den vorherigen Versuchen vergessen müßten. Bei Systemen aus maschinellen Aufgabenträgern könnte zwar grundsätzlich der ursprüngliche Systemzustand auf der betrachteten Ebene wiederhergestellt werden, allerdings resultiert in diesem Fall die Unschärfe der Prüfungsinformationen, wie oben ausgeführt wurde[56], nicht aus indeterministischem Verhalten des Systems, sondern aus der fehlenden genauen Kenntnis der tatsächlich ausgeführten Funktionen. Diese werden jedoch bei gleichem Ausgangszustand und gleichem Systeminput zu jeweils identischen Abbildungen führen, so daß dann keine Zufallsfolge von Ereignissen, hier Abbildungen entsteht. In den in der Realität anzutreffenden datenerzeugenden Systemen werden Funktionen regelmäßig sowohl von personellen als auch von maschinellen Aufgabenträgern ausgeführt. Insoweit kann wegen der beteiligten personellen Aufgabenträger keine physikalische Unabhängigkeit der Versuche vorliegen. Selbst wenn demnach die Schwierigkeit, längere Ereignisfolgen zu erhalten, überwunden wird, bleibt das Problem des Nachweises der statistischen Unabhängigkeit der einzelnen Ereignisse bestehen, bevor objektive Wahrscheinlichkeiten verwendet werden könnten. Insgesamt bleibt festzustellen, daß, soweit der Möglichkeitsraum durch alle möglichen oder alle durch das datenerzeugende System möglichen Abbildungen von U' gebildet wird, die Verwendung objektiver Wahrscheinlichkeiten nicht gerechtfertigt ist.

Zu einer anderen Interpretation führt die Annahme, daß der Möglichkeitsraum aus allen Abbildungen, die von datenerzeugenden Systemen eines räumlich und zeitlich abgegrenzten Bereichs erzeugt werden, besteht. Die räumliche Abgrenzung würde sich dabei auf solche Systeme beziehen, für die hinsichtlich der für die Abbildung der Realität geltenden Regeln identische Funktionen π_X, ω_X und b_X unterstellt werden könnten. Die zeitliche Abgrenzung müßte in der Weise erfolgen, daß die Bedingungen, die zur Zufallsfolge von Abbildungen führen, als konstant angenommen werden könnten. Bei dieser Interpretation

[56] Siehe oben Kapitel B. II. 1. im Ersten Teil der Arbeit.

bestünde zwar zumindest im Hinblick auf das zu betrachtende Gesamtmaß, das für das Eintreten der einzelnen Umweltzustände aus S anzugeben ist, die Möglichkeit objektive Wahrscheinlichkeiten abzuleiten, allerdings ginge damit eine Veränderung der Fragestellung einher. Die Größe $pr(O)$ würde dann nicht mehr die Wahrscheinlichkeit für die tatsächliche Ordnungsmäßigkeit von X, also das Eintreten von s_1 bezeichnen, sondern die Wahrscheinlichkeit bei der Prüfung eines Unternehmens des betrachteten Bereichs eine ordnungsmäßige Abbildung vorzufinden. Ob diese beiden Wahrscheinlichkeiten gleichgesetzt werden dürfen, kann empirisch nicht geprüft werden, da mit der betrachteten Prüfungsgesamtheit nur eine einzige Realisation vorliegt. Wenn die ursprüngliche Fragestellung beibehalten werden soll, rechtfertigt somit auch dieser Ansatz nicht die Verwendung objektiver Wahrscheinlichkeiten als Risikomaß.

Nachdem deutlich wird, daß es sich bei dem insgesamt zu bestimmenden Risikomaß nicht um ein Wahrscheinlichkeitsmaß im Sinne einer objektiven Interpretation von Wahrscheinlichkeiten handeln kann, soll untersucht werden, inwieweit zumindest einzelne Formen der Unschärfe von Prüfungsinformationen durch objektive Wahrscheinlichkeiten gemessen werden können. Diese Untersuchung ist insbesondere aus zwei Gründen gerechtfertigt. Sollte sich herausstellen, daß es sich bei dem Gesamtmaß um ein subjektives Unschärfemaß handelt, dann liefern die folgenden Überlegungen eventuell Ansätze, um einzelne Ursachen für die Subjektivität auszuschließen. Darüber hinaus bieten sie, soweit im einzelnen objektive Wahrscheinlichkeiten gemessen werden können, Ansätze zur späteren Aufspaltung des Gesamtrisikos auf objektive und subjektive Bestandteile. Im folgenden soll daher, ohne auf die Prüfungsmethoden einzugehen[57], geprüft werden, inwieweit die einzelnen festgestellten Formen von Unschärfe durch objektive Wahrscheinlichkeiten modelliert werden können.

Eine Form der Unschärfe von Prüfungsinformationen, die im Rahmen indirekter Messungen und bei der Verwendung von Stichprobenverfahren mit bewußter Auswahl der Stichprobenelemente auftritt, resultiert aus der nicht uneingeschränkten Gültigkeit von unterstellten Schlußfolgerungen. Diese wurde oben für den Fall eindeutiger Zusammenhänge durch eine logische Implikation abgebildet. Um zu untersuchen, ob diese Form der Unschärfe durch Wahrscheinlichkeiten abgebildet werden kann, müssen die Schlußfolgerungen zunächst als Mengen dargestellt werden. Für die Übertragung auf die men-

[57] Die Anwendung der hier dargestellten Unschärfemaße im Rahmen der einzelnen Prüfungsmethoden erfolgt im Dritten Teil der Arbeit.

gentheoretische Abbildung soll gelten, daß die Wahrheit der Aussagen über den Zustand des Ersatztatbestandes, die oben mit z bezeichnet wurde, die Menge E als Teilmenge von Ω, das hier als die Menge aller möglichen Abbildungen aufgefaßt werden soll, abgrenzt. Wird durch z beispielsweise die Aussage getroffen, daß das datenerzeugende System „zuverlässig" ist, so enthält E alle Abbildungen, die durch ein als „zuverlässig" gekennzeichnetes datenerzeugendes System erzeugt werden können. Bezeichnet O wiederum die Menge der ordnungsmäßigen Abbildungen, so läßt sich die logische Implikation $z \rightarrow o$[58] durch die Relation $E \subseteq O$ mengentheoretisch abbilden, da im Beispiel bei strenger Gültigkeit der durch die Implikation ausgedrückten Gesetzmäßigkeit jede Abbildung, die durch ein „zuverlässiges" datenerzeugendes System erzeugt wird, auch in der Menge der ordnungsmäßigen Abbildungen O enthalten sein muß. Ist demnach bekannt, daß die Prüfungsgesamtheit X in E enthalten ist, also z logisch wahr ist, dann müßte, falls ein Wahrscheinlichkeitsmaß für das Eintreten von O existiert, gelten: $pr(O|E)=1$. Für den Fall, daß keine strenge Gültigkeit der betrachteten Gesetzmäßigkeit unterstellt werden kann, existieren demnach Abbildungen, die Bestandteil von E nicht jedoch von O sind. Wird aber zumindest eine tendenzielle Gültigkeit des Zusammenhangs angenommen, so muß die relative Häufigkeit der ordnungsmäßigen Abbildungen in E größer sein als in Ω. Es muß demnach gelten:

(2.27) $$pr(O) < pr(O|E) \leq 1.$$

Die Modellierung der Unschärfe von Gesetzmäßigkeiten, die bei strenger Gültigkeit als logische Implikationen formuliert wurden, kann demnach grundsätzlich durch die Verwendung von bedingten Wahrscheinlichkeiten zur Messung der Stärke des Zusammenhanges zwischen Vorder- und Hintersatz der Implikation erfolgen. Da die bedingten Wahrscheinlichkeiten für den gleichen Ereigniskörper wie die absoluten Wahrscheinlichkeiten definiert sind, setzt, wie die Definition in (2.7) erkennen läßt, die Existenz objektiver bedingter Wahrscheinlichkeiten voraus, daß objektive absolute Wahrscheinlichkeiten existieren. Da dies für das vorliegende Modell nicht zutrifft, ist zumindest die Verwendung eines objektiven Wahrscheinlichkeitsmaßes zur Abbildung unscharfer Gesetzmäßigkeiten nicht möglich.

---

[58] Da ausschließlich die grundsätzliche Eignung von Wahrscheinlichkeitsmaßen zur Abbildung der beim indirekten Messen auftretenden Unschärfen geprüft werden soll, wird an dieser Stelle auf die im Ersten Teil der Arbeit unterstellte Gesetzmäßigkeit zwischen Ersatztatbestand und Fehleranteil bzw. Fehlerausmaß noch nicht eingegangen.

Eine weitere Form der Unschärfe kann aus der Unvollständigkeit der Prüfungsinformationen resultieren. Unterschiedliche Folgen von Prüfungsergebnissen sind dabei ausschließlich dann zu erwarten, wenn die Stichprobenelemente zufällig ausgewählt werden. Der Möglichkeitsraum Ω wird dann durch alle für den gewählten Stichprobenumfang und das unterstellte Auswahlverfahren[59] möglichen Kombinationen von Stichprobenelementen aus der Prüfungsgesamtheit X gebildet (Stichprobenraum). Als Versuchsanordnung, der die objektiven Wahrscheinlichkeiten als theoretische Eigenschaft zugeschrieben werden können, ist hier das Auswahlverfahren für die Stichprobenelemente zu betrachten. Dieses muß demnach die oben genannten Anforderungen, die an die Versuchsanordnung gestellt werden, erfüllen[60].

Die Definition des Ereigniskörpers ist von der Fragestellung, anhand derer die Stichprobe ausgewertet werden soll, abhängig. Bei der Verwendung von Schätzmethoden enthält der Ereigniskörper $\mathbb{K}$ neben dem sicheren und dem unmöglichen Ereignis die Menge der Stichproben, die bei vorgegebenem Verfahren zur Konstruktion des Konfidenzintervalls aus dem Stichprobenergebnis dazu führen, daß der wahre Parameter der Grundgesamtheit mindestens mit einer vorzugebenden Wahrscheinlichkeit von $1-\alpha$ im Konfidenzintervall enthalten ist, sowie das Komplement dieser Menge. Bei Anwendung von Testmethoden enthält $\mathbb{K}$ neben dem sicheren und dem unmöglichen Ereignis ebenfalls zwei disjunkte Mengen[61]. Die eine Menge entspricht dem Ablehnungsbereich der Alternativhypothese, der mit dem Annahmebereich der Nullhypothese gleichgesetzt wird, die andere Menge wird durch den Ablehnungsbereich der Nullhypothese, der dem Annahmebereich der Alternativhypothese entspricht, gebildet. Die Eintrittswahrscheinlichkeiten für die beiden Mengen können bestimmt werden, sie sind außer von der unterstellten Grenzverteilung auch davon abhängig, welche der beiden Hypothesen jeweils als wahr unterstellt wird. Die bei der Anwendung von Zufallsstichprobenverfahren auftretenden Unschärfen sind, soweit die Sicherheit der getroffenen Aussagen gemeint sind, demnach durch objektive Wahrscheinlichkeiten meßbar, wenn das der Zu-

---

59 Bei Prüfungen erfolgt die Auswahl der Stichprobenelemente regelmäßig ohne Zurücklegen der bereits gezogenen Elemente.

60 Bei der für die Prüfung unterstellten Auswahl der Stichprobenelemente ohne Zurücklegen bereits gezogener Elemente ist die statistische Unabhängigkeit der einzelnen Ziehungen nicht gewährleistet. Diese schwache Abhängigkeit kann jedoch für genügend kleine Auswahlsätze vernachlässigt werden bzw. im Fall dichotomer Grundgesamtheiten durch Verwendung der hypergeometrischen Verteilung an Stelle der Binomialverteilung berücksichtigt werden.

61 Siehe oben Kapitel B. I. 1. im Ersten Teil der Arbeit und die dort in Fn. 43 gemachten Anmerkungen.

fallsauswahl zugrundeliegende Verfahren zumindest prinzipiell einer empirischen Überprüfung zugänglich ist.

Als letzte Form der Unschärfe, für die untersucht werden soll, ob sie durch objektive Wahrscheinlichkeiten gemessen werden kann, werden unscharfe Wertangaben auf metrischen Skalen betrachtet. Diese können zum einen beim Hintersatz der als Majorprämisse verwendeten Implikation für die Prüfung des datenerzeugenden Systems, zum anderen bei der Ableitung des Sollobjektes für einzelne Elemente der Prüfungsgesamtheit und der sich anschließenden Fehlerfeststellung durch den Prüfer auftreten. Die Grundstruktur des Problems ist in allen Fällen identisch: Eine verbale Angabe kann auf einer metrischen Skala zumindest in einem Grenzbereich nicht eindeutig zugeordnet werden. Den Möglichkeitsraum Ω bildet hier demnach die zugrundeliegende Skala, jeder Wert der Skala wird als ein Ereignis betrachtet[62]. Der darüber zu bildende Ereigniskörper $\mathbb{K}$ wird durch geeignet definierte Teilmengen von Ω gebildet. Im vorliegenden Fall handelt es sich dabei um Intervalle der zugrundeliegenden Skala. Neben den Grenzbereichen, in denen keine eindeutige Wertzuordnung auf der zugrundeliegenden Skala erfolgen kann, existieren Bereiche, für die eindeutig erklärt werden kann, ob auf sie die verbale Angabe zutrifft oder nicht. Für die Zuordnung von Wahrscheinlichkeit wäre demnach zu fordern, daß der Menge der Werte, die mit Sicherheit durch die verbale Angabe erfaßt werden, die Wahrscheinlichkeit Eins, der Menge der Werte, die von der Angabe sicher nicht betroffen sind, die Wahrscheinlichkeit Null zugeordnet wird. Für die Grenzbereiche müßte die Wahrscheinlichkeit Werte zwischen Null und Eins annehmen. Wird z. B. für die Angabe „sehr hoher" Fehleranteil festgelegt, daß Fehleranteile von 20 % und mehr unstrittig sehr hoch sind und Fehleranteile, die weniger als 16 % betragen, sicher nicht mehr als sehr hoch eingestuft werden sollen, wäre zu fordern, daß die Wahrscheinlichkeit für das Intervall $[0;0{,}16]$ gleich Null und für das Intervall $[0{,}2;1]$ gleich Eins ist. Während die erste Forderung mit den Axiomen in (2.4) bis (2.6) vereinbar ist, gilt dies für die zweite ausschließlich dann, wenn gemäß (2.6) auch die Wahrscheinlichkeit für den Bereich zwischen 16 % und 20 % Null gesetzt wird. Diese Zuordnung ist hier aber gerade nicht gewünscht.

62 Probleme, die sich aus dem Umstand ergeben, daß der Möglichkeitsraum im vorliegenden Fall nicht mehr endlich ist, wie oben in Kapitel B. I. 1. a) gefordert wird, sollen zunächst nicht beachtet werden. Als Unterschied ergibt sich, daß keine relativen Häufigkeiten mehr zu den einzelnen Ereignissen angegeben werden können, sondern die Wahrscheinlichkeitsdichte als stetige Funktion über dem Möglichkeitsraum definiert ist. Wahrscheinlichkeiten können dann ausschließlich für Fraktile des Möglichkeitsraumes angegeben werden. Vgl. *Fisz*, Wahrscheinlichkeitsrechnung, 1978, S. 54.

DOI https://doi.org/10.3790/978-3-428-48909-1

Ein zweiter schwerwiegender Einwand ergibt sich ebenfalls aus dem in (2.6) formulierten Axiom. Dieses fordert, daß die Wahrscheinlichkeit einer Menge von Ereignissen der Summe der Wahrscheinlichkeiten bei einer Zerlegung in disjunkte Teilmengen entspricht. Im betrachteten Beispiel bedeutet dies, daß die Wahrscheinlichkeit, die den Intervallen $[0,2;0,5]$ und $(0,5;1]$ zugeordnet wird, zumindest für eines der beiden Intervalle kleiner als Eins sein muß. Für beide Intervalle soll aber nach der oben angegebenen Festlegung gelten, daß sie die verbale Angabe sicher erfüllen. Auch hier kann demnach die Struktur der Unschärfe nicht durch das Wahrscheinlichkeitsmaß abgebildet werden. Insgesamt ist festzustellen, daß die hier untersuchte Form der Unschärfe verbaler Angaben zu metrischen Skalen generell nicht durch Wahrscheinlichkeiten gemessen werden kann. Sie wird daher auch im nächsten Abschnitt über die subjektiven Wahrscheinlichkeiten nicht mehr untersucht.

b) Subjektive Wahrscheinlichkeiten

Bei der subjektiven Interpretation von Wahrscheinlichkeitsmaßen wird die Wahrscheinlichkeit eines Ereignisses oder einer Aussage mit dem Grad der Überzeugung, die eine bestimmte Person (Subjekt) über die Wahrheit dieser Aussage oder über den Eintritt des Ereignisses hat, gleichgesetzt[63]. Dabei muß die Wahrscheinlichkeit, die einem Ereignis zugeordnet wird, nicht für alle Personen gleich sein. Die Regeln der Wahrscheinlichkeitstheorie dienen dann der Kombination einzelner Überzeugungsgrade. Insoweit wird bei dieser Interpretation die Wahrscheinlichkeitstheorie als normative[64], nicht wie bei der objektiven Interpretation, als empirische Theorie verstanden. Da unterstellt werden kann, daß Personen ihren Überzeugungsgrad über den Eintritt eines Ereignisses regelmäßig nicht in Form reeller Zahlen zwischen Null und Eins angeben können[65], ist zunächst die Frage zu beantworten, wie und unter welchen weiteren Voraussetzungen subjektive Wahrscheinlichkeiten, die den oben genannten Axiomen in (2.4) bis (2.6) folgen, gemessen werden können, bevor untersucht wird, inwieweit sich das Risiko in der hier unterstellten Prüfungssituation durch subjektive Wahrscheinlichkeiten messen läßt.

Die Quantifizierung subjektiver Wahrscheinlichkeiten setzt voraus, daß der von einer Person einzelnen Ereignissen zugeordnete Überzeugungsgrad be-

---

63 Vgl. *Savage*, Foundations, 1972, S. 3; *Stegmüller*, Wahrscheinlichkeit, 1. Halbband, 1973, S. 110.

64 Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 47.

65 Vgl. *Tversky/Kahneman*, Judgement, 1974, S. 1124-1131.

stimmte Rationalitätskriterien erfüllt. Im folgenden bezeichnet Ω weiterhin die Menge der Möglichkeiten, $\mathbb{K}$ einen darauf definierten Ereigniskörper[66] sowie A, B, und C Elemente von $\mathbb{K}$. Für die subjektive Wahrscheinlichkeit soll ebenfalls pr[67] als Bezeichner verwendet werden. Die Wahrscheinlichkeitsbewertung pr einer Person wird ausschließlich dann als rational bezeichnet, wenn sie zumindest kohärent ist[68], d. h., die folgenden Voraussetzungen für alle $A,B,C \in \mathbb{K}$ erfüllt sind[69]:

(2.28) $$pr(A) \leq pr(B) \vee pr(B) \leq pr(A);$$

(2.29) $$pr(A) \leq pr(B) \wedge pr(B) \leq pr(C) \rightarrow pr(A) \leq pr(C);$$

(2.30) $$pr(\varnothing) \leq pr(A);$$

(2.31) $$(A \cup B) \cap C = \varnothing \rightarrow \left(pr(A) \leq pr(B) \leftrightarrow pr(A \cup C) \leq pr(B \cup C)\right).$$

Die beiden ersten Axiome stellen sicher, daß die subjektiven Wahrscheinlichkeiten der betrachteten Ereignisse schwach transitiv geordnet werden können. Strittig ist hier, ob diese Ordnung tatsächlich über alle Ereignisse aus $\mathbb{K}$ hergestellt werden muß[70], oder ob Angaben nur insoweit erforderlich sind, bis die vollständige Ordnung feststeht[71]. Wird die Theorie der subjektiven Wahrscheinlichkeit als empirische Theorie aufgefaßt, müßte der ersten Auffassung zugestimmt werden. Wird jedoch unterstellt, daß es sich um eine normative Theorie handelt, führt dies zur zweiten Auffassung. Diese kann, da sie rationale Wahrscheinlichkeitsbeurteilungen unterstellt, nicht mit eventuell den Axiomen widersprechenden Beurteilungen umgehen. Insbesondere das Problem, daß ein Ereigniskörper $\mathbb{K}$ über n einzelne Ereignisse 2^n Elemente enthält, deren im Hinblick auf die Axiome (2.28) und (2.29) widerspruchsfreie Ordnung zu überprüfen wäre, macht deutlich, daß die Voraussetzungen, die an eine rationale Wahrscheinlichkeitsbeurteilung gestellt werden, keineswegs trivial sind. Damit ist die Verwendung subjektiver Wahrscheinlichkeiten auf Fälle von Ereigniskör-

---

66 Die Existenz von Ω und $\mathbb{K}$ sei zunächst unterstellt. Wie Ω und $\mathbb{K}$ im vorliegenden Modell abzugrenzen sind, erfolgt in der anschließenden Untersuchung über die Formen von Unschärfe, die durch subjektive Wahrscheinlichkeiten abgebildet werden können.

67 Die Funktion pr ist an dieser Stelle noch keine quantitative, sondern ausschließlich als komparative Wahrscheinlichkeitsbewertung zu verstehen. D. h., es existiert noch keine Zuweisung $pr(A) = r$.

68 Vgl. *Stegmüller*, Wahrscheinlichkeit, 1. Halbband, 1973, S. 396.

69 Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 50. Die dort als Axiom 5 angegebene Bedingung wird hier nicht benötigt, da Ω endlich ist. Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 51.

70 Vgl. *Schneider*, Meßbarkeitsstufen, 1979, S. 100.

71 Vgl. *Schneeweiß*, Bemerkungen, 1977, S. 221.

pern mit geringer Kardinalität beschränkt. Außerdem müssen Verfahren gefunden werden, die in der Lage sind, Widersprüche in den Wahrscheinlichkeitsbeurteilungen aufzulösen.

Das dritte Axiom (2.30) legt fest, daß alle Ereignisse mindestens so wahrscheinlich sind wie das unmögliche Ereignis. Da durch das unmögliche Ereignis alle nicht im Möglichkeitsraum berücksichtigten Ereignisse zusammengefaßt sind, legt das Axiom die Anforderungen fest, die bei der Abgrenzung des Möglichkeitsraumes zu berücksichtigen sind. Bei der späteren Metrisierung der subjektiven Wahrscheinlichkeiten wird, damit die Axiome in (2.4) bis (2.6) eingehalten werden, die Wahrscheinlichkeit für das unmögliche Ereignis Null gesetzt. Daher muß die Abgrenzung von Ω in einer Form, die ex post Überraschungen, d. h. das Auftreten von Ereignissen, die nicht in Ω enthalten sind, ausschließt, erfolgen [72]. Diese Beschränkung ist, wie die folgenden Überlegungen zeigen werden, für die Existenz subjektiver Wahrscheinlichkeiten sehr wichtig. Da die subjektiven Wahrscheinlichkeiten, die einzelne Personen bestimmten Ereignissen zuordnen, immer aus den ihnen zu einem bestimmten Zeitpunkt zur Verfügung stehenden Informationen abgeleitet werden, sind diese als bedingte Wahrscheinlichkeiten in bezug auf die vorliegenden Informationen zu verstehen, die ausschließlich durch neue Informationen nach (2.14) verändert werden dürfen[73]. Da die a-posteriori-Wahrscheinlichkeiten nicht definiert sind, wenn ein Ereignis, dessen absolute Wahrscheinlichkeit Null beträgt, eintritt, müssen durch die Definition von Ω und $\mathbb{K}$ ex post Überraschungen auch in einem engeren Sinn, nämlich als Auftreten von Ereignissen, deren Wahrscheinlichkeit Null ist, ausgeschlossen werden.

Das Unabhängigkeitsaxiom in (2.31) stellt sicher, daß die Einschätzung der Wahrscheinlichkeit von Ereignissen sich nicht ändert, wenn Ereignisse, die mit ihnen keine Elementarereignisse aus Ω gemeinsam haben, in die Beurteilung einbezogen oder aus der Beurteilung ausgeschlossen werden. D. h., die Wahr-

[72] *Schneider* folgert aus dem Umstand, daß aus Vereinfachungsgründen nicht alle Ereignisse bei der Planung berücksichtigt werden können, daß eine Quantifizierung subjektiver Wahrscheinlichkeiten nicht möglich ist, da bei dieser Vorgehensweise ex post Überraschungen im oben angeführten Sinn nicht ausgeschlossen werden könnten. Vgl. *Schneider*, Informationstheorie, 1995, S. 95. Er unterscheidet dabei aber anscheinend nicht zwischen dem Möglichkeitsraum Ω und dem darauf definierten Ereigniskörper $\mathbb{K}$. Vereinfachungen des Modells werden nicht durch die Definition des Möglichkeitsraumes, sondern bei der Festlegung des Ereigniskörpers berücksichtigt. Dabei müssen auch die aus Ω bei der Planung nicht explizit berücksichtigten Ereignisse als ein mögliches Ereignis in $\mathbb{K}$ aufgenommen werden, so daß die Notwendigkeit zu vereinfachen nicht im Widerspruch zur Quantifizierung von subjektiven Wahrscheinlichkeiten steht.

[73] Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 85f.; *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 375.

scheinlichkeitsbeurteilung zweier Ereignisse soll unabhängig von nicht relevanten Ereignissen sein. Diese Form der Unabhängigkeit, wie sie hier gefordert wird, ist als Rationalitätskriterium unbestritten[74].

Werden die vier genannten Axiome erfüllt, so liegt eine qualitative Wahrscheinlichkeitsbeurteilung vor. Fraglich ist jetzt, welche weiteren Anforderungen eingehalten werden müssen, damit eine qualitative Wahrscheinlichkeitsbeurteilung durch quantitative, subjektive Wahrscheinlichkeiten abgebildet werden kann. Die hinzukommende Anforderung richtet sich auf die Struktur des der Wahrscheinlichkeitsbeurteilung zugrundeliegenden Möglichkeitsraumes. Existiert zu jedem $n \geq 1$ eine gleichförmige Zerlegung von Ω in n disjunkte Teilmengen A_i^n, so daß gilt:

$$\forall i \neq j: A_i^n \cap A_j^n = \varnothing \wedge pr\left(A_i^n\right) = pr\left(A_j^n\right) \wedge \bigcup_{i=1}^{n} A_i^n = \Omega, \tag{2.32}$$

und kann für alle $A, B \in \mathbb{K}$, für die $pr(A) < pr(B)$ gilt, ein n und ein r bestimmt werden mit:

$$pr(A) < pr\left(\bigcup_{i=1}^{r} A_i^n\right) \leq pr(B), \tag{2.33}$$

so existiert eine eindeutige quantitative Funktion *pr*, welche die zugrundeliegende qualitative Wahrscheinlichkeitsbeurteilung so abbildet, daß diese die oben angeführten Axiome (2.4) bis (2.6) erfüllt[75]. Die Idee ist demnach, daß die Struktur des Möglichkeitsraumes fein genug ist, um eine genügend kleine Menge von Ereignissen aus Ω zu finden, deren Wahrscheinlichkeit sich als Einheit für die Wahrscheinlichkeitsbeurteilungen der interessierenden Ereignisse aus $\mathbb{K}$ eignet[76]. Erfüllt der zugrundeliegende Möglichkeitsraum Ω die in (2.32) angegebene Bedingung nicht, läßt sich jedoch auch in diesem Fall eine quantitative Funktion *pr*, die allerdings nicht eindeutig sein muß, ableiten[77].

---

[74] Vgl. *Schneider*, Informationstheorie, 1995, S. 84f.; *Krantz* et al., Measurement, 1971, S. 209f.

[75] Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 52f.; *Savage*, Foundations, 1972, S. 34-38.

[76] Vgl. *de Finetti*, Theory, 1974, S. 199f. Diese Interpretation wird auch im Beweis des Metrisierungstheorems (2.33) deutlich, der hier nicht wiederholt werden soll, da die dargestellten Axiome ausschließlich zur Kenntlichmachung der bei der Quantifizierung von subjektiven Wahrscheinlichkeiten auftretenden Probleme dienen. Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 522-525.

[77] Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 57.

Bevor auf die möglichen Verfahren zur Messung subjektiver Wahrscheinlichkeiten eingegangen werden kann, müssen demnach der Möglichkeitsraum Ω und der Ereigniskörper $\mathbb{K}$, die im hier betrachteten Prüfungsmodell einer subjektiven Wahrscheinlichkeitsbeurteilung zugrundeliegen, definiert werden. Da ex post Überraschungen aus den oben genannten Gründen ausgeschlossen werden müssen, ist der Möglichkeitsraum weit zu definieren. Da letztendlich eine Wahrscheinlichkeitsaussage über den tatsächlichen Zustand der Prüfungsgesamtheit gefordert ist, soll Ω aus allen möglichen Abbildungen von U', deren Anzahl aus Gründen der formalen Einfachheit als endlich unterstellt wird, bestehen.

Nach den Ausführungen zur objektiven Wahrscheinlichkeitsinterpretation verbleibt für die Risikomessung durch Wahrscheinlichkeiten ausschließlich die eventuell bei Gesetzmäßigkeiten, die für die Festlegung des Auswahlkriteriums bei bewußter Auswahl der Stichprobenelemente oder im Rahmen indirekter Messungen herangezogen werden, auftretende Unschärfe. Die in Frage kommenden Gesetzmäßigkeiten enthalten regelmäßig Verknüpfungen von Merkmalen der einzelnen Elementarereignisse aus Ω. Im einfachsten Fall sind diese z. B. von der Art: „Wenn eine Abbildung von einer bestimmten Form eines datenerzeugenden Systems generiert wurde, dann ist sie meistens ordnungsmäßig." Demnach muß der Ereigniskörper $\mathbb{K}$, dessen Elemente der Wahrscheinlichkeitsbeurteilung unterzogen werden, neben den schon definierten Mengen O und $\overline{O}$ alle Mengen enthalten, die durch die in den verwendeten Gesetzmäßigkeiten vorkommenden Merkmale gebildet werden, sowie die Mengen, die notwendig sind, daß die an $\mathbb{K}$ gestellten Bedingungen (2.1) bis (2.3) erfüllt sind[78].

Für die praktische Ableitung von subjektiven Wahrscheinlichkeiten dürfte der Weg über das in (2.33) gegebene Axiom ähnlich problematisch sein wie die direkte Angabe quantitativer Wahrscheinlichkeiten. Es treten dabei insbesondere zwei Probleme auf. Zum einen dürfte die Struktur des hier unterstellten Möglichkeitsraumes zwar grundsätzlich fein genug sein, um den Anforderungen in (2.32) zu genügen, dem Prüfer dürfte es aber regelmäßig nicht möglich sein, die gleichwahrscheinlichen Teilmengen für ein genügend großes n zu identifizieren. Die Annahme, daß alle Elementarereignisse in Ω gleich wahrscheinlich sind, führt zu den schon oben gezeigten Problemen, die aus einer absoluten Wahrscheinlichkeit für O sehr nahe bei Null resultieren[79]. Zum anderen

---

78 Eine genauere Abgrenzung dieser Mengen erfolgt im Dritten Teil der Arbeit.

79 Siehe oben in Kapitel B. I. 2. a) in diesem Teil der Arbeit.

dürfte, selbst wenn unterstellt wird, daß sich die Zerlegung von Ω durch den Prüfer durchführen läßt, der Prüfer spätestens bei der Bestimmung des relevanten Wertes von r in (2.33) für jedes Ereignis überfordert sein, wenn man davon ausgeht, daß ihm auch direkte Zahlenangaben zu subjektiven Wahrscheinlichkeiten nicht möglich sind.

Diese Probleme treten dann in den Hintergrund, wenn die aktuelle Wahrscheinlichkeitsbeurteilung nicht wie bisher als absolute Wahrscheinlichkeit, sondern als bedingte Wahrscheinlichkeit bei gegebenen sicheren Erkenntnissen, die aus den Erfahrungsdaten des Prüfers bestehen können, verstanden wird. Wird weiterhin gefordert, daß eine rationale Wahrscheinlichkeitsbeurteilung ausschließlich dann geändert werden darf, wenn neue Informationen gewonnen werden[80], so läßt sich zeigen, daß für einen genügend großen Umfang von Erfahrungsdaten die subjektiven Wahrscheinlichkeiten weitestgehend von diesen Erfahrungsdaten, insbesondere von den dabei beobachteten relativen Häufigkeiten abhängen[81]. Diese Anforderung an eine rationale Wahrscheinlichkeitsbeurteilung impliziert allerdings, daß die Person, welche die Wahrscheinlichkeitsbeurteilung vornimmt, über ein perfektes Gedächtnis verfügt[82]. Auch wenn die Unterstellung eines rational urteilenden Prüfers im vorliegenden Modell gerechtfertigt erscheint, dürfte die Annahme eines Prüfers mit perfektem Gedächtnis zu realitätsfern sein, um dem vorliegenden Prüfungsmodell praktische Relevanz zu ermöglichen. Soweit allerdings Erfahrungsdaten dokumentiert sind, also z. B. empirische Erhebungen zu bestimmten Gesetzmäßigkeiten vorliegen, können die dort ermittelten relativen Häufigkeiten als subjektive Wahrscheinlichkeiten verwendet werden. Liegen demnach Beobachtungen für Unternehmen eines räumlich und zeitlich abgegrenzten Bereichs mit identischen Abbildungsregeln vor, so können diese, soweit der Prüfer ihre Anwendbarkeit in der vorliegenden Prüfungssituation für zulässig hält, zur Ermittlung subjektiver Wahrscheinlichkeiten verwendet werden.

Ein anderer Zugang zur Messung von subjektiven Wahrscheinlichkeiten geht von der Idee aus, daß sich der Überzeugungsgrad für das Eintreten eines bestimmten Ereignisses, den eine Person hat, aus den Entscheidungen dieser Person ablesen lassen müßte[83]. Dies führt zu der Idee, subjektive Wahrscheinlichkeiten für Ereignisse in $\mathbb{K}$ aus einem System hypothetischer Wetten (Versicherungen), die eine Person auf den Eintritt der Ereignisse abschließen wür-

80 Vgl. *Stegmüller*, Wahrscheinlichkeit, 1. Halbband, 1973, S. 401.

81 Vgl. *v. Kutschera*, Wissenschaftstheorie, 1972, S. 85f.

82 Vgl. *Stegmüller*, Wahrscheinlichkeit, 1. Halbband, 1973, S. 402.

83 Vgl. *de Finetti*, Theory, 1974, S. 76.

de, abzuleiten[84]. Die einfachste Form eines solchen Wettsystems besteht aus einer Wette mit dem Einsatz e_A auf das Ereignis A aus $\mathbb{K}$ und einem Gewinn von g, wenn A eintritt, und einer zweiten Wette mit dem Einsatz $e_{\overline{A}}$ und ebenfalls einem Gewinn von g, wenn das komplementäre Ereignis $\overline{A}$ eintritt. Der Quotient aus Einsatz und Gewinn soll als Wettquotient bezeichnet werden. Wird unterstellt, daß die Einsätze direkt subjektive Nutzenwerte repräsentieren, so wird die erste Wette nur dann abgeschlossen, wenn gilt[85]:

$$pr(A) \geq \frac{e_A}{g}. \tag{2.34}$$

Sind die einzelnen Wetten „fair" in dem Sinn, daß die betrachtete Person auch die Position des Wettpartners einnehmen würde[86], sind also die Wettquotienten maximal, so muß gelten:

$$pr(A) = \frac{e_A}{g} \text{ und } pr(\overline{A}) = \frac{e_{\overline{A}}}{g}. \tag{2.35}$$

Ein Wettsystem, das auf Wahrscheinlichkeitsbeurteilungen beruht, die der in (2.35) genannten Bedingung gehorchen, bei dem somit die Wettquotienten den subjektiven Wahrscheinlichkeiten entsprechen, ist kohärent, wenn ein sicherer Verlust, d. h. ein Verlust, der unabhängig vom tatsächlich eintretenden Ereignis entsteht, ausgeschlossen ist[87]. Außerdem gilt, daß eine subjektive Wahrscheinlichkeitsbeurteilung die Axiome in (2.28) bis (2.31) und (2.33) dann erfüllt, wenn kein auf diesen Wahrscheinlichkeitsbeurteilungen beruhendes Wettsystem existiert, das zu einem sicheren Verlust führt[88]. Dies läßt sich in dem hier angeführten, einfachen Wettsystem leicht zeigen, wenn die Gewinne für die Situationen, in denen A bzw. $\overline{A}$ eintritt, wie in Abb. 2.1 zusammengefaßt werden.

Der Gewinn bei dem hier unterstellten Wettsystem ist unabhängig vom tatsächlich eintretenden Ereignis. Damit das Wettsystem kohärent ist, muß der Gewinn also in beiden Fällen mindestens Null sein. D. h., es muß gelten:

---

[84] Vgl. *de Finetti*, Theory, 1974, S. 85f.; *v. Kutschera*, Wissenschaftstheorie, 1972, S. 68; *Stegmüller*, Wahrscheinlichkeit, 1. Halbband, 1973, S. 394.

[85] Vgl. *Keynes*, Treatise, 1963, S. 22; *v. Kutschera*, Wissenschaftstheorie, 1972, S. 69.

[86] Vgl. *de Finetti*, Theory, 1974, S. 81.

[87] Vgl. *de Finetti*, Theory, 1974, S. 87.

[88] Vgl. *de Finetti*, Theory, 1974, S. 191f.; *v. Kutschera*, Wissenschaftstheorie, 1972, S. 70; *Stegmüller*, Wahrscheinlichkeit, 1. Halbband, 1973, S. 396.

$$g - e_A - e_{\overline{A}} = g\left(1 - \frac{e_A}{g} - \frac{e_{\overline{A}}}{g}\right) \geq 0. \tag{2.36}$$

Werden in (2.36) die Wettquotienten durch die subjektiven Wahrscheinlichkeiten gemäß (2.35) ersetzt, ist zu erkennen, daß die Bedingung in (2.36) ausschließlich dann eingehalten wird, wenn die Summe der Wahrscheinlichkeiten höchstens Eins ergibt. Soweit die Summe der Wahrscheinlichkeiten allerdings kleiner als Eins ist, wäre die Wahrscheinlichkeitsbeurteilung nicht kohärent, da die Position des Wettpartners, welche die betrachtete Person gemäß der in (2.35) zugrundegelegten Definition jederzeit einnehmen würde, dann ein Wettsystem mit sicherem Verlust darstellt. Insoweit ist die Wahrscheinlichkeitsbeurteilung ausschließlich dann kohärent, also rational im oben angesprochenen Sinn, wenn die Summe der Wahrscheinlichkeiten in Übereinstimmung mit (2.5) und (2.6) Eins ergibt.

	A	$\overline{A}$
1. Wette	$g - e_A$	$-e_A$
2. Wette	$-e_{\overline{A}}$	$g - e_{\overline{A}}$
Σ	$g - e_A - e_{\overline{A}}$	$g - e_A - e_{\overline{A}}$

Abb. 2.1: Gewinne bei einem einfachen Wettsystem

Auch wenn mit dem Ansatz, subjektive Wahrscheinlichkeiten über hypothetische Wetten zu bestimmen, die Messung der Wahrscheinlichkeiten praktikabel erscheint, bleiben für das hier betrachtete Modell einige Probleme, die den Anwendungsbereich subjektiver Wahrscheinlichkeiten für die Abbildung der betrachteten Formen von Unschärfe einschränken. Dabei sollen die Einwendungen gegen diesen Ansatz, die sich auf Probleme der Nutzenmessung für die Einsätze beziehen, außer acht gelassen, werden, da diese Probleme auch im unterstellten Entscheidungsmodell ausgeklammert wurden. Es bleiben insbesondere drei Problembereiche. Wie in (2.35) deutlich wird, ist die Angabe des Wettquotienten, zu dem eine Wette eingegangen würde, identisch mit der Angabe der subjektiven Wahrscheinlichkeit für das Ereignis, auf das gewettet wird. Soweit unterstellt wird, daß der Prüfer quantitative Wahrscheinlichkeiten nicht angeben kann, ist fraglich, warum er dann Wettquotienten angeben können soll. Hierbei ist allerdings zu beachten, daß aus der Annahme der Gültigkeit von Axiomen, die rationale Entscheidungen unter Risiko nach dem *Ber-*

noulli-Prinzip ermöglichen, zwingend folgt, daß der Entscheidende subjektive Wahrscheinlichkeiten angeben kann[89]. Dies wird insbesondere deutlich, wenn das Stetigkeitsaxiom, welches besagt, daß der Entscheidende zu zwei Ergebnissen, deren Eintrittswahrscheinlichkeiten bekannt sind, ein Sicherheitsäquivalent angeben kann[90], betrachtet wird. Soweit also die Gültigkeit des Stetigkeitsaxioms unterstellt wird, können aus einer gegebenen Indifferenzsituation auch die subjektiven Wahrscheinlichkeiten abgeleitet werden.

Darüber hinaus wird, um beliebig gewählte Wettquotienten zu vermeiden, gefordert, daß die betrachteten Wetten in endlicher Zeit entschieden werden können[91]. Dies führt im vorliegenden Prüfungsmodell dann zu Problemen, wenn für die Feststellung des Wettergebnisses Aussagen über den tatsächlichen Zustand der Prüfungsgesamtheit X benötigt werden, wie dies bei den zu betrachtenden unscharfen Gesetzmäßigkeiten regelmäßig der Fall sein wird, da diese auf Zusammenhänge zwischen tatsächlichen Parametern der Prüfungsgesamtheit und Ersatztatbeständen und/oder Annahmen über die Realität abstellen. Zwar wurde im zugrundegelegten Entscheidungsmodell angenommen, daß der tatsächliche Zustand von X immer bekannt wird, allerdings handelt es sich dabei um eine Vereinfachung im Modell, um den maximalen Teil des Risikos zu ermitteln, den der Prüfer durch Prüfungshandlungen beeinflussen kann. Für die Ermittlung der Wettquotienten kann diese Vereinfachung nicht gelten, so daß subjektive Wahrscheinlichkeiten ausschließlich für solche Ereignisse bestimmt werden können, deren Eintreten oder Nichteintreten tatsächlich festgestellt werden kann.

Ein dritter möglicher Einwand bei der Verwendung von subjektiven Wahrscheinlichkeiten betrifft die Form, in der Unwissen über einzelne Wahrscheinlichkeiten modelliert wird. Angenommen Ω wird in n disjunkte Teilmengen A_i zerlegt. Ist dem Prüfer ausschließlich die Wahrscheinlichkeit $pr(A_1)$ bekannt, so wird nach dem Satz des unzureichenden Grundes[92] die verbleibende Wahrscheinlichkeit $1-pr(A_1)$ gleichmäßig auf die übrigen Ereignisse verteilt, da sich die Summe der Wahrscheinlichkeiten gemäß (2.5) und (2.6) zu Eins addieren müssen. Es gilt demnach:

$$pr(A_\mathrm{i})=\frac{1-pr(A_1)}{\mathrm{n}-1} \text{ für } \mathrm{i}=2,\ldots,\mathrm{n}. \tag{2.37}$$

---

89 Vgl. z. B. *Fishburn*, Utility Theory, 1970, S. 201f.

90 Vgl. *Luce/Raiffa*, Games, 1957, S. 27.

91 Vgl. *de Finetti*, Theory, 1974, S. 212.

92 Vgl. *Keynes*, Treatise, 1963, S. 375-377.

Wird nun z. B. die Menge A_n in zwei disjunkte Ereignisse A'_n und A'_{n+1} zerlegt, also die neue Zerlegung $A_1,\dots,A_{n-1},A'_n,A'_{n+1}$ betrachtet, so ändern sich, wenn aus den gleichen Gründen wie oben unterstellt wird, daß die verbleibende Wahrscheinlichkeit $1-pr(A_1)$ gleichmäßig auf die verbleibenden Ereignisse zu verteilen ist, die Wahrscheinlichkeiten der Ereignisse $A_2,\dots,A_{n-1}$, obwohl die Ereignisse selbst unverändert geblieben sind. Dies widerspricht insoweit dem Grundsatz, daß bei rationalem Verhalten Wahrscheinlichkeitsbeurteilungen nur durch neue Erfahrungsdaten beeinflußt werden sollen, da fraglich ist, ob die Umstrukturierung der für die Wahrscheinlichkeitsbeurteilung relevanten Ereignisse ein neues Erfahrungsdatum im Sinne des unterstellten Prinzips des „Lernens aus Erfahrung“[93] darstellt.

Die Probleme resultieren insbesondere aus der in (2.6) geforderten Additivität von Wahrscheinlichkeiten, die eine Wahrscheinlichkeitszuordnung erfordert, auch wenn keine Erkenntnisse über den Eintritt des betrachteten Ereignisses vorliegen[94]. Für die Interpretation von Wahrscheinlichkeiten als Wettquotienten bedeutet dies, daß eine Person, die eine Wette auf das Ereignis A mit dem Wettquotienten $pr(A)$ abschließt, zwingend auch eine Wette auf $\overline{A}$ mit dem Wettquotienten $1-pr(A)$ akzeptiert. Der Fall, daß die zweite Wette von der betrachteten Person nicht abgeschlossen würde, also $pr(\overline{A})=0$ gilt, kann bei der Verwendung von Wahrscheinlichkeiten nicht abgebildet werden, auch wenn dieses Verhalten nicht notwendigerweise unvernünftig ist[95].

Der letzte Einwand ist, wie die folgenden Überlegungen zeigen, auch für das hier unterstellte Prüfungsmodell relevant. Wie oben gezeigt wurde[96], besteht prinzipiell die Möglichkeit unscharfe Gesetzmäßigkeiten durch Wahrscheinlichkeiten abzubilden. Allerdings muß dann akzeptiert werden, daß die betrachteten Gesetzmäßigkeiten sowohl die Einschätzung der Wahrscheinlichkeit für die Ordnungsmäßigkeit als auch für die Nichtordnungsmäßigkeit beeinflussen. D. h., falls der in (2.27) unterstellte Zusammenhang zwischen O und E gilt, muß auch akzeptiert werden, daß gilt:

(2.38) $$pr(\overline{O}) > pr(\overline{O}|E) \geq 0.$$

Dies mag für den Fall einleuchten, daß E bedeutet: „Es wurde ein zuverlässiges datenerzeugendes System vorgefunden.“, da diese Information durchaus

---

93 Siehe oben S. 100 und Fn. 80.

94 Vgl. *Shafer*, Theory, 1976, S. 22f.

95 Vgl. *Shafer/Srivastava*, Bayesian, 1990, S. 489; *Schneider*, Informationstheorie, 1995, S. 91.

96 Siehe oben Seite 92 in diesem Teil der Arbeit.

dazu führen kann, daß der Prüfer die Wahrscheinlichkeit für die Nichtordnungsmäßigkeit der Prüfungsgesamtheit jetzt geringer einschätzt als vor Erhalt der Information. Dies muß allerdings nicht bei allen verwendeten Gesetzmäßigkeiten der Fall sein. Falls E bedeutet, daß im Rahmen von analytischen Prüfungshandlungen keine Abweichung zwischen prognostizierter Kenngröße und tatsächlich festgestellter Kenngröße gefunden wurde, könnte zwar tendenziell die in (2.38) angeführte Relation begründet werden, da einige Fehlermöglichkeiten ausgeschlossen werden. Wie oben ausgeführt wurde[97], kann daraus aber keine Aussage über die Wahrscheinlichkeit für die Ordnungsmäßigkeit der Prüfungsgesamtheit abgeleitet werden, so daß die in (2.27) geforderte Konsequenz aus den im Rahmen von analytischen Prüfungshandlungen unterstellten Gesetzmäßigkeiten regelmäßig nicht folgen wird.

Soweit die Modellierung von Unschärfe durch Wahrscheinlichkeiten im vorliegenden Modell auf subjektive Wahrscheinlichkeiten angewiesen ist, erscheinen die Anforderungen, die an eine kohärente und konsistente Wahrscheinlichkeitsbeurteilung zu stellen sind, insgesamt als zu streng[98]. Im folgenden Abschnitt soll daher ein Unschärfemaß, welches das Axiom in (2.6) abschwächt, und die sich daraus für die Risikomessung ergebenden Konsequenzen dargestellt werden.

II. Glaubwürdigkeits– und Plausibilitätsmaß

1. Grundbegriffe

a) Axiome

Wie im vorigen Abschnitt über das Wahrscheinlichkeitsmaß soll die Existenz eines Möglichkeitsraumes Ω und eines darüber definierten Ereigniskörpers $\mathbb{K}$ vorausgesetzt werden[99]. Unter einem Glaubwürdigkeitsmaß

---

97 Siehe oben Kapitel B. II. 2. im Ersten Teil der Arbeit.

98 Siehe auch die Auswertung empirischer Studien zu individuellen Schätzleistungen subjektiver Wahrscheinlichkeiten bei *Gans*, Prüfungen, 1986, S. 162f.

99 *Shafer* definiert sein Maß für die Glaubwürdigkeit von Ereignissen nicht, wie hier unterstellt werden soll, auf einem Möglichkeitsraum, der alle irgendwie möglichen Zustände der Welt beinhaltet, sondern auf einem „frame of discernment". Dieser beinhaltet die Menge der Ereignisse, die bei gegebenem Wissenstand von der die Glaubwürdigkeitszuordnung vornehmenden Person unterschieden werden können. Der für die Glaubwürdigkeitszuordnung relevante Ereigniskörper $\mathbb{K}$ entspricht demnach der Potenzmenge des frame of discernment. Vgl. *Shafer*, Theory, 1976, S. 36. Auch wenn die Annahmen *Shafers* eher der Realität entsprechen, soll für die formale Darstellung die Existenz eines Möglichkeitsraumes weiter unterstellt werden. Die Konsequenzen werden unten in Kapitel B. II. 2. in diesem Teil der Arbeit diskutiert.

$bel: \mathbb{K} \to [0;1]$[100] wird dann jedes Maß verstanden, das den folgenden Axiomen gehorcht[101]:

(2.39) $$bel(\varnothing) = 0;$$

(2.40) $$bel(\Omega) = 1;$$

(2.41) $$\forall A, B \in \mathbb{K}: bel(A \cup B) \geq bel(A) + bel(B) - bel(A \cap B).$$

Durch $bel(A) = \mathrm{r}$ wird demnach ausgedrückt, daß der Eintritt des Ereignisses A im Maße r für glaubwürdig gehalten wird. Entsprechend kann der Zweifel am Eintritt des Ereignisses A durch die Glaubwürdigkeit seines Komplements, also $bel(\overline{A})$ ausgedrückt werden. Der Grad, in dem an A nicht gezweifelt wird, wird mit $pl(A)$ bezeichnet, und drückt aus, wie plausibel das Eintreten von A ist. Nach dem bisher Gesagten gilt demnach[102]:

(2.42) $$pl(A) = 1 - bel(\overline{A}).$$

Da aus (2.41) abgeleitet werden kann, daß

(2.43) $$bel(A) + bel(\overline{A}) \leq 1$$

ist[103], muß folgende Bedingung immer gelten[104]:

(2.44) $$pl(A) \geq bel(A).$$

D. h., die Plausibilität eines Ereignisses ist mindestens so groß wie seine Glaubwürdigkeit[105]. Wegen (2.44) wird $pl(A)$ auch als obere Wahrscheinlichkeit von A bezeichnet[106]. Wie sich zeigen läßt[107], ist außerdem jedes Wahrscheinlichkeitsmaß auch ein spezielles Glaubwürdigkeitsmaß, für das gilt:

(2.45) $$pr(A) = pl(A) = bel(A).$$

Wegen (2.43) wird bei der Verwendung eines Glaubwürdigkeitsmaßes zur Risikomessung das oben angeführte Wahrscheinlichkeitskonzept um die Mög-

---

100 *bel* steht für „belief".

101 Vgl. *Shafer*, Theory, 1976, S. 39. Dabei wurde das dritte Axiom zur besseren Vergleichbarkeit mit (2.6) nur für den Fall der Vereinigung zweier Mengen angeführt. Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 159.

102 Vgl. *Shafer*, Theory, 1976, S. 43.

103 Vgl. *Klir/Folger*, Fuzzy Sets, 1988, S. 111.

104 Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 159.

105 Vgl. *Dubois/Prade*, Possibility, 1988, S. 120.

106 Vgl. *Shafer*, Theory, 1976, S. 43.

107 Vgl. *Shafer*, Theory, 1976, S. 54f.

lichkeit erweitert, Unkenntnis über die Glaubwürdigkeit des Eintritts eines Ereignisses A explizit durch die Zuweisung $bel(A)=0$ auszudrücken. Dadurch besteht zumindest prinzipiell auch die Möglichkeit, Gesetzmäßigkeiten in das Modell aufzunehmen, die ausschließlich Kenntnis über den Einfluß auf die Glaubwürdigkeit eines beliebigen Ereignisses A ausdrücken, ohne daß sie Veränderungen der Glaubwürdigkeit von $\overline{A}$ implizieren.

b) Kombination von Glaubwürdigkeitseinschätzungen

Nachdem im vorigen Abschnitt der Glaubwürdigkeits- und Plausibilitätsgrad in Form von Axiomen eingeführt wurde, muß, um die Kombination unterschiedlicher Glaubwürdigkeitszuordnungen zu ermöglichen, zunächst analysiert werden, wie sich die einem einzelnen Ereignis A insgesamt zugeordnete Glaubwürdigkeit $bel(A)$ zusammensetzt. Aus dem in (2.41) angegebenen Axiom für das Glaubwürdigkeitsmaß folgt, daß Fälle existieren können, in denen ein bestimmter Teil der einer Menge von Ereignissen insgesamt zugeordneten Glaubwürdigkeit nur dieser Menge selbst zugeordnet und nicht auf ihre Teilmengen verteilt werden kann. Dieser Teil der Glaubwürdigkeit einer Menge A, der ausschließlich A selbst zugeordnet werden kann, wird „grundlegende Wahrscheinlichkeitszuordnung" auf A genannt und soll mit $m(A)$ bezeichnet werden[108]. Wird die grundlegende Wahrscheinlichkeitszuordnung der aus den beiden unterschiedlichen, nicht leeren Mengen A und B gebildeten Vereinigungsmenge betrachtet, muß also gelten:

(2.46) $$m(A \cup B) = bel(A \cup B) - bel(A) - bel(B) + bel(A \cap B).$$

Falls ein Glaubwürdigkeitsmaß den Axiomen (2.39) bis (2.41) gehorcht, gilt für die grundlegende Wahrscheinlichkeitszuordnung [109]:

(2.47) $$m(\varnothing) = 0;$$

(2.48) $$\sum_{\forall A \in \mathbb{K}} m(A) = 1.$$

[108] Der Unterschied zu dem in Kapitel B. I. behandelten Wahrscheinlichkeitsmaß *pr* besteht darin, daß es Axiom (2.6) erforderlich macht, daß jede Wahrscheinlichkeit, die einer Menge A, die aus mehr als einem Element besteht, zugewiesen wird, genau auf die einzelnen Elemente, aus denen A sich zusammensetzt, zugeordnet werden kann. Diese Zuordnung auf einzelne Teilmengen verbietet die hier betrachtete grundlegende Wahrscheinlichkeitszuordnung definitionsgemäß. Vgl. *Srivastava/Shafer*, Belief-Function, 1992, S. 256.

[109] Vgl. *Shafer*, Theory, 1976, S. 38.

Es handelt sich demnach bei m nicht um ein Wahrscheinlichkeitsmaß auf Ω, sondern auf $\mathbb{K}$[110]. Die gesamte Glaubwürdigkeit $bel(A)$ für eine Menge von Ereignissen A ergibt sich als Summe der Glaubwürdigkeiten ihrer Teilmengen, da diese implizieren, daß auch A gilt, und der Glaubwürdigkeit, die ausschließlich A zugeordnet werden kann[111]:

$$(2.49) \qquad bel(A) = \sum_{\forall B \subseteq A} m(B).$$

Wird die Menge aller Ereignisse A, für die bei einer gegebenen Glaubwürdigkeitsbeurteilung $m(A) > 0$ gilt, mit $\mathbb{E}$ bezeichnet, so kann eine bestimmte Glaubwürdigkeitszuordnung auf $\mathbb{K}$ auch durch das Tupel $\langle \mathbb{E}, m \rangle$ beschrieben werden. Das Tupel $\langle \mathbb{E}, m \rangle$ wird auch als „body of evidence" bezeichnet[112], da es sämtliche Informationen, die zur Bestimmung der Glaubwürdigkeitsziffern für alle Elemente aus $\mathbb{K}$ notwendig sind, enthält.

Existieren zwei voneinander unabhängige[113] Glaubwürdigkeitszuordnungen bel_1 und bel_2 zu Ω und damit auch $\langle \mathbb{E}_1, m_1 \rangle$ und $\langle \mathbb{E}_2, m_2 \rangle$ mit $m_1(A_i) > 0$ $(i = 1, \ldots, k)$ und $m_2(B_j) > 0$ $(j = 1, \ldots, l)$, so lassen sich diese nach „*Dempster's* rule of combination" zusammenfassen[114]. Die gemeinsame grundlegende Wahrscheinlichkeitszuordnung für alle $A \in \mathbb{K}$, wenn $A \neq \varnothing$ gilt, ergibt sich nach:

$$(2.50) \qquad m(A) = \frac{\sum_{\forall A_i, \forall B_j : A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{\forall A_i, \forall B_j : A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j)},$$

wenn

$$(2.51) \qquad \sum_{\forall A_i, \forall B_j : A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j) < 1$$

---

110 Vgl. *Dubois/Prade*, Properties, 1987, S. 162.

111 Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 156.

112 Vgl. *Dubois/Prade*, Properties, 1987, S. 163.

113 Die Unabhängigkeit von Glaubwürdigkeitszuordnungen wird, analog der physikalischen Unabhängigkeit von Versuchen bei der Verwendung objektiver Wahrscheinlichkeiten, eher intuitiv bestimmt. Sie ist dann gegeben, wenn die Quellen, aus denen die in den Glaubwürdigkeitszuordnungen zusammengefaßten Erkenntnisse stammen, voneinander unabhängig sind. Vgl. *Dempster*, Probabilities, 1967, S. 335; *Shafer*, Probability, 1987, S. 9; *Halpern/Fagin*, View, 1992, S. 289. Eine notwendige aber nicht hinreichende Voraussetzung für die Unabhängigkeit zweier Glaubwürdigkeitszuordnungen ist dann, daß für alle A_i und B_j gemäß (2.18) $m(A_i \cap B_j) = m_1(A_i) m_2(B_j)$ gilt. Vgl. *Klir/Folger*, Fuzzy Sets, 1988, S. 117.

114 Vgl. *Dempster*, Probabilities, 1967, S. 335-337; *Shafer*, Theory, 1976, S. 60; *Klir/Folger*, Fuzzy Sets, 1988, S. 114.

gilt[115]. Die Glaubwürdigkeit für alle Ereignisse aus $\mathbb{K}$ kann dann nach (2.49) bestimmt werden, ihre Plausibilität erhält man nach (2.42). Falls die Bedingung (2.51) nicht erfüllt wird, existiert das gemeinsame Glaubwürdigkeitsmaß nicht, da sich die zugrundegelegten Glaubwürdigkeitszuordnungen vollständig widersprechen.

Die Normierung von $m(A)$ in (2.50) wird notwendig, da gemäß (2.47) der leeren Menge keine grundlegende Wahrscheinlichkeit zugeordnet wird. Darüber hinaus erhält man mit:

(2.52) $$\frac{1}{1-\sum\limits_{\forall A_i,\forall B_j:A_i\cap B_j=\varnothing} m_1(A_i)m_2(B_j)}$$

ein Maß für den Widerspruch zwischen den beiden Glaubwürdigkeitszuordnungen bel_1 und bel_2[116]. Wenn daher im weiteren von miteinander verträglichen Erkenntnissen gesprochen wird, soll ausgedrückt werden, daß die Bedingung in (2.51) erfüllt ist.

Sollen mehr als zwei Glaubwürdigkeitszuordnungen zusammengefaßt werden, so kann dies durch wiederholte Anwendung von (2.50) in beliebiger Reihenfolge erfolgen. Falls die Bedingung (2.51) in einer Ebene der Zusammenfassung nicht erfüllt wird, existiert keine Reihenfolge, in der dieser Fall vermieden werden kann[117].

c) Bedingte Glaubwürdigkeit

Analog der Vorgehensweise für Wahrscheinlichkeiten lassen sich auch bedingte Glaubwürdigkeiten angeben. Gesucht ist demnach die Glaubwürdigkeit $bel(A|B)$ für eine Menge von Ereignissen $A \in \mathbb{K}$, wenn bekannt ist, daß ein Ereignis $B \in \mathbb{K}$ schon eingetreten ist. Die ursprüngliche Glaubwürdigkeitszuordnung für A soll mit $bel_1(A)$ und die zugehörige grundlegende Wahrscheinlichkeitszuordnung mit m_1 bezeichnet werden. Für ein sicheres Ereignis B erhält man $m_2(B)=1$ und damit $bel_2(B_j)=1$ für alle B_j, für die $B \subseteq B_j$ gilt. Bedingte Glaubwürdigkeiten können demnach als ein Spezialfall der Kombination von Glaub-

115 Vgl. *Shafer*, Theory, 1976, S. 60.
116 Vgl. *Shafer*, Theory, 1976, S. 65.
117 Vgl. *Shafer*, Theory, 1976, S. 62f.

würdigkeitseinschätzungen formuliert werden. Nach (2.50) und (2.49) gilt dann[118]:

(2.53) $$bel(A|B)=\frac{bel_1(A\cup\overline{B})-bel_1(\overline{B})}{1-bel_1(\overline{B})} \text{ für } bel_1(\overline{B})<1.$$

Die Bedingung, unter der die bedingte Glaubwürdigkeit existiert, kann auch als $pl_1(B)>0$ gelesen werden und ist insofern intuitiv einsichtig, da die bedingte Glaubwürdigkeit eines Ereignisses A gegeben B nur dann existieren kann, wenn das Eintreten von B nach den bisherigen Erkenntnissen zumindest plausibel ist. Nach (2.42) erhält man den zu $bel(A|B)$ gehörenden bedingten Plausibilitätsgrad:

(2.54) $$pl(A|B)=\frac{pl_1(A\cap B)}{pl_1(B)} \text{ für } pl_1(B)>0.$$

Die Übereinstimmung von (2.54) und (2.7) ist offensichtlich. Da außerdem das Wahrscheinlichkeitsmaß pr ein spezielles Glaubwürdigkeitsmaß und (2.54) die Folgerung aus einem speziellen Fall von „*Dempster's* rule of combination" ist, kann das Theorem von *Bayes* als Spezialfall der Kombination von Glaubwürdigkeitszuordnungen aufgefaßt werden[119].

2. *Interpretation*

Nachdem bisher ausschließlich die formale Spezifikation eines Glaubwürdigkeits- und Plausibilitätsmaßes dargestellt wurde, soll im folgenden eine Interpretation dieser Maße im Hinblick auf das hier vorgestellte Prüfungsmodell und die darin festgestellten Formen von Unschärfe versucht werden.

a) Abbildung unscharfer Prüfungsinformationen

Da gemäß (2.45) das Wahrscheinlichkeitsmaß pr ein spezielles Glaubwürdigkeitsmaß ist, müssen ausschließlich die Formen von Unschärfe untersucht werden, die bisher nicht durch Wahrscheinlichkeiten abgebildet werden konnten. Dies sind zum einen bestimmte unscharf formulierte Gesetzmäßigkeiten, die im Rahmen des indirekten Messens oder bei der Überprüfung der Voraussetzungen für bestimmte Auswahlmethoden bei der bewußten Auswahl von Stichprobenelementen Verwendung finden. Zum anderen ist noch keine Möglichkeit der Abbildung von unscharfen, verbalen Angaben zu metrisch skalier-

---

[118] Vgl. *Dempster*, Probabilities, 1967, S. 334; *Shafer*, Theory, 1976, S. 67.

[119] Vgl. *Shafer/Srivastava*, Bayesian, 1990, S. 510-512.

ten Größen gefunden. Da die letzteren, wie oben gezeigt wurde[120], ein Maß erfordern, das es zuläßt, daß mehreren echten Teilmengen des Möglichkeitsraumes Ω der Wert Eins zugeordnet wird, dieses aber gegen (2.41) verstößt, kommt eine Messung dieser Form von Unschärfe auch nicht durch ein Glaubwürdigkeitsmaß in Betracht. Für die unscharf formulierten Gesetzmäßigkeiten wurde angenommen, daß diese ausschließlich die Glaubwürdigkeit eines Ereignisses A verändern, ohne gleichzeitig die Glaubwürdigkeit des Komplements $\overline{A}$ zu beeinflussen. Wie schon angedeutet wurde, lassen die Axiome in (2.39) bis (2.41) die Abbildung solcher Gesetzmäßigkeiten prinzipiell zu. Um die Gründe für diesen Unterschied zum Wahrscheinlichkeitsmaß deutlich zu machen, müssen zunächst die Unterschiede zwischen dem Möglichkeitsraum, über dem pr definiert ist, und dem frame of discernment aufgezeigt werden.

Bisher wurde für die Beschreibung des Glaubwürdigkeitsmaßes unterstellt, daß ein Möglichkeitsraum, der alle möglichen Zustände der (Modell-)Welt enthält existiert. Wie schon oben mehrfach betont wurde[121], kennt der Prüfer nicht alle Elemente von Ω, somit kann er weder alle Elemente voneinander unterscheiden, noch kann er sie zumindest hinsichtlich ihrer Glaubwürdigkeit vollständig unterscheiden. Insoweit stimmt der Möglichkeitsraum nicht mit dem frame of discernment, auf dem das Glaubwürdigkeitsmaß üblicherweise definiert wird, überein. Allerdings wird für den bisher zum Glaubwürdigkeitsmaß betrachteten Ereigniskörper regelmäßig $\mathbb{K} \subset \mathbb{P}(\Omega)$ gelten. Wird jetzt das Element aus $\mathbb{K}$ mit paarweise disjunkten Elemente aus Ω, dessen Kardinalität maximal ist, mit Θ bezeichnet, so entspricht Θ dem frame of discernment. $\mathbb{K}$ ist wegen der Bedingungen (2.1) bis (2.3) die Potenzmenge von Θ. Die Existenz des Möglichkeitsraumes stellt für das hier betrachtete Modell somit eine notwendige Bedingung dar, um gegebenenfalls eine Verfeinerung von Θ zu ermöglichen[122]. Diese erzeugt aus $\mathbb{K}$ einen Ereigniskörper $\mathbb{K}'$, über dem alle zu kombinierenden Glaubwürdigkeitszuordnungen definiert sind. Die Existenz des Möglichkeitsraumes soll also nicht die Forderung einschließen, daß der Prüfer alle Elemente aus Ω kennt bzw. unterscheiden kann, sondern ausschließlich die notwendigen Verfeinerungen des frame of discernment garantieren.

Der Unterschied zwischen Möglichkeitsraum Ω und frame of discernment Θ ist allerdings nicht nur formaler Natur. Dahinter verbergen sich zwei unterschiedliche Ansätze, wie neue Erkenntnisse (Evidenz) im Modell von Glaub-

---

[120] Siehe oben in Kapitel B. I. 2. a) auf S. 95.

[121] Siehe oben S. 99 und Fn. 99 in diesem Teil der Arbeit.

[122] Zum Begriff der Verfeinerung eines frame of discernment vgl. *Shafer*, Theory, 1976, S. 115f.

würdigkeits- bzw. Wahrscheinlichkeitszuordnungen verarbeitet werden. Die Verwendung eines Wahrscheinlichkeitsmaßes erfordert, daß bei der Festlegung des relevanten Ereigniskörpers entweder alle möglichen Ereignisse und die Zusammenhänge unter den einzelnen Ereignissen, soweit sie deren Wahrscheinlichkeitsbeurteilung betreffen, gewußt werden, oder daß die neuen Erkenntnisse auch die Zusammenhänge mit allen Ereignissen, die im bisher verwendeten Ereigniskörper enthalten sind, beinhalten[123]. Dies wird anhand der folgenden formalen Argumentation deutlicher. Ausgangspunkt sei ein Ereigniskörper $\mathbb{K}_1$ und damit ein zugehöriger frame of discernment Θ_1. Werden neue Prüfungsinformationen gewonnen, die über dem Ereigniskörper $\mathbb{K}_2$ und damit Θ_2 definiert sind, dann ist, soweit vorausgesetzt werden kann, daß die Erkenntnisse miteinander verträglich sind, die Kombination dieser Erkenntnisse auf einer Verfeinerung der beiden frames of discernment $\Theta = \Theta_1 \times \Theta_2$ definiert[124]. Der entstehende Ereigniskörper $\mathbb{K}$, für den alle Wahrscheinlichkeiten bestimmbar sein müssen, ist dann definiert durch:

(2.55) $$\mathbb{K} = \mathbb{P}(\Theta) = \mathbb{K}(\{(x_1, x_2) | \forall x_1 \in \Theta_1, \forall x_2 \in \Theta_2\}).$$

Da eine Wahrscheinlichkeitszuordnung gemäß (2.28) erfordert, daß grundsätzlich für jedes Element des zugrundegelegten Ereigniskörpers eine Wahrscheinlichkeitsbeurteilung erfolgen kann, müssen die neuen Erkenntnisse nicht nur die Wahrscheinlichkeiten für die Elemente in Θ_2, sondern auch die notwendigen Informationen, um diese Wahrscheinlichkeiten auf die in Θ enthaltenen Elemente zu verteilen, beinhalten. Betrachtet man als einfaches Beispiel:

(2.56) $$\Theta_1 = \{x_1; x_2\} \text{ und } \Theta_2 = \{y_1; y_2\}$$

mit den zugehörigen Wahrscheinlichkeitsmaßen pr_1 und pr_2, so genügt es für die Kombination der darin enthaltenen Erkenntnisse bei Verwendung einer Wahrscheinlichkeitszuordnung nicht, wenn z. B.

(2.57) $$pr_2(\{y_1\}) = pr(\{(x_1; y_1); (x_2; y_1)\})$$

bekannt ist. Es müssen darüber hinaus aus auch

(2.58) $$pr(\{(x_1; y_1)\}) \text{ und } pr(\{(x_2; y_1)\})$$

bestimmt werden können. Während also bei Verwendung einer Wahrscheinlichkeitsbeurteilung die Anzahl der Wahrscheinlichkeitszuordnungen, die in den neuen Erkenntnissen enthalten sein müssen, $|\Theta_1| \cdot |\Theta_2|$ beträgt, genügen bei

[123] Vgl. *Shafer*, Theory, 1976, S. 120.

[124] Vgl. *Shafer*, Theory, 1976, S. 124.

Verwendung von Glaubwürdigkeitsbeurteilungen die $|\Theta_1|+|\Theta_2|$ Glaubwürdigkeitszuordnungen, wobei hier vorausgesetzt wird, daß die Erkenntnisse vollständig verträglich sind. Ansonsten müssen noch die miteinander unverträglichen Elemente aus $\mathbb{K}_1$ und $\mathbb{K}_2$ bestimmt werden[125]. Die Verwendung eines Wahrscheinlichkeitmaßes setzt demnach präzise Erkenntnisse über die einzelnen Elemente des Möglichkeitsraumes voraus, während diese für die Verwendung des Glaubwürdigkeits- und Plausibilitätsmaßes nicht erforderlich sind[126].

Dieser Unterschied, den neue Erkenntnisse im Hinblick auf ihren Inhalt haben müssen, erklärt auch, weshalb sich die oben angesprochene Gruppe von unscharf formulierten Gesetzmäßigkeiten nicht mittels eines Wahrscheinlichkeitsmaßes, wohl aber durch ein Glaubwürdigkeitsmaß abbilden läßt. Bei der Verwendung eines Wahrscheinlichkeitsmaßes muß, wenn die Erkenntnis darin besteht, daß ein Ereignis aus E eingetreten ist, und dies im Hinblick auf eine Wahrscheinlichkeitsaussage über das Ereignis $\overline{O}$ berücksichtigt werden soll[127], die bedingte Wahrscheinlichkeit für $\overline{O}$ gegeben E, also

(2.59) $$pr(\overline{O}|E)=\frac{pr(\overline{O}\cap E)}{pr(E)}=\frac{pr(\overline{O}\cap E)}{pr(\overline{O}\cap E)+pr(O\cap E)}$$

bestimmt werden können. Wie in (2.59) zu erkennen ist, impliziert dies, daß eine exakte Aufteilung der Prüfungsinformationen, welche durch die Menge E repräsentiert werden, auf die Fälle, in denen sie die Ordnungsmäßigkeit der Prüfungsgesamtheit stützt $(O\cap E)$, und auf solche, in denen sie die Nichtordnungsmäßigkeit stützt $(\overline{O}\cap E)$, erfolgen kann. Dabei impliziert die Festlegung eines der beiden Werte wegen (2.6) auch den anderen Wert. Soweit also demjenigen, der die Wahrscheinlichkeitsbeurteilung durchführt, die Zusammenhänge nicht in der geforderten exakten Form bekannt sind, können diese nicht mittels Wahrscheinlichkeiten abgebildet werden. Die bisher als Beispiel angeführten unscharfen Gesetzmäßigkeiten im Rahmen von analytischen Prüfungshandlungen waren nun genau so beschaffen, daß ausschließlich für $\overline{O}\cap E$ Informationen abgeleitet werden konnten, ohne daß diese die Einschätzung der Ordnungsmäßigkeit verändern sollten. Die angeführten Einschränkungen gelten bei der Verwendung eines Glaubwürdigkeitsmaßes nicht.

---

125 Dieser Fall kann ausschließlich dann auftreten, wenn nicht $\Theta=\Theta_1\times\Theta_2$ als frame of discernment gewählt wird. Dabei ist zu beachten, daß es bei der Verwendung eines Wahrscheinlichkeitsmaßes nicht zulässig ist, $\mathbb{K}\neq\Theta_1\times\Theta_2$ zu wählen.

126 Vgl. auch *Dubois/Prade*, Statistical Data, 1986, S. 349.

127 Siehe oben die Formeln (2.38) und (2.27).

b) Bestimmung von Glaubwürdigkeitswerten

Zur Bestimmung von konkreten Glaubwürdigkeitswerten soll im folgenden untersucht werden, inwieweit die in (2.39) bis (2.41) bzw. die in (2.49) gegebene Definition des Glaubwürdigkeitsmaßes geeignet ist, um zu quantitativen Glaubwürdigkeiten zu gelangen. Da die erste Definition, ebenso wie die Axiomatik von *Kolmogoroff* für Wahrscheinlichkeitsmaße, ausschließlich Anforderungen festlegt, denen eine Glaubwürdigkeitszuordnung gehorchen muß, bietet es sich an zu fragen, welche Anforderungen eine qualitative Glaubwürdigkeitszuordnung erfüllen muß, damit ihre strukturerhaltende quantitative Abbildung die Axiome in (2.39) bis (2.41) erfüllt. Diese Anforderungen sind im folgenden angeführt[128]. Für alle $A, B, C \in \mathbb{K}$ muß gelten[129]:

(2.60) $$bel(A) \leq bel(B) \vee bel(B) \leq bel(A);$$

(2.61) $$bel(A) \leq bel(B) \wedge bel(B) \leq bel(C) \rightarrow bel(A) \leq bel(C);$$

(2.62) $$bel(\varnothing) \leq bel(A);$$

(2.63) $$(A \cup B) \cap C = \varnothing \rightarrow (bel(A) \leq bel(B) \rightarrow bel(A \cup C) \leq bel(B \cup C)).$$

Der einzige Unterschied in den Anforderungen besteht in einer schwächeren Fassung des Unabhängigkeitsaxioms. Während in (2.31) gefordert wird, daß die Beurteilung des Eintritts von Ereignissen sich nicht ändern darf, wenn davon unabhängige Ereignisse in die Beurteilung einbezogen oder aus der Beurteilung ausgeschlossen werden, fordert (2.63) nur, daß sich die Beurteilung bei der Einbeziehung eines unabhängigen Ereignisses nicht ändern darf, die Glaubwürdigkeitsbeurteilung insoweit also monoton ist. Eine Überprüfung, ob Prüfer tatsächlich bei der Beurteilung von Prüfungsgesamtheiten diese Anforderungen erfüllen, kann nur empirisch erfolgen. Hier soll weiterhin im Sinne einer normativen Theorie der Messung von Risiko bei Prüfungen vom Prüfer gefordert werden, daß seine Glaubwürdigkeitseinschätzungen diese Anforderungen erfüllen. Sie erscheinen allerdings wirklichkeitsnäher, da sie weniger streng als die für die Existenz subjektiver Wahrscheinlichkeiten notwendigen Anforderungen sind.

Als nachteilig ist zu beurteilen, daß zumindest bisher für Glaubwürdigkeitszuordnungen kein Axiom bekannt ist, das, vergleichbar dem in (2.33) angege-

[128] Auch wenn in den folgenden Axiomen *bel* eine komparative und keine, wie bisher unterstellt, quantitative Glaubwürdigkeitsbeurteilung darstellt, soll die Bezeichnung beibehalten werden, da aus dem Kontext klar wird, welche Art von Glaubwürdigkeitszuordnung gemeint ist.

[129] Vgl. *Dubois/Prade*, Fuzzy Sets, 1989, S. 145.

benen Axiom, die Existenz einer eindeutigen, strukturerhaltenden quantitativen Glaubwürdigkeitszuordnung zu einer qualitativen Glaubwürdigkeitszuordnung garantiert. Insoweit kann eine vorliegende Glaubwürdigkeitszuordnung ausschließlich daraufhin überprüft werden, ob sie die Axiome in (2.60) bis (2.63) erfüllt. Einzelne Festlegungen von Glaubwürdigkeitsziffern sind demgegenüber rein subjektiv, also intersubjektiv nicht überprüfbar. Dieser Nachteil kommt allerdings im Vergleich mit dem auf subjektiven Wahrscheinlichkeiten basierenden Unschärfemaß nicht zum Tragen, da für die vorliegende Modellwelt, wie oben gezeigt wurde[130], die Bedingung in (2.33) regelmäßig nicht erfüllt sein wird.

Wie bei der Ermittlung subjektiver Wahrscheinlichkeiten, können auch Glaubwürdigkeiten aus tatsächlichen oder hypothetischen Entscheidungen (Wetten, Versicherungen) einer Person abgeleitet werden. Ist demnach eine Person bereit, einen Einsatz e_A für eine Wette, die einen Gewinn von g auszahlt, wenn das Ereignis A eintritt, zu zahlen, so gilt für die Glaubwürdigkeit von A analog zu (2.34):

$$(2.64) \qquad bel(A) \geq \frac{e_A}{g}.$$

Für die Ableitung von Glaubwürdigkeiten wird allerdings nicht unterstellt, daß die betrachtete Person auch die Position der Gegenseite einnehmen würde. Der Einsatz bei einer Wette, bei der der Gewinn ebenfalls g beträgt, und die gewonnen wird, wenn A nicht eintritt, wird mit $e_{\overline{A}}$ bezeichnet. Es gilt dann allenfalls

$$(2.65) \qquad \frac{e_{\overline{A}}}{g} \leq 1 - \frac{e_A}{g},$$

da ansonsten ein sicherer Verlust eingegangen wird. Werden die auf diese Weise ermittelten Wettquotienten als Glaubwürdigkeiten interpretiert[131], gilt in Übereinstimmung mit (2.43)[132]:

$$(2.66) \qquad bel(A \cup \overline{A}) = bel(\Omega) = 1 \geq bel(A) + bel(\overline{A}).$$

[130] Siehe oben in Kapitel B. I. 2. b) auf S. 99.

[131] Diese Interpretation ist selbst dann nicht falsch, wenn man wie *Keynes* den Fall betrachtet, daß für die subjektiven Glaubwürdigkeiten selbst gar keine genauen Werte existieren. Vgl. *Keynes*, Treatise, 1963, S. 22.

[132] Vgl. *Shafer*, Nonadditive Probability, 1985, S. 272.

Die auf diese Weise ermittelten Glaubwürdigkeiten erfüllen zwar die Axiome (2.39) bis (2.41), allerdings stellen sie, gemäß ihrer Definition ausschließlich Untergrenzen für die tatsächliche Glaubwürdigkeit dar. Im Gegensatz zu den bei der Ermittlung subjektiver Wahrscheinlichkeiten unterstellten Annahmen, wird somit nicht sichergestellt, daß maximale Werte für die Höhe der Glaubwürdigkeiten ermittelt werden.

Neben der direkten Angabe von Glaubwürdigkeitsziffern, können nach (2.49) die Glaubwürdigkeiten auch aus der grundlegenden Wahrscheinlichkeitszuordnung auf $\mathbb{K}$ bestimmt werden. Für die Festlegung der grundlegenden Wahrscheinlichkeiten besteht neben der subjektiven Festlegung durch den Prüfer auch die Möglichkeit, diese aus objektiven Wahrscheinlichkeiten abzuleiten[133]. Für eine formale Darstellung der zuletzt genannten Alternative soll der interessierende Ereigniskörper, für den keine Glaubwürdigkeitszuordnungen bekannt sind, mit $\mathbb{K}$, der zugehörige frame of discernment mit Θ bezeichnet werden. Existiert ein zweiter frame of discernment Θ', auf dem ein Wahrscheinlichkeitsmaß *pr* definiert ist, und existiert eine Kompatibilitätsrelation $\Gamma:\Theta' \to \mathbb{K}$, die für alle $x \in \Theta'$ und $A \in \mathbb{K}$ die folgende Bedingung

$$\Gamma(x) = A \tag{2.67}$$

nur dann erfüllt, wenn x mit allen Elementen in A verträglich ist[134], dann ergibt sich die grundlegende Wahrscheinlichkeitszuweisung für Θ nach[135]:

$$\forall A \in \mathbb{K}, A \neq \varnothing : m(A) = \frac{\sum\limits_{\forall x \in \Theta' : \Gamma(x) = A} pr(x)}{1 - \sum\limits_{\forall x \in \Theta' : \Gamma(x) = \varnothing} pr(x)} \text{ und } \sum_{\forall x \in \Theta' : \Gamma(x) = \varnothing} pr(x) < 1 \tag{2.68}$$

$$m(\varnothing) = 0$$

c) Risikomessung

Zum Schluß bleibt die Frage, wie das Risiko einer Entscheidung zu messen ist, falls die Ungewißheit über den Eintritt einzelner Umweltzustände aus S

---

[133] Vgl. *Srivastava/Shafer*, Belief-Function, 1992, S. 257.

[134] Der Vorteil dieser Vorgehensweise liegt darin, daß nicht alle Elemente aus der Menge $\Theta \times \Theta'$ unterschieden werden müssen, sondern ausschließlich für alle Elemente aus Θ' die Verträglichkeit mit den Elementen aus $\mathbb{K}$ angegeben werden muß.

[135] Vgl. *Srivastava/Shafer*, Belief-Function, 1992, S. 257. Da hier auch der Fall zugelassen wird, daß $\exists x \in \Theta' : \Gamma(x) = \varnothing$ gilt, müssen die dort angegebenen grundlegenden Wahrscheinlichkeitszuweisungen normalisiert werden.

durch ein Glaubwürdigkeits- bzw. ein Plausibilitätsmaß gemessen wird. Dazu ist zunächst zu prüfen, welches der beiden Maße zur Messung der Ungewißheit über den Eintritt der Umweltzustände im Rahmen des Entscheidungsmodells zu verwenden ist. Prinzipiell wäre es denkbar die Entscheidung für eines der beiden Maße von der Risikoeinstellung des Entscheidenden abhängig zu machen. Dabei entspricht wegen (2.44) die Verwendung des Plausibilitätsmaßes einer optimistischen, die Verwendung des Glaubwürdigkeitsmaßes einer pessimistischen Schätzung für die Eintrittswahrscheinlichkeit des betrachteten Umweltzustandes. Gegen diese Vorgehensweise spricht jedoch, daß ein hoher Plausibilitätsgrad für einen Umweltzustand nicht durch Erkenntnisse zustande kommt, die für das Eintreten dieses Umweltzustandes sprechen, sondern, wie (2.42) zeigt, ausschließlich aus den fehlenden Kenntnissen über das Eintreten der anderen Umweltzustände resultiert. Hier soll davon ausgegangen werden, daß die Werte für das zu verwendende Ungewißheitsmaß ausschließlich von der Erkenntnis über das Eintreten des jeweiligen Umweltzustandes abhängen sollen. Insoweit gilt also $\forall s \in S: p(s) = bel(s)$.

Für die Risikomessung stehen, falls z. B. die Handlungsalternative a_1 gewählt wird, grundsätzlich zwei Größen, namentlich das Glaubwürdigkeitsmaß $bel(s_2)$ oder das Plausibilitätsmaß $pl(s_2) = 1 - bel(s_1)$, zur Auswahl, da ein Fehlurteil ausschließlich dann abgegeben wird, wenn s_2 eintritt. Da die erste Größe auch dann Null wird, wenn die vorliegenden Erkenntnisse keine Hinweise auf die Glaubwürdigkeit von s_2 geben, ist sie für die Risikomessung ungeeignet. Entsprechend der oben vorgebrachten Argumentation darf aus geringen oder fehlenden Erkenntnissen auch nicht abgeleitet werden, daß ein geringes oder kein Risiko besteht. Die zweite Alternative setzt das Risiko einer Handlungsalternative mit dem Grad der Plausibilität des Umweltzustandes, der zu dem unerwünschten Ergebnis führt, gleich. Dieser entspricht der Obergrenze für die Wahrscheinlichkeit dieses Umweltzustandes beim gegebenen Stand der Erkenntnisse. Falls keine Erkenntnisse vorliegen, die für das Eintreten von s_1 sprechen, gilt $pl(s_2) = 1$. Dies wäre der Fall des maximalen Risikos für die Wahl der Aktion a_1. Der zweiten Alternative ist demnach der Vorzug zu geben. Die Argumentation gilt bei der Wahl von Handlungsalternative a_2 entsprechend, so daß insgesamt das Risiko bei Wahl von a_1 durch $pl(s_2)$ bei der Wahl von a_2 durch $pl(s_1)$ gemessen wird.

III. Allgemeine Unschärfemaße

Auch die allgemeinen Unschärfemaße sollen zunächst axiomatisch definiert werden, um die Unterschiede zum Wahrscheinlichkeits- bzw. dem Glaubwürdigkeits- und Plausibilitätsmaß deutlich zu machen. Bevor anschließend die Interpretation von zwei gegenüber verallgemeinerten Unschärfemaßen vorgestellt wird, sollen der Begriff der unscharfen Menge sowie die für unscharfe Mengen definierten Operatoren dargestellt werden, da diese Grundlage für die Interpretation sind.

1. Axiome

Es sei wiederum unterstellt, daß ein Möglichkeitsraum Ω und ein darüber definierter Ereigniskörper $\mathbb{K}$ existiert. Von einem allgemeinen Unschärfemaß $f:\mathbb{K}\to[0;1]$, das den Grad für das Eintreten eines Ereignisses aus $\mathbb{K}$ aufgrund bestimmter unscharfer Informationen mißt, wird gefordert, daß es den folgenden Axiomen gehorcht[136]:

(2.69) $$f(\emptyset)=0;$$

(2.70) $$f(\Omega)=1;$$

(2.71) $$\forall A,B\in\mathbb{K}: A\subseteq B\to f(A)\leq f(B).$$

Als Folgerungen aus (2.71) ergibt sich insbesondere[137]:

(2.72) $$\forall A,B\in\mathbb{K}: f(A\cup B)\geq max(f(A),f(B)) \text{ und}$$

(2.73) $$\forall A,B\in\mathbb{K}: f(A\cap B)\leq min(f(A),f(B)).$$

Die Axiome sind insgesamt schwächer als die an Wahrscheinlichkeits- bzw. Glaubwürdigkeitsmaße gestellten Anforderungen. Die beiden ersten Axiome (2.69) und (2.70) sollen sicherstellen, daß f auf das Einheitsintervall normiert wird. Darüber hinaus erfordert (2.71) nur, daß die Unschärfemessung insoweit konsistent ist, als f nicht geringer werden darf, wenn das betrachtete Ereignis allgemeiner wird. Wie sich leicht zeigen läßt, ist auch das durch (2.4) bis (2.6) definierte Wahrscheinlichkeitsmaß ein spezielles unscharfes Maß[138]. Wahr-

---

136 Vgl. *Dubois/Prade*, Possibility, 1988, S. 7; *Bandemer/Gottwald*, Einführung, 1993, S. 148.

137 Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 151.

138 Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 148.

scheinlichkeiten sind, wie schon gezeigt wurde[139], dann zur Abbildung von Unschärfe geeignet, wenn stark strukturiertes Wissen vorliegt.

Das durch (2.39) bis (2.41) definierte Glaubwürdigkeitsmaß gehorcht ausschließlich dann der Anforderung in (2.71) und stellt damit einen anderen speziellen Fall eines unscharfen Maßes dar, wenn die der Glaubwürdigkeitszuordnung zugrundeliegenden Erkenntnisse konsonant sind[140]. Eine Glaubwürdigkeitszuordnung ist dann konsonant, wenn ein positiver Glaubwürdigkeitsgrad entweder einem Ereignis oder seinem Komplement zugeordnet wird, aber nicht beiden. Ein Glaubwürdigkeitsmaß ist demnach dann ein allgemeines Unschärfemaß, wenn gilt[141]:

(2.74) $$\forall A \in \mathbb{K} : min\left(bel(A); bel(\overline{A})\right) = 0 .$$

2. *Unscharfe Mengen*

a) Definition

In (1.3) wurde X als Teilmenge von U durch eine Zugehörigkeitsfunktion π_X definiert, die für jedes Element x aus U angibt, ob es zu X gehört, oder nicht. Eine Menge kann demnach außer durch die Aufzählung ihrer Elemente auch durch die Angabe eines Grundbereichs, für die Prüfungsgesamtheit X ist dies U, und die Festlegung einer geeigneten Zugehörigkeitsfunktion definiert werden. Soweit die Zugehörigkeitsfunktion für alle Elemente des Grundbereichs ausschließlich die Werte Null für die Nichtzugehörigkeit und Eins für die Zugehörigkeit annehmen kann, wird eine scharf abgegrenzte oder gewöhnliche Menge definiert. Dieses Konzept kann jedoch dadurch verallgemeinert werden, daß auch Zugehörigkeitswerte zwischen Null und Eins zugelassen werden. Im folgenden soll als Grundbereich wie bei den bisherigen Unschärfemaßen die gewöhnliche Menge Ω gewählt werden. Ist x ein Element des Grundbereiches so ist eine unscharfe Menge (fuzzy set) $\tilde{A}$ durch ihre Zugehörigkeitsfunktion:

(2.75) $$\pi_{\tilde{A}} : \Omega \to [0;1]$$

definiert[142]. Für $x \in \Omega$ gibt der Ausdruck $\pi_{\tilde{A}}(x) = r$ an, daß x der Menge $\tilde{A}$ mit dem Grad r angehört. Eine unscharfe Menge $\tilde{A}$ heißt normalisiert, wenn gilt[143]:

[139] Siehe oben Kapitel B. II. 2. a) auf S. 113.

[140] Vgl. *Dubois/Prade*, Statistical Data, 1986, S. 349.

[141] Vgl. *Shafer*, Theory, 1976, S. 220f.

[142] Vgl. *Zadeh*, Fuzzy Sets, 1965, S. 339.

[143] Vgl. *Zimmermann*, Fuzzy Set Theory, 1991, S. 13.

(2.76) $$\exists x \in \Omega : \pi_{\tilde{A}}(x) = 1.$$

Für die Zugehörigkeitsfunktionen von Ω und $\varnothing$ soll gelten[144]:

(2.77) $$\forall x \in \Omega : \pi_{\Omega}(x) = 1,$$

(2.78) $$\forall x \in \Omega : \pi_{\varnothing}(x) = 0.$$

Probleme der geeigneten Definition bzw. effizienten Speicherung von Zugehörigkeitsfunktionen sollen hier nicht betrachtet werden, da es um die generelle Eignung der Theorie unscharfer Mengen zur Abbildung der aufgezeigten Formen von Unschärfe im vorliegenden Prüfungsmodell geht. Soweit bisher bekannt ist, hat allerdings die konkrete Ausgestaltung der Zugehörigkeitsfunktionen, soweit sie die qualitative Ordnung der betrachteten Objekte nicht verändern, geringen Einfluß auf die erzielten Ergebnisse[145].

Als Kenngrößen unscharfer Mengen werden im weiteren insbesondere die Kardinalität einer endlichen unscharfen Menge $\tilde{A}$ benötigt. Diese ist definiert durch[146]:

(2.79) $$\left|\tilde{A}\right| = \sum_{\forall x \in \Omega} \pi_{\tilde{A}}(x).$$

Darüber hinaus soll noch als spezielle Menge A^{α} über Ω, die sich aus $\tilde{A}$ ableiten läßt, die Menge der Elemente definiert werden, deren Zugehörigkeitsgrad zu $\tilde{A}$ mindestens α beträgt:

(2.80) $$A^{\alpha} = \left\{x \in \Omega \middle| \pi_{\tilde{A}}(x) \geq \alpha\right\}.$$

Die Menge A^{α} wird als scharfer α-Schnitt von $\tilde{A}$ bezeichnet und ist eine gewöhnliche Menge[147]. Für zwei unscharfe Mengen $\tilde{A}$ und $\tilde{B}$ und ihre α-Schnitte gelten die folgenden Relationen[148]:

(2.81) $$\tilde{A} = \tilde{B} \leftrightarrow \left(\forall \alpha \in (0;1] : A^{\alpha} = B^{\alpha}\right) \leftrightarrow \left(\forall x \in \Omega : \pi_{\tilde{A}}(x) = \pi_{\tilde{B}}(x)\right);$$

(2.82) $$\tilde{A} \subseteq \tilde{B} \leftrightarrow \left(\forall \alpha \in (0;1] : A^{\alpha} \subseteq B^{\alpha}\right) \leftrightarrow \left(\forall x \in \Omega : \pi_{\tilde{A}}(x) \leq \pi_{\tilde{B}}(x)\right).$$

---

144 Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 19f.

145 Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 38.

146 Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 22. Die Kardinalität unendlicher unscharfer Mengen wird im vorliegenden Modell nicht benötigt.

147 Vgl. *Dubois/Prade*, Possibility, 1988, S. 14.

148 Vgl. *Zadeh*, Fuzzy Sets, 1965, S. 340; *Bandemer/Gottwald*, Einführung, 1993, S. 24f.

b) Mengenoperationen

Für unscharfe Mengen sind analog den üblichen Mengen die Operationen der Komplementbildung bezogen auf Ω, der Vereinigung und der Durchschnittsbildung definiert. Gilt wiederum $\tilde{A},\tilde{B}\subseteq\Omega$, so wird eine Verallgemeinerung der für gewöhnliche Mengen geltenden Operationen durch die folgenden Vereinbarungen erreicht[149]:

Komplement:

(2.83) $$\forall x\in\Omega:\pi_{\overline{\tilde{A}}}(x)=1-\pi_{\tilde{A}}(x);$$

Vereinigung:

(2.84) $$\forall x\in\Omega:\pi_{\tilde{A}\cup\tilde{B}}(x)=max(\pi_{\tilde{A}}(x);\pi_{\tilde{B}}(x));$$

Durchschnitt:

(2.85) $$\forall x\in\Omega:\pi_{\tilde{A}\cap\tilde{B}}(x)=min(\pi_{\tilde{A}}(x);\pi_{\tilde{B}}(x)).$$

Aus den Definitionen in (2.83) und (2.85) folgt insbesondere, daß sich ein unscharfes Ereignis repräsentiert durch die unscharfe Menge $\tilde{A}$ und seine Verneinung, also das Komplement bezogen auf Ω regelmäßig nicht gänzlich ausschließen werden, und der Grundbereich nicht durch die Vereinigung einer unscharfen Menge und ihrem Komplement entsteht. Es gilt demnach[150]:

(2.86) $$\exists x:0<\pi_{\tilde{A}}(x)<1\rightarrow\tilde{A}\cap\overline{\tilde{A}}\neq\varnothing,$$

(2.87) $$\exists x:0<\pi_{\tilde{A}}(x)<1\rightarrow\tilde{A}\cup\overline{\tilde{A}}\neq\Omega.$$

Während für die Bildung des Komplements regelmäßig die in (2.83) angegebene Funktion Verwendung findet, können für die Vereinigung und den Durchschnitt von Mengen neben dem *max*- und dem *min*-Operator auch andere Operatoren verwendet werden. Diese sollen hier, bis auf eine grobe Skizze ihrer Eigenschaften, nicht weiter behandelt werden[151]. Eine erste Unterscheidung der Operatoren kann danach getroffen werden, ob einer der Zugehörigkeitswerte der beteiligten unscharfen Mengen erhalten bleibt oder nicht. Bei den ersteren handelt es sich um nicht interaktive Operatoren, die zuletzt genannten werden als interaktive Operatoren bezeichnet. Die in (2.84) und (2.85) verwen-

---

[149] Vgl. *Zadeh*, Fuzzy Sets, 1965, S. 340f.; *Bandemer/Gottwald*, Einführung, 1993, S. 39.

[150] Vgl. *Demant*, Fuzzy-Theorie, 1993, S. 17f.

[151] Eine Zusammenfassung der im einzelnen diskutierten Operatoren findet sich in *Zimmermann*, Fuzzy Set Theory, 1991, S. 29-39.

deten Operatoren sind die einzigen nicht interaktiven Operatoren[152]. Bei den interaktiven Operatoren sind noch zwei wichtige Klassen zu unterscheiden. Zum einen existiert eine Reihe von t-Normen für die Definition der Vereinigung von unscharfen Mengen und zugehörige t-Conormen zur Durchschnittsbildung mit bestimmten gewünschten mathematischen Eigenschaften, wie Monotonie, Kommutativität und Assoziativität[153]. Für alle Operatoren aus dieser Klasse gilt, daß die resultierenden Zugehörigkeitswerte bei der Vereinigung mindestens so hoch wie der Wert des *max*-Operators nach (2.84) und bei der Durchschnittsbildung höchstens so hoch wie der Wert des *min*-Operators nach (2.85) sind[154].

Da sich dieses Verhalten der Mengenoperatoren insbesondere beim Einsatz in Systemen zur Entscheidungsunterstützung als unbefriedigend erwiesen hat[155], wurden Operatoren für die Vereinigung und Durchschnittsbildung untersucht, die Zugehörigkeitswerte zwischen den Zugehörigkeitswerten der aggregierten unscharfen Mengen liefern[156]. Sie werden als Durchschnittsoperatoren bezeichnet und zeigen kompensatorisches Verhalten in dem Sinn, daß bei der Durchschnittsbildung nicht ausschließlich der schlechteste und bei der Vereinigung nicht ausschließlich der beste Zugehörigkeitsgrad berücksichtigt wird. Als letztes Unterscheidungsmerkmal soll schließlich die Möglichkeit, das Verhalten der Operatoren durch Parameter zu steuern, angeführt werden. Beispiel für einen kompensatorischen Operator, dessen Verhalten über einen Parameter zwischen Vereinigungsbildung und Durchschnittsbildung „stufenlos" eingestellt werden kann, ist das „kompensatorische-und". Wird diese Art der Verknüpfung zweier unscharfer Mengen $\tilde{A},\tilde{B} \subseteq \Omega$ mit $\tilde{A} \circ \tilde{B}$ bezeichnet, dann ist das kompensatorische-und für $0 \leq \gamma \leq 1$ definiert durch[157]:

$$(2.88) \qquad \forall x \in \Omega: \pi_{\tilde{A}\circ\tilde{B}}(x) = \left(\pi_{\tilde{A}}(x)\pi_{\tilde{B}}(x)\right)^{1-\gamma}\left(1-\left(1-\pi_{\tilde{A}}(x)\right)\left(1-\pi_{\tilde{B}}(x)\right)\right)^{\gamma}.$$

Anwendungsprobleme solcher parametrisierter Operatoren ergeben sich allerdings durch die fehlende Interpretation der Parameter im Einzelfall, so daß keine direkte empirische Ermittlung in Betracht kommt[158]. Die Anwendung ist demnach auf solche Fälle beschränkt, in denen Beobachtungen sowohl über

[152] Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 44.
[153] Vgl. *Kruse* et al., Fuzzy-Systeme, 1995, S. 23f.
[154] Vgl. *Zimmermann*, Fuzzy Set Theory, 1991, S. 38.
[155] Vgl. die Übersicht bei *Rommelfanger*, Decision, 1994, S. 26ff.
[156] Vgl. *Zimmermann*, Fuzzy Set Theory, 1991, S. 36.
[157] Vgl. *Zimmermann/Zysno*, Connectives, 1980, S. 47.
[158] Vgl. *Zimmermann/Zysno*, Decisions, 1983, S. 256f.

einzelne Zugehörigkeitsgrade für die zu aggregierenden unscharfen Mengen als auch für die Zugehörigkeitsfunktion der aggregierten unscharfen Menge vorliegen. Der relevante Wert für den Parameter kann dann geeignet geschätzt werden[159]. Allerdings können dann Probleme bei der Anwendung des auf diese Weise ermittelten Parameters auftreten, wenn nicht in den Beobachtungsdaten vorhandene Zugehörigkeitswerte zu aggregieren sind, oder mit der Veränderung des Parameters im Zeitablauf zu rechnen ist.

3. Interpretationen

Die Interpretation eines allgemeinen Unschärfemaßes kann, falls Erkenntnisse durch unscharfe Mengen und deren Verknüpfung repräsentiert werden, von zwei unterschiedlichen Standpunkten aus erfolgen. Zum einen kann für ein bestimmtes Element aus dem Grundbereich nach dem Grad der Möglichkeit gefragt werden, daß es zu einer bestimmten gewöhnlichen Menge $A \in \mathbb{P}(\Omega)$ gehört, wenn ausschließlich Erkenntnisse in Form unscharfer Ereignisse (Mengen) aus $\mathbb{P}(\Omega)$ bekannt sind. Diese Fragestellung führt zu einem Möglichkeits- und Notwendigkeitsmaß über $\mathbb{P}(\Omega)$ oder über einem geeignet gewählten Ereigniskörper $\mathbb{K}$. Zum anderen kann, wie bei der Definition unscharfer Menge in (2.75), die Frage nach dem Zugehörigkeitsgrad eines Elementes aus dem Grundbereich zu einer bestimmten unscharfen Menge gestellt werden[160]. Dieser Standpunkt soll anschließend bei der Diskussion des Konzepts der linguistischen Variable eingenommen werden.

a) Möglichkeits- und Notwendigkeitsmaß

Bezeichnet Ω weiterhin den Möglichkeitsraum und $\mathbb{K}$ einen darüber definierten Ereigniskörper, so sollen für ein Maß $pos: \mathbb{K} \to [0;1]$[161], das den Grad der Möglichkeit für den Eintritt eines Ereignisses angibt, die folgenden Vereinbarungen gelten[162]:

(2.89) $$pos(\varnothing) = 0;$$

(2.90) $$pos(\Omega) = 1;$$

---

[159] Vgl. *Zimmermann/Zysno*, Connectives, 1980, S. 48.

[160] Vgl. *Dubois/Prade*, a, 1988, S. 14f.

[161] *pos* steht für „possibility".

[162] Vgl. *Zadeh*, Possibility, 1978, S. 9; *Dubois/Prade*, Possibility, 1988, S. 8; *Bandemer/Gottwald*, Einführung, 1993, S. 151.

DOI https://doi.org/10.3790/978-3-428-48909-1

(2.91) $$\forall A,B \in \mathbb{K}: pos(A \cup B) = max(pos(A); pos(B)).$$

Wie zu erkennen ist, handelt es sich bei dem in (2.89) bis (2.91) definierten Möglichkeitsmaß um ein allgemeines Unschärfemaß im Sinne von (2.69) bis (2.71). Analog der Vorgehensweise, die bei dem oben definierten Glaubwürdigkeitsmaß gewählt wurde, existiert zu jedem Möglichkeitsmaß ein Notwendigkeitsmaß $nec: \mathbb{K} \to [0;1]$[163], für das gilt[164]:

(2.92) $$\forall A \in \mathbb{K}: nec(A) = 1 - pos(\overline{A});$$

(2.93) $$\forall A,B \in \mathbb{K}: nec(A \cap B) = min(nec(A); nec(B)).$$

Die beiden Maße *pos* und *nec* stellen demnach genau die Grenzfälle der Anforderungen, die an allgemeine Unschärfemaße in (2.72) und (2.73) gestellt wurden, dar. Für $A \in \mathbb{K}$ spiegelt demnach bei gegebenem Erkenntnisstand die Zuweisung $pos(A) = \mathrm{r}$ wider, daß der Grad, in dem der Eintritt von A für möglich gehalten wird, beim gegebenen Erkenntnisstand r beträgt. Entsprechend gibt $nec(A) = \mathrm{r}$ an, daß die Notwendigkeit für das Eintreten von A beim gegebenen Erkenntnisstand dem Grad r entspricht. Es gilt insbesondere[165]:

(2.94) $$\forall A \in \mathbb{K}: pos(A) \geq nec(A).$$

Dies ist insoweit intuitiv einleuchtend, da ein Ereignis eher für möglich gehalten werden kann, als sein Eintreten für notwendig erachtet wird.

Bevor die Ansätze zur Bestimmung von Möglichkeits- und Notwendigkeitswerten diskutiert werden, soll wiederum untersucht werden, welche Anforderungen an eine komparative Möglichkeitszuordnung *pos* zu stellen sind, damit ihre strukturerhaltende quantitative Abbildung die Vereinbarungen in (2.89) bis (2.91) erfüllt. Dies ist genau dann der Fall, wenn für alle $A,B,C \in \mathbb{K}$ gilt[166]:

(2.95) $$pos(A) \leq pos(B) \vee pos(B) \leq pos(A);$$

(2.96) $$pos(A) \leq pos(B) \wedge pos(B) \leq pos(C) \to pos(A) \leq pos(C);$$

(2.97) $$pos(\varnothing) \leq pos(A);$$

---

[163] *nec* steht für „necessity".

[164] Vgl. *Dubois/Prade*, Possibility, 1988, S. 10; *Klir/Folger*, Fuzzy Sets, 1988, S. 122; *Bandemer/Gottwald*, Einführung, 1993, S. 153.

[165] Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 154; *Kruse* et al., Fuzzy-Systeme, 1995, S. 87f.

[166] Vgl. *Dubois/Prade*, Fuzzy Sets, 1989, S. 145.

DOI https://doi.org/10.3790/978-3-428-48909-1

(2.98) $$pos(A) \leq pos(B) \rightarrow pos(A \cup C) \leq pos(B \cup C).$$

Gegenüber den Anforderungen in (2.60) bis (2.63), die an eine qualitative Glaubwürdigkeitszuordnung gestellt werden, besteht der einzige Unterschied in der Formulierung des Axioms (2.98). Während in (2.63) für Glaubwürdigkeits- und in (2.31) für Wahrscheinlichkeitszuordnungen gefordert wird, daß diese monoton sind, wenn ein unabhängiges Ereignis in die Betrachtung einbezogen wird, wird die Monotoniebedingung in (2.98) für Möglichkeitsmaße auf jede gleichförmige Erweiterungen des betrachteten Bereichs ausgedehnt. Es wird insoweit von demjenigen, der die Möglichkeitszuordnung erstellt, nicht gefordert, daß er disjunkte Ereignisse aus $\mathbb{K}$ identifizieren können muß, wie dies für Glaubwürdigkeits- und Wahrscheinlichkeitszuordnungen notwendig ist. Insoweit sind also die Anforderungen, die an das Wissen über die Struktur des betrachteten Ereigniskörpers gestellt werden geringer als bei der Verwendung eines Wahrscheinlichkeits- oder Glaubwürdigkeitsmaßes.

Die Ableitung einer quantitativen Möglichkeitszuordnung kann dann erfolgen, wenn die vorhandenen unscharfen Erkenntnisse in Form einer grundlegenden Möglichkeitszuordnung $pos_{\tilde{E}} : \Omega \rightarrow [0;1]$[167] vorliegen. Diese soll normalisiert sein, d. h., es muß gelten:

(2.99) $$\exists x \in \Omega : pos_{\tilde{E}}(\{x\}) = 1.$$

Aus (2.91) lassen sich dann die zugehörigen Möglichkeits- und Notwendigkeitswerte für alle Ereignisse aus $\mathbb{K}$ ableiten mit[168]:

(2.100) $$\forall A \in \mathbb{K} : pos(A) = \max_{\forall x \in A} \left(pos_{\tilde{E}}(\{x\}) \right) \text{ und}$$

(2.101) $$\forall A \in \mathbb{K} : nec(A) = \min_{\forall x \notin A} \left(1 - pos_{\tilde{E}}(\{x\}) \right).$$

Es stellt sich nun ebenso, wie bei dem oben angeführten Glaubwürdigkeits- und Plausibilitätsmaß, die Frage, wie die Risikomessung bei der Wahl einer Handlungsalternative zu erfolgen hat, wenn die Ungewißheit über den Eintritt einzelner Umweltzustände durch ein Möglichkeits- und Notwendigkeitsmaß abgebildet wird. Damit verbunden ist zunächst die Beantwortung der Frage, ob die Ungewißheit über den Eintritt einzelner Umweltzustände aus S im hier unterstellen Entscheidungsmodell durch ein Möglichkeits- oder ein Notwendigkeitsmaß zu messen ist. Dabei erscheint es für die Messung der Ungewißheit

---

[167] Es soll hier keine Unterscheidung zwischen der im Schrifttum diskutierten „possibility distribution" und einer Möglichkeitszuordnung auf Einermengen aus $\mathbb{K}$ gemacht werden.

[168] Vgl. *Klir/Folger*, Fuzzy Sets, 1988, S. 122f.; *Kruse* et al., Fuzzy-Systeme, 1995, S. 86.

aus den gleichen Gründen, die oben für das Glaubwürdigkeitsmaß angeführt wurden[169], plausibler, auf den Grad der Notwendigkeit, mit dem ein Ereignis eintreten muß, als auf den Grad der Möglichkeit, mit dem ein Ereignis eintreten kann, abzustellen. Auch hier wird zur Messung der Ungewißheit das pessimistischere Maß verwendet. Es soll daher unterstellt werden, daß für den hier betrachteten Fall $\forall s \in S: p(s) = nec(s)$ gilt. Aus (2.90), (2.91) und (2.92) folgt insbesondere[170]:

(2.102) $$\forall A \in \mathbb{K}: min(nec(A); nec(\overline{A})) = 0,$$

(2.103) $$\forall A \in \mathbb{K}: pos(A) < 1 \leftrightarrow pos(\overline{A}) = 1 \leftrightarrow nec(A) = 0 \text{ und}$$

(2.104) $$\forall A \in \mathbb{K}: nec(A) > 0 \leftrightarrow nec(\overline{A}) = 0 \leftrightarrow pos(A) = 1.$$

Bei der Wahl von a_1 stehen, da ein Fehlurteil ausschließlich dann abgegeben wird, wenn s_2 eintritt, wiederum prinzipiell $nec(s_2)$ oder $pos(s_2) = 1 - nec(s_1)$ für die Risikomessung zur Auswahl. Wegen (2.104) wird $nec(s_2)$ zwingend dann Null, wenn der Grad der Notwendigkeit für den Eintritt von s_1 von Null abweicht. Da das Risiko der Entscheidung für a_1 somit schon dann Null wäre, wenn die geringste Erkenntnis für die Notwendigkeit der Ordnungsmäßigkeit der Prüfungsgesamtheit vorliegt, erscheint $nec(s_2)$ nicht zur Risikomessung der Aktion a_1 geeignet. Wird dagegen das Risiko durch den Grad der Möglichkeit, mit dem s_2 eintreten kann, gemessen, so führt dieses Vorgehen für den Fall, daß a_1 gewählt wird, obwohl die Notwendigkeit von s_1 Null ist, gemäß (2.103) zum maximalen Risiko von Eins. In den übrigen Fällen spiegelt das gemessene Risiko wider, wieviel bei den gegebenen Erkenntnissen zur Notwendigkeit des Eintretens von s_1 fehlt. Soweit die Unsicherheit bezüglich des relevanten Umweltzustandes durch ein Notwendigkeits- und Möglichkeitsmaß gemessen wird, entspricht demnach das Risiko einer gewählten Handlungsalternative dem Grad der Möglichkeit für das Eintreten des Umweltzustandes, der zu einem Fehlurteil führt.

Offen ist demnach die Frage, wie die oben angesprochene grundlegende Möglichkeitszuordnung aus den vorhandenen unscharfen Erkenntnissen abzuleiten ist. Wird zunächst der Grenzfall betrachtet, daß die Erkenntnisse in der Aussage bestehen, daß $E \in \mathbb{K}$ eingetreten ist, dann folgt für die Möglichkeitzuordnung[171]:

---

[169] Siehe oben Kapitel B. II. 2. c) in diesem Teil der Arbeit.

[170] Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 154.

[171] Vgl. *Dubois/Prade*, Fuzzy Sets, 1989, S. 137.

(2.105) $$A \in \mathbb{K}: pos_E(A) = \begin{cases} 1, \text{wenn } E \cap A \neq \emptyset \\ 0, \text{sonst} \end{cases}.$$

Möglichkeitszuordnungen zwischen Null und Eins können demnach ausschließlich dann auftreten, wenn die Zugehörigkeit einzelner Elemente zu dem bekannten Ereignis nicht eindeutig ist. Dies ist gleichbedeutend mit dem Umstand, daß die Erkenntnisse in Form einer nicht scharfen Beschränkung $\tilde{E}$ mit $\pi_{\tilde{E}}(x)$ für alle $x \in \Omega$ des Möglichkeitsraumes Ω vorliegen. Betrachtet man z. B. den Fall, daß der Möglichkeitsraum den möglichen Fehleranteilen in einer Prüfungsgesamtheit, also den Zahlen zwischen Null und Eins entspricht, und die unscharfe Erkenntnis $\tilde{E}$, daß der Fehleranteil gering sein muß, vorliegt. Existiert dann eine Zugehörigkeitsfunktion $\pi_{\tilde{E}}(x)$ mit $x \in \Omega$ für die unscharfe Menge der geringen Fehleranteile, so kann als grundlegende Möglichkeitszuordnung

(2.106) $$\forall x \in \Omega: pos_{\tilde{E}}(\{x\}) = \pi_{\tilde{E}}(x)$$

gewählt werden[172]. Aus dieser grundlegenden Möglichkeitszuordnung kann dann gemäß (2.100) z. B. der Möglichkeitsgrad für einen Fehleranteil zwischen 0,01 und 0,03 bestimmt werden. Die Betrachtung von Erkenntnissen, die in Form einer Angabe wie „Der Fehleranteil ist gering" vorliegen, führt direkt zum Konzept der linguistischen Variable. Dieses soll daher im nächsten Abschnitt behandelt werden.

b) Linguistische Variable

Unter einer linguistischen Variable wird eine Variable, deren Ausprägungen nicht numerische Ausdrücke sind, verstanden. Als Beispiel soll die linguistische Variable „FEHLERANTEIL" mit der Menge ihrer möglichen Ausprägungen {„sehr hoch"; „hoch"; „mittel"; „gering"; „sehr gering"} betrachtet werden. Eine Abbildung durch unscharfe Mengen kommt dann in Betracht, wenn die Werte der Variablen grundsätzlich auch über eine metrische Skala (Grundbereich) definiert werden könnten. Im obigen Beispiel wäre dies das Intervall der reellen Zahlen zwischen Null und Eins. Jede mögliche Ausprägung der linguistischen Variable kann dann durch eine über dem Grundbereich definierte Zugehörigkeitsfunktion angegeben werden. Abb. 2.2 zeigt eine Alternative, wie die Zugehörigkeitsfunktionen für die einzelnen im Beispiel genannten Ausprägungen definiert werden könnten.

---

172 Vgl. *Zadeh*, Possibility, 1978, S. 11f.

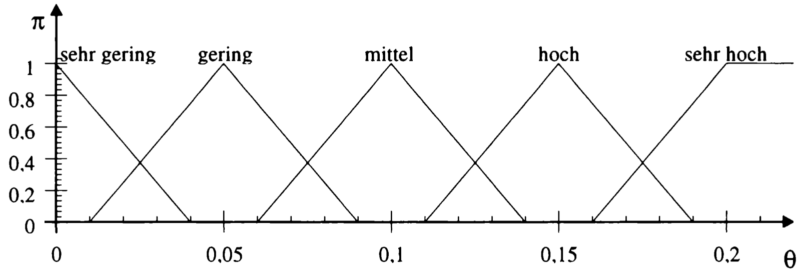

Abb. 2.2: Zugehörigkeitsfunktionen für die möglichen Ausprägungen der linguistischen Variable FEHLERANTEIL

Eine linguistische Variable ist demnach

- durch ihren Namen,
- durch den Grundbereich Ω, aus dem sie grundsätzlich Werte annehmen könnte,
- durch die Menge von Ausprägungen $\mathbb{A}$ in Form natürlichsprachlicher Ausdrücke, welche die Variable annehmen kann, und
- durch die Zugehörigkeitsfunktionen $\pi_{\tilde{A}}$ auf dem Grundbereich Ω für die einzelnen Ausprägungen $\tilde{A} \in \mathbb{A}$

eindeutig definiert. Eine Erweiterung könnte dieses Konzept durch die Einführung von syntaktischen Regeln erfahren, die eine Modifikation einzelner Ausdrücke durch Zusätze, wie „sehr" oder „mehr oder weniger" zulassen[173]. Im angeführten Beispiel würde dann als Menge von Ausprägungen { „hoch"; „mittel"; „gering" } genügen, die beiden fehlenden Ausprägungen könnten dann über die anzugebenden syntaktischen Regeln abgeleitet werden. Für die Modifikation einer Ausprägung durch „sehr" bzw. „mehr oder weniger" werden insbesondere vorgeschlagen[174]:

(2.107) $$\pi_{\text{sehr hoch}}(x) = \pi_{\text{hoch}}(x)^2 \text{ und}$$

(2.108) $$\pi_{\text{mehr oder weniger hoch}}(x) = \sqrt{\pi_{\text{hoch}}(x)}.$$

Die Auswirkungen auf den Verlauf der Zugehörigkeitsfunktion sind in Abb. 2.3 dargestellt. Insgesamt sind diese Formen der Modifikation von linguistischen Variablen jedoch theoretisch wenig untersucht[175]. Über die Zulässigkeit ihrer Anwendung kann daher ausschließlich für jeden einzelnen Anwendungsfall gesondert entschieden werden. Unproblematisch ist die logische Verknüp-

---

[173] Vgl. *Rommelfanger*, Decision, 1994, S. 66; *Zimmermann*, Fuzzy Set Theory, 1991, S. 132.

[174] Vgl. *Zadeh*, Concept, Part 2, 1975, S. 292 und 294.

[175] Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 104.

fung von linguistischen Wertzuweisungen, wie „Der Fehleranteil ist gering oder sehr gering“ durch die für unscharfe Mengen definierten Mengenoperatoren, die dann auf die relevanten Zugehörigkeitsfunktionen der zu verknüpfenden Ausprägungen der linguistischen Variable anzuwenden wären[176].

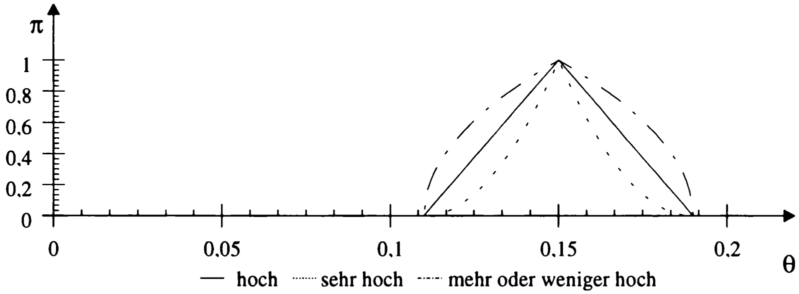

Abb. 2.3: Auswirkungen auf die Zugehörigkeitsfunktion bei Modifikation der Ausprägung „hoch“ durch „sehr“ und „mehr oder weniger“

Auch wenn die Einschränkung des Möglichkeitsraumes Ω durch unscharf formulierte Ausprägungen linguistischer Variablen als intuitiv einleuchtendes Konzept erscheint, bleibt die Frage offen, wie für den konkreten Anwendungsfall die jeweiligen Zugehörigkeitsfunktionen für die verwendeten linguistischen Variablen zu ermitteln sind[177]. Zur Lösung dieser Frage werden insbesondere zwei Ansätze diskutiert. Zunächst kann die Ableitung der Zugehörigkeitsfunktionen $\pi_{\tilde{A}}$ für die Ausprägung $\tilde{A}$ einer linguistischen Variablen individuell aus den Einschätzungen der Person abgeleitet werden, welche die Variable verwendet. Damit eine eindeutige Abbildung $\pi_{\tilde{A}}:\Omega \rightarrow [0;1]$ existiert, muß zum einen der Grad der Zugehörigkeit zur Ausprägung $\tilde{A}$, den die Person einem Wert x aus Ω zuordnet, auf einer Intervallskala gemessen werden können[178]. Dies erfordert insbesondere, daß die Differenzen der Zugehörigkeitsgrade von jeweils zwei Elementen aus Ω schwach transitiv geordnet werden können. Zum anderen muß zumindest ein Element aus Ω existieren, das die Ausprägung $\tilde{A}$ eindeutig erfüllt, und zumindest ein Element aus Ω darf die Ausprägung $\tilde{A}$ eindeutig nicht erfüllen. Der Grad der Zugehörigkeit muß demnach sowohl nach oben als auch nach unten begrenzt sein[179]. Auch wenn intervallskalierte Zugehörigkeitsgrade für eine Ableitung der Zugehörigkeitsfunktion ausreichen, ist

176 Siehe oben Kapitel B. III. 2. b) in diesem Teil der Arbeit.

177 Vgl. *Dubois/Prade*, Fuzzy Sets, 1980, S. 256f.

178 Zu den Axiomen, die insgesamt erfüllt sein müssen, vgl. *Krantz*, et al., Measurement, 1971, S. 151.

179 Vgl. *Norwich/Turksen*, Model, 1984, S. 3.

das resultierende, auf das Einheitsintervall normierte Zugehörigkeitsmaß $\pi_{\tilde{A}}(x)$ verhältnisskaliert, da es die Relation zwischen der Differenz aus dem Grad der Zugehörigkeit von x und der Untergrenze für den Zugehörigkeitsgrad und der Differenz zwischen Ober- und Untergrenze des Zugehörigkeitsgrades ausdrückt[180].

Über diese grundlegenden meßtheoretischen Untersuchungen zur Interpretation von Zugehörigkeitsgraden hinausgehende Axiomatisierungen existieren bisher nicht[181]. Die Mehrzahl der im Schrifttum diskutierten Ansätze zur Bestimmung von Zugehörigkeitsfunktionen ist eher pragmatisch orientiert. Dabei wird ausgehend von groben Vorstellungen über den Verlauf der Zugehörigkeitsfunktion für einen bestimmten Kontext diese durch eine geeignet zu parametrisierende Funktion approximiert[182].

Bei den bisherigen Ansätzen zur Festlegung der Zugehörigkeitsfunktionen wurde unterstellt, daß Unschärfe im Hinblick auf die Zugehörigkeit einzelner Werte aus dem Grundbereich zu den einzelnen Ausprägungen einer linguistischen Variable bei einer einzelnen Person besteht. Eine zweite Möglichkeit besteht darin, daß zwar einzelne Personen eindeutige Zuordnungen von Elementen des Grundbereiches zu einer Ausprägung der betrachteten linguistischen Variable treffen können, diese aber nicht für alle Personen einheitlich sind. Unschärfe entsteht in diesem Fall daraus, daß die Zuordnung von Elementen des Grundbereichs abhängig ist vom Kontext der jeweiligen Person. Die Menge der Kontexte, die für die abzuleitende Zugehörigkeitsfunktion der Ausprägung $\tilde{A}$ einer betrachteten linguistischen Variable zu berücksichtigen sind, soll mit $\mathbb{C}_{\tilde{A}}$ bezeichnet werden. Existiert ein Wahrscheinlichkeitsmaß $pr(c)$ für $c \in \mathbb{C}_{\tilde{A}}$ über der Potenzmenge von $\mathbb{C}_{\tilde{A}}$ und eine Abbildung $\Gamma_{\tilde{A}}: \mathbb{C}_{\tilde{A}} \to \mathbb{P}(\Omega)$, dann läßt sich die Zugehörigkeitsfunktion $\pi_{\tilde{A}}$ ableiten nach[183]:

$$\forall x \in \Omega : \pi_{\tilde{A}}(x) = \sum_{x \in \Gamma_{\tilde{A}}(c)} pr(c). \tag{2.109}$$

Dabei wird das Wahrscheinlichkeitsmaß $pr(c)$ ausschließlich zur Gewichtung der einzelnen Kontexte verwendet, ohne daß damit eine subjektive oder

180 Vgl. *Norwich/Turksen*, Model, 1984, S. 21f.

181 Vgl. *Zimmermann*, Fuzzy Set Theory, 1991, S. 344; *Kruse* et al., Fuzzy-Systeme, 1995, S. 43.

182 Vgl. *Dubois/Prade*, Possibility, 1988, S. 19f.; *Dubois/Prade*, Fuzzy Sets, 1989, S. 149. Ergebnisse empirischer Untersuchungen zu dieser Art der Festlegung von Zugehörigkeitsfunktionen finden sich bei *Zimmermann*, Fuzzy Set Theory, 1991, S. 346-355.

183 Vgl. *Kruse* et al., Fuzzy-Systeme, 1995, S. 45.

„sehr hoher" Fehleranteil $c \in \mathbb{C}_{\text{sehr hoch}}$	relative Häufigkeit $pr(c)$
$\geq 15\%$	0,1
$\geq 17{,}5\%$	0,25
$\geq 20\%$	0,5
$\geq 25\%$	0,15

Abb. 2.4: Beispieldaten zur Ableitung einer Zugehörigkeitsfunktion für die Ausprägung „sehr hoch" der linguistischen Variable FEHLERANTEIL

objektive Interpretation des Wahrscheinlichkeitsmaßes verbunden wird[184]. Die resultierende Zugehörigkeitsfunktion erfüllt allerdings ausschließlich dann die in (2.99) geforderte Eigenschaft der Normalisiertheit, wenn zumindest ein Element aus Ω in allen Kontexten $c \in \mathbb{C}_{\tilde{A}}$ enthalten ist[185].

Die Vorgehensweise soll abschließend an einem kurzen Beispiel erläutert werden. Befragt man eine Gruppe von Prüfern, welchen Fehleranteil sie im Rahmen einer Prüfung als sehr hoch bezeichnen würden, und mißt dabei die in Abb. 2.4 angeführten relativen Häufigkeiten der Antworten, so resultiert für die Ausprägung „sehr hoch" der linguistischen Variable FEHLERANTEIL nach Anwendung von (2.109) die in Abb. 2.5 graphisch dargestellte Zugehörigkeitsfunktion.

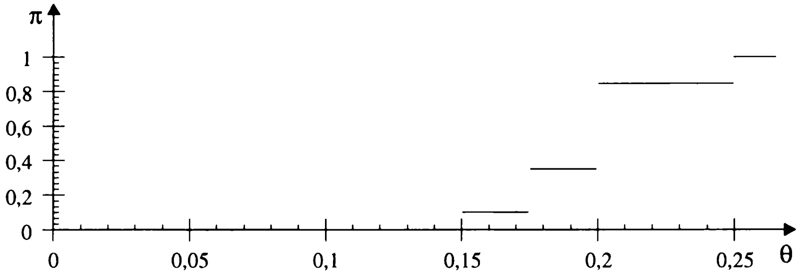

Abb. 2.5: Aus den Beispieldaten in Abb. 2.1 resultierende Zugehörigkeitsfunktion für die Ausprägung „sehr hoch" der linguistischen Variable FEHLERANTEIL

184 Vgl. *Kruse* et al., Fuzzy-Systeme, 1995, S. 45.

185 Vgl. auch *Dubois/Prade*, Statistical Data, 1986, S. 349, die durch die Verwendung geschachtelter Elemente in der Kontextmenge sicherstellen, daß die resultierende Zugehörigkeitsfunktion normiert ist.

C. Abgrenzung von anderen risikoorientierten Prüfungsansätzen

Wie die Analyse der vorhandenen Maße, die für einzelne Formen von Unschärfe der Prüfungsinformationen in Betracht kommen, gezeigt hat, kommt für die Risikomessung nicht nur ein einzelnes Unschärfemaß in Frage. Dabei unterscheiden sich die einzelnen Risikomaße nicht nur durch den mathematischen Apparat, sondern auch bei den möglichen Interpretationen und den Anforderungen, die an die einzelnen Erkenntnisse bzw. Prüfungsinformationen, aus denen eine Maßgröße für das Risiko abgeleitet wird, zu stellen sind. Da die bisher im Schrifttum zum Prüfungswesen diskutierten Ansätze zur Risikomessung, soweit sie eine Quantifizierung des Prüfungsrisikos versuchen, im wesentlichen auf ein einziges Unschärfemaß, namentlich auf das Wahrscheinlichkeitsmaß abstellen, soll schon an dieser Stelle eine Abgrenzung zum hier noch weiter zu entwickelnden Ansatz der Risikomessung bei der Prüfung gemacht werden. Relevante Ansätze, die nicht nur auf eigenes Ermessen des Prüfers im Hinblick auf die Urteilsqualität im Sinne eines ausreichend geringen Risikos abstellen, sind dabei der Audit Risk Ansatz sowie solche Ansätze, die entweder das Theorem von *Bayes* verwenden, oder die Entscheidung für die Beurteilung anhand von likelihoods treffen. Die letzteren werden zusammen als entscheidungslogische Ansätze diskutiert.

I. Audit Risk Ansatz

Der Audit Risk Ansatz, wie er hier untersucht werden soll, geht auf das SAS No. 47 des *AICPA* zurück[186]. Auch wenn der dort vorgestellte Ansatz im wesentlichen zur Prüfungsplanung entwickelt wurde[187], soll hier untersucht werden, inwieweit er zur Risikomessung, die als notwendige Voraussetzung einer risikoorientierten Prüfungsplanung betrachtet werden muß, geeignet ist. Zur Risikomessung werden beim Audit Risk Ansatz Wahrscheinlichkeiten als einziges Unschärfemaß verwendet, ohne daß deutlich wird, welche der möglichen Interpretationen, objektiv oder subjektiv, für die ermittelten Wahrscheinlichkeiten unterstellt wird. Es soll angenommen werden, daß zumindest für das zu ermittelnde gesamte Prüfungsrisiko keine objektive Wahrscheinlichkeit gemeint ist, da diese aus den oben genannten Gründen nicht verwendet werden kann[188]. Obwohl gezeigt wurde, daß auch die Verwendung subjektiver Wahr-

---

[186] Vgl. *AICPA*, SAS No. 47, 1994, AU Sec. 312.

[187] Vgl. *AICPA*, SAS No. 47, 1994, AU Sec. 312 § 1; *Wiedmann*, Prüfungsansatz, 1993, S. 15.

[188] Siehe oben Kapitel B. I. 2. a) in diesem Teil der Arbeit.

scheinlichkeiten zur Messung des Gesamtrisikos erhebliche Probleme mit sich bringt[189], sollen dennoch die weiteren Annahmen des Audit Risk Ansatzes kritisch untersucht werden, sowie die Unterschiede zum vorliegenden Ansatz aufgezeigt werden[190].

1. Risiko- und Fehlerbegriff

Risiko wird im Rahmen des Audit Risk Ansatzes definiert als die Wahrscheinlichkeit, daß es der Prüfer unwissentlich versäumt, seinen Bestätigungsvermerk zu versagen oder einzuschränken, obwohl die Prüfungsgesamtheit tatsächlich wesentliche Fehler enthält[191]. Der Ansatz stimmt demnach dann mit dem hier vorgeschlagenen Prüfungsmodell überein,

- wenn unter einem wesentlichen Fehler in der Prüfungsgesamtheit die Tatsache, daß die Bedingung in (1.27) nicht erfüllt ist, verstanden wird,
- wenn als Handlungsalternative für den Prüfer ausschließlich a_1 betrachtet wird, und
- wenn als Unschärfemaß ein Wahrscheinlichkeitsmaß im Sinne von (2.4) bis (2.6) verwendet wird.

Ein Fehler wird beim Audit Risk Ansatz dann als wesentlich betrachtet, wenn sein Auftreten zur Ablehnung der Ordnungsmäßigkeit der Prüfungsgesamtheit führt. Auch wenn dabei der Schwerpunkt auf die wertmäßige Abweichung zwischen Soll- und Istobjekten der Prüfung, also $\Delta(X,Y)$ gelegt wird[192], und der in (1.18) dargestellte Fall, daß eine Ablehnung der Ordnungsmäßigkeit auch bei einer Vielzahl wertmäßig nicht wesentlicher Fehler erfolgt, nicht explizit berücksichtigt wird, soll angenommen werden, daß die Kriterien, die zu einer Ablehnung der Prüfungsgesamtheit führen, in beiden Prüfungsansätzen weitestgehend übereinstimmen, bzw. beide Ansätze in der Lage sind, die Prü-

---

[189] Siehe oben Kapitel B. I. 2. b) in diesem Teil der Arbeit. Eine Untersuchung über die Auswirkungen von Fehlern bei der Bestimmung der einzelnen subjektiven Wahrscheinlichkeiten, die im Rahmen des Audit Risk Ansatzes benötigt werden, haben *Cushing/Loebbecke* durchgeführt. Vgl. *Cushing/Loebbecke*, Audit Risk, 1983, S. 30-32. Die Ergebnisse werden hier nicht benötigt, da die im folgenden vorgebrachte Kritik am Audit Risk Ansatz grundlegender ist.

[190] Dabei wird auf das wohl am weitesten verbreitete mathematische Modell des *AICPA* aus dem Anhang des SAS No. 39 abgestellt. Vgl. *AICPA*, SAS No. 39, 1994, AU Sec. 350 § 48. Die darüber hinaus diskutierten Varianten verwenden alle ebenfalls absolute Wahrscheinlichkeiten, die multiplikativ verknüpft werden. Die hier im folgenden vorgebrachte Kritik trifft sie demnach in gleichem Maße. Zu einer Übersicht der verwendeten Risikoformeln vgl. *Leslie*, Analysis, 1984, S. 90-92.

[191] Vgl. *AICPA*, SAS No. 47, 1994, AU Sec. 312 § 2; *Diehl*, Prüfungsvorgehen, 1991, S. 192; *Dörner*, Audit Risk, 1992, Sp. 82.

[192] Vgl. *Wiedmann*, Prüfungsansatz, 1993, S. 20f.

fungsrealität ausreichend abzubilden. Wie allerdings schon oben festgestellt wurde, erscheint die Begrenzung der Risikomessung auf den Fall der Erteilung eines Bestätigungsvermerks der hier unterstellten Entscheidungssituation des Prüfers nicht angemessen[193].

Ein weiteres Problem des Audit Risk Ansatzes liegt darin, daß, wie unten noch deutlicher wird[194], die vorgenommenen Wahrscheinlichkeitszuordnungen unabhängig von der angewandten Prüfungsmethode direkt auf das Vorliegen wesentlicher Fehler zielen. Damit gehen zum einen bei der Verwendung empirischer Daten zur Ermittlung von Wahrscheinlichkeiten jeweils zwei Zufallsgrößen in die Schätzung von Fehlerhäufigkeiten ein, namentlich die festgestellten Abweichungen selbst und die jeweilige Bezugsgröße für die Bestimmung, ob ein wesentlicher Fehler vorliegt. Solche Verhältnisschätzer sind eher verzerrt als ein einfacher Mittelwertschätzer, der ausschließlich zur Ermittlung von wertmäßigen Abweichungen bei bestimmten vorliegenden Bedingungen eingesetzt wird. Hinzu kommt, daß in vielen Fällen Abweichungen unabhängig von den für die Bestimmung der Materiality-Grenze verwendeten Bezugsgrößen sind. Zum anderen führen Korrekturen an der Materiality-Grenze dazu, daß alle Wahrscheinlichkeitszuordnungen neu festgelegt werden müssen. Da die Materiality-Grenze für jede Prüfungsgesamtheit individuell festgelegt wird, erscheinen daher Ansätze, die zunächst versuchen, die Ausprägung der für die Beurteilung der vorliegenden Prüfungsgesamtheit relevanten Kriterien zu ermitteln, und dann aus diesen zusammen mit den vorgegebenen Grenzwerten θ^* und Δ^* ein Risikomaß ableiten, eher für die Verwendung empirischer und damit objektivierter Daten sowie für die Analyse von Einflüssen der Materiality-Grenze auf das gemessene Risiko geeignet.

2. Bestimmung von Risikokomponenten

Eine erste Annahme des Audit Risk Ansatzes über die Struktur des Prüfungsrisikos besteht darin, daß sich das zu ermittelnde Gesamtrisiko aus zunächst zwei voneinander unabhängigen Komponenten zusammensetzt, namentlich dem Risiko (Fehlerwahrscheinlichkeit), daß die Prüfungsgesamtheit wesentliche Fehler enthält, und dem Risiko (Entdeckungsrisiko), daß diese Fehler vom Prüfer im Rahmen seiner Prüfungshandlungen nicht aufgedeckt werden[195]. Dabei wird von Vertretern des Audit Risk Ansatzes davon ausge-

---

[193] Siehe oben Kapitel A. III. in diesem Teil der Arbeit.

[194] Siehe unten Kapitel C. I. 2. in diesem Teil der Arbeit.

[195] Vgl. *Dörner*, Audit Risk, 1992, Sp. 82; *Wiedmann*, Prüfungsansatz, 1993, S. 17; *Stibi*,

gangen, daß ausschließlich das zuletzt genannte Risiko durch Prüfungshandlungen bzw. deren Ergebnisse beeinflußt wird[196]. Insoweit können Risiken, die aus der bei der Erfassung und Beurteilung des datenerzeugenden Systems verbleibenden Unschärfe resultieren, bei diesem Ansatz nicht gesondert berücksichtigt werden. Darüber hinaus werden durch diese Zerlegung zwei unterschiedliche Fragestellungen miteinander vermischt. Während für eine Entscheidung über die Ordnungsmäßigkeit der Prüfungsgesamtheit die Kenntnis der Fehlerwahrscheinlichkeit ausreicht, werden die aufgedeckten Fehler ausschließlich zur Begründung des Prüfungsurteils benötigt. Die vorgenommene Zerlegung erscheint demnach ausschließlich dann zweckmäßig, wenn nach Korrektur der aufgedeckten Fehler die Beurteilung der Prüfungsgesamtheit als ordnungsmäßig erfolgen soll.

Für die Fehlerwahrscheinlichkeit wird angenommen, daß sie sich aus dem „inherent risk" und dem „control risk" ableiten läßt. Dabei wird unter „inherent risk" die Wahrscheinlichkeit für eine nicht ordnungsmäßige Prüfungsgesamtheit verstanden, die ermittelt würde, wenn ausschließlich Informationen über die Verarbeitungsfunktionen des datenerzeugenden Systems sowie über den Dateninput bekannt sind. Das „control risk" beschreibt dann die Wahrscheinlichkeit, daß wesentliche Fehler vom internen Kontrollsystem, welches in das datenerzeugende System integriert ist, nicht aufgedeckt werden[197]. Probleme bei der Interpretation der zu messenden Wahrscheinlichkeiten ergeben sich dabei durch den Umstand, daß der zugrundeliegende Möglichkeitsraum, dessen Abgrenzung sowohl für die Aussagekraft der ermittelten Wahrscheinlichkeiten als auch für Überlegungen hinsichtlich der Zulässigkeit der Verwendung eines Wahrscheinlichkeitsmaßes eine Rolle spielt, für den Audit Risk Ansatz nicht definiert wird. Da der Audit Risk Ansatz, wie das hier vorgestellte Prüfungsmodell, eine Aussage über die tatsächliche Ausprägung der Prüfungsgesamtheit X aus den durch Prüfungsinformationen gewonnen Informationen anstrebt, soll davon ausgegangen werden, daß auch hier der Möglichkeitsraum durch alle möglichen Abbildungen der realen Sachverhalte in bezug auf das betrachtete Unternehmen definiert ist.

Damit die Zerlegung der Fehlerwahrscheinlichkeit möglich ist, muß demnach zunächst die Menge der Abbildungen, die bei dem gegebenen Verarbeitungssystem möglich sind, identifiziert werden. Diese Menge soll mit E_1 be-

---

Prüfungsrisikomodell, 1995, S. 103.

[196] Vgl. *Diehl*, Prüfungsvorgehen, 1991, S. 195; *Wiedmann*, Prüfungsansatz, 1993, S. 17; *Stibi*, Prüfungsrisikomodell, 1995, S. 103.

[197] Vgl. *Diehl*, Prüfungsvorgehen, 1991, S. 193f.; *Dörner*, Audit Risk, 1992, Sp. 83.

zeichnet werden. Entsprechend wird die Menge der Abbildungen, die bei dem festgestellten Kontrollsystem möglich sind, E_2 genannt. Dabei darf nicht übersehen werden, daß diese Aufteilung hohe Anforderungen an das Abstraktionsvermögen des Prüfers stellt, da in realen datenerzeugenden Systemen eine Trennung in Verarbeitungs- und Kontrollsystem nur gedanklich vollzogen werden kann. Insoweit erscheint es vorteilhafter eine gemeinsame Wahrscheinlichkeitsbeurteilung der beiden Risiken vorzunehmen[198]. Soweit hierzu Einflußfaktoren angeführt werden, die bei dem zu prüfenden Unternehmen selbst liegen, wie Art, Größe und Lage des Unternehmens oder die Verwertbarkeit von Vermögensgegenständen oder Schulden[199], wurde jedoch gezeigt, daß eine theoretisch begründete Ableitung der Fehlerwahrscheinlichkeit nur sehr eingeschränkt möglich ist[200].

Ausgehend von der im Rahmen des Audit Risk Ansatzes gewählten Kombination der beiden Risiken, muß für die Definition des inherent risk formal $pr(\overline{O} \cap E_1)$ und für das control risk $pr(\overline{O} \cap E_2)$ gelten. Ausschließlich dann ergibt sich nämlich unter der Annahme, daß die beiden Risiken unabhängig im Sinne von (2.18) sind, die im Schrifttum[201] abgeleitete Fehlerwahrscheinlichkeit mit:

(2.110) $$pr(\overline{O} \cap E_1 \cap E_2) = pr(\overline{O} \cap E_1)\,pr(\overline{O} \cap E_2).$$

Das Problem des Audit Risk Ansatzes liegt dabei nicht im erforderlichen Abstraktionsvermögen des Prüfers oder in der notwendigen Unabhängigkeit der beiden Risikokomponenten[202], der Ansatz selbst ist der unterstellten Prüfungssituation nicht angemessen. Die in (2.110) aus inherent risk und control risk abgeleitete Fehlerwahrscheinlichkeit entspricht der a-priori-Wahrscheinlichkeit für das Eintreten wesentlicher Fehler, wenn das durch E_1 und E_2 definierte datenerzeugende System eingesetzt wird[203]. Da dem Prüfer jedoch bereits bekannt ist, daß sowohl E_1 als auch E_2 eingetreten sind, ist die a-priori-Wahrscheinlichkeit für die vorliegende Situation nicht mehr relevant. Diese be-

---

198 Vgl. *Kreutzfeldt*, Response, 1990, S. 143f.

199 Vgl. *IDW*, HFA 1/1988, 1988, S. 241.

200 Vgl. *Obermeier*, Abschlussprüfung, 1983, S. 82-86.

201 Vgl. *Wiedmann*, Prüfungsansatz, 1993, S. 18; *Stibi*, Prüfungsrisikomodell, 1995, S. 156.

202 Vgl. *Cushing/Loebbecke*, Audit Risk, 1983, S. 28. Dabei ist die Unabhängigkeit von $\overline{O} \cap E_1$ und $\overline{O} \cap E_2$ dann nicht entscheidend, wenn gemäß (2.7) die jeweils relevante bedingte Wahrscheinlichkeit verwendet wird. Vgl. *Stibi*, Prüfungsrisikomodell, 1995, S. 145f. Allerdings bleibt fraglich, wie dann die relevanten bedingten Wahrscheinlichkeiten durch den Prüfer ermittelt werden sollen. Vgl. *Stibi*, Prüfungsrisikomodell, 1995, S. 152.

203 Vgl. *Quick*, Risiken, 1996, S. 160.

rücksichtigt noch, daß das Eintreten von E_1 bzw. E_2 nicht sicher ist. Um zu einer adäquaten Abbildung der vorliegenden Informationen zu kommen, ist das inherent risk mit $pr(\overline{O}|E_1)$, das control risk mit $pr(\overline{O}|E_2)$ anzusetzen. Dabei gilt wegen (2.7) folgende Relation:

$$pr(\overline{O}|E_1) \geq pr(\overline{O} \cap E_1). \tag{2.111}$$

Die Gleichheit der beiden Wahrscheinlichkeiten ist dabei nach (2.7) ausschließlich dann gegeben, wenn $pr(E_1)=1$ ist. Dies wäre gleichbedeutend mit einer Einschränkung des Möglichkeitsraumes auf die beim vorliegenden Verarbeitungssystem möglichen Abbildungen der Realität oder der Annahme, daß alle nicht durch das vorliegende Verarbeitungssystem erzeugbaren Abbildungen die Wahrscheinlichkeit Null haben. Selbst wenn eine dieser Annahmen zutreffen sollte, wird, da die gleiche Argumentation für E_2 gilt, zumindest das control risk falsch gemessen. Würde auch $pr(E_2)=1$ gesetzt, dann müßte $E_2 \subseteq E_1$ gelten. Für den Fall, daß die beiden Mengen identisch sind, würden zwar die Einzelrisiken richtig gemessen, allerdings wäre die Ermittlung der Fehlerwahrscheinlichkeit nach (2.110) falsch, da dann gelten muß:

$$pr(\overline{O} \cap E_1 \cap E_2) = pr(\overline{O} \cap E_1) = pr(\overline{O} \cap E_2). \tag{2.112}$$

Auch wenn E_2 eine echte Teilmenge von E_1 sein sollte, muß diese Relation gelten, da dann wegen $pr(E_2)=1$ und (2.6) alle nicht ordnungsmäßigen Abbildungen, die nicht in E_2 liegen, die Wahrscheinlichkeit Null haben müssen. Insoweit bleibt bis zu diesem Punkt festzustellen, daß die zur Messung der einzelnen Komponenten der Fehlerwahrscheinlichkeit verwendeten Wahrscheinlichkeiten im Audit Risk Ansatz in keinem Fall der Entscheidungssituation des Prüfers angemessen sind.

Werden für die Messung des inherent risk und des control risk die bedingten Wahrscheinlichkeiten verwendet, ergibt sich nach (2.20) für die Fehlerwahrscheinlichkeit:

$$pr(\overline{O}|E_1 \cap E_2) = \frac{pr(E_2|\overline{O} \cap E_1)\, pr(\overline{O}|E_1)}{pr(E_2|E_1)}. \tag{2.113}$$

Diese kann, wie zu erkennen ist, selbst wenn statistische Unabhängigkeit von E_1 und E_2 nach (2.18) sowie die bedingte Unabhängigkeit der beiden Ereignisse nach (2.21) im Kontext $\overline{O}$ angenommen wird, nicht durch eine multiplikative Verknüpfung der beiden Wahrscheinlichkeiten ermittelt werden, da dann wegen (2.10) gilt:

$$(2.114) \qquad \frac{pr(E_2|\overline{O})pr(\overline{O}|E_1)}{pr(E_2)} = \frac{pr(\overline{O}|E_1)}{pr(\overline{O})} pr(\overline{O}|E_2) \geq pr(\overline{O}|E_1)pr(\overline{O}|E_2).$$

Somit ist auch die gewählte Verknüpfung der beiden Komponenten der Fehlerwahrscheinlichkeit falsch, da sie ausschließlich dann zum gleichen Ergebnis kommt, wenn die a-priori-Wahrscheinlichkeit für die Nichtordnungsmäßigkeit Eins ist. Falls aber die a-priori-Wahrscheinlichkeit von $\overline{O}$ auf Eins gesetzt wird, kann auch die bedingte Wahrscheinlichkeit für die Nichtordnungsmäßigkeit nicht geringer als Eins werden. Es bleibt demnach festzustellen, daß die im Rahmen des Audit Risk Ansatzes ermittelte Fehlerwahrscheinlichkeit, die eine Komponente des zu ermittelnden Gesamtrisikos darstellt, die Situation des Prüfers in keinem Fall richtig abbildet.

Für das Entdeckungsrisko wird ebenfalls angenommen, daß es sich aus mindestens zwei Komponenten zusammensetzt, namentlich der Wahrscheinlichkeit, daß die durchgeführten analytischen Prüfungshandlungen wesentliche Fehler nicht aufdecken konnten, und der Wahrscheinlichkeit, daß wesentliche Fehler im Rahmen der ergebnisorientierten Einzelfallprüfung nicht aufgedeckt werden[204]. Auch diese beiden Komponenten werden wiederum multiplikativ verknüpft, so daß die oben vorgebrachte Kritik auch hier zutrifft, da es sich bei den beiden zu bestimmenden Wahrscheinlichkeiten zumindest dann, wenn der Prüfer seine Entscheidung auf der Basis der vorliegenden Prüfungsinformationen treffen will, um bedingte Wahrscheinlichkeiten handeln muß. Darüber hinaus treten bei der Bestimmung der relevanten Wahrscheinlichkeitszuordnungen Probleme auf.

Während gegen die Messung des nach der Durchführung von Einzelfallprüfungen verbleibenden Risikos durch Wahrscheinlichkeiten zumindest bis zu dieser Stelle keine Einwendungen zu erheben sind, ist diese Vorgehensweise für die Bestimmung des nach der Durchführung analytischer Prüfungshandlungen verbleibenden Risikos problematisch. Wie schon gezeigt wurde, läßt die Struktur der logischen Schlußform, die der Auswertung von analytischen Prüfungshandlungen zugrundeliegt, regelmäßig nur eine Risikoverminderung im Hinblick auf die Handlungsalternative a_2 zu, wenn Abweichungen zwischen ermittelter und prognostizierter Kennzahl gefunden werden[205]. Für den Fall,

[204] Vgl. *Wiedmann*, Prüfungsansatz, 1993, S. 18; *Stibi*, Prüfungsrisikomodell, 1995, S. 103; *AICPA*, SAS No. 39, 1994, AU Sec. 350.48 § 4.

[205] Siehe oben Kapitel B. II. 2. im Ersten Teil der Arbeit.

daß im Rahmen der analytischen Prüfungshandlungen jedoch keine Abweichungen festgestellt werden, können diese zwar bestimmte Fehler ausschließen und sprechen insoweit gegen das Vorliegen einer nicht ordnungsmäßigen Prüfungsgesamtheit, aber sie können nicht die Wahrscheinlichkeit erhöhen, daß die Prüfungsgesamtheit ordnungsmäßig ist. Insoweit können sie also nicht zur Risikoverminderung der Handlungsalternative a_1 beitragen. Diese Zusammenhänge lassen sich wegen der Verwendung von Wahrscheinlichkeiten als Unschärfemaß nicht abbilden. Soweit dort Sicherheitsbeiträge aus analytischen Prüfungshandlungen unterstellt werden[206], gelten diese entweder ausschließlich für die Versagung des Bestätigungsvermerkes, oder es wird, ohne die den verwendeten Gesetzmäßigkeiten zugrundeliegende Struktur zu beachten, die Wahrscheinlichkeit für die Ordnungsmäßigkeit der Prüfungsgesamtheit erhöht, ohne daß dies in irgendeiner Form theoretisch begründet werden könnte.

Insgesamt bleibt festzustellen, daß der Audit Risk Ansatz nicht geeignet erscheint, das Risiko, welches der Prüfer mit der Entscheidung für eines der beiden möglichen Urteile eingeht, zu ermitteln, zumal für die Unschärfe von Prüfungsinformationen, die grundsätzlich nicht durch Wahrscheinlichkeiten gemessen werden kann, keinerlei Möglichkeit zur Berücksichtigung bei der Risikomessung besteht. Darüber hinaus ist, wie die Analyse des Ansatzes gezeigt hat, nicht nur die Bestimmung der einzelnen Wahrscheinlichkeiten, die in die Messung des Gesamtrisikos eingehen, problematisch, sondern auch die Verknüpfung der verwendeten Wahrscheinlichkeiten nicht der unterstellten Prüfungssituation angemessen, so daß auch das Argument, der Ansatz bilde zumindest näherungsweise die Zusammenhänge im Hinblick auf die Ermittlung des Gesamtrisikos ab[207], nicht greift.

3. Auswahl von Prüfungshandlungen

Zur Ableitung von Handlungsanweisungen werden im Schrifttum zum Audit Risk Ansatz den einzelnen Prüfungsmethoden bestimmte Sicherheitsbeiträge, die ein Maß für die Risikoverringerung je Prüfungshandlung bzw. Prüfungszeiteinheit darstellen sollen, zugeordnet, ohne daß allerdings gezeigt wird, wie diese analytisch oder empirisch abzuleiten wären. Sind die Sicherheitsbeiträge der zur Verfügung stehenden Prüfungsmethoden bekannt, läßt sich die optimale Methodenkombination einfach bestimmen[208]. Schon die bisher angeführ-

---

206 Vgl. *Dörner*, Audit Risk, 1992, Sp. 87; *Wiedmann*, Prüfungsansatz, 1993, S. 18.

207 Vgl. z. B. *AICPA*, SAS No. 39, 1994, AU Sec. 350.48 § 4.

208 Vgl. z. B. *Dörner*, Audit Risk, 1992, Sp. 87f.; *Wiedmann*, Prüfungsansatz, 1993, S. 18f.

ten Einwendungen gegen den Audit Risk Ansatz zeigen, daß die Ableitung von Handlungsanweisungen für die Prüfungsdurchführung aus den ermittelten Risikomaßen nicht angemessen erscheint.

Die Eignung der im Rahmen des Audit Risk Ansatzes vorgeschlagenen Methoden zur Prüfungsplanung ist aber auch aus einem zweiten schwerwiegenderen Grund in Frage zu stellen. Wie im oben formulierten Entscheidungsmodell für den Prüfer deutlich wurde, hängt das Risiko nicht von der angewandten Prüfungsmethode, sondern von den dadurch erzielten Prüfungsergebnissen, im hier unterstellten Prüfungsmodell von der Ausprägung der erhaltenen Information, ab[209]. Da Asymmetrie im Hinblick auf die erreichbare Sicherheit für die Beurteilung der Prüfungsgesamtheit in dem Sinne besteht, daß zwar in bestimmten Fällen mit Sicherheit festgestellt werden kann, ob die Prüfungsgesamtheit nicht ordnungsmäßig ist, aber nie eine sichere Aussage über die Ordnungsmäßigkeit erfolgen kann[210], müssen die Sicherheitsbeiträge einzelner Prüfungshandlungen auch davon abhängen, ob die ermittelten Ergebnisse eher für die Ordnungsmäßigkeit oder für die Nichtordnungsmäßigkeit der Prüfungsgesamtheit sprechen. Diese müssen zumindest im ersten Fall mit zunehmender Urteilssicherheit für die Annahme der Ordnungsmäßigkeit unabhängig von der eingesetzten Prüfungsmethode gegen Null gehen.

Insoweit kann ohne Berücksichtigung der möglichen Ergebnisse von Prüfungsmethoden keine Abschätzung ihrer Sicherheitsbeiträge erfolgen. Dies gilt unabhängig davon, ob, wie beim Audit Risk Ansatz, als Prüfungsrisiko ausschließlich die Unschärfe bei Wahl der Handlungsalternative a_1 betrachtet wird, oder ob die beiden möglichen Fehlurteile in die Betrachtung einbezogen werden. Demnach ist auch die Art und Weise, wie der Audit Risk Ansatz zur Prüfungsplanung eingesetzt wird, der vorliegenden Prüfungssituation nicht angemessen.

II. Entscheidungslogische Ansätze

Die beiden im folgenden gemeinsam untersuchten Gruppen von Ansätzen zur Risikoermittlung im Rahmen von Prüfungen sind keine umfassenden Ansätze in dem Sinn, daß sie eine Risikomessung unabhängig von der verwendeten Prüfungsmethode für das im Rahmen der Prüfung abzugebende Gesamturteil anstreben. Sie finden vielmehr nach der Durchführung indirekter Messun-

[209] Siehe insbesondere oben Kapitel C. II. im Ersten Teil der Arbeit.

[210] Siehe dazu auch die Ausführungen zu Beginn von Kapitel B. I. im Ersten Teil der Arbeit.

DOI https://doi.org/10.3790/978-3-428-48909-1

gen im Rahmen von Einzelfallprüfungen Anwendung. Im folgenden sollen zum einen Ansätze, welche die Entscheidung des Prüfers für eine der beiden Handlungsalternativen auf der Basis des Theorems von *Bayes* unterstützen, zum anderen Ansätze der statistischen Testtheorie[211] unterschieden werden. Während die zuerst genannten Ansätze direkt in das oben formulierte Entscheidungsmodell für die Urteilsfindung[212] integriert werden können, verwenden die zuletzt genannten Ansätze Entscheidungsregeln, die ausschließlich auf likelihoods beruhen. Zuvor soll jedoch das den beiden Ansätzen gemeinsame Risikomaß und der unterstellte Fehlerbegriff untersucht werden.

1. Risikomaß und Fehlerbegriff

Soweit die Ansätze objektive Wahrscheinlichkeiten verwenden, die durch die Anwendung mathematisch-statistischer Stichprobenverfahren ermittelt wurden, stimmen sie hinsichtlich des für die Risikomessung verwendeten Ansatzes weitgehend mit dem hier unterstellten Prüfungsmodell überein. Die Darstellung der entsprechenden Prüfungsmethoden ist zumindest in den Grundzügen schon erfolgt[213], die Messung des Risikos im Rahmen dieser Prüfungsmethoden ist Gegenstand des nächsten Teils der Arbeit. Unterschiede ergeben sich dabei durch die Art und Weise, wie die übrigen nicht aus Zufallsstichproben stammenden Prüfungsinformationen in das jeweilige Modell integriert werden. Bei den auf dem Theorem von *Bayes* beruhenden Ansätzen werden Prüfungsinformationen aus indirekten Messungen zur Ableitung der benötigten a-priori-Wahrscheinlichkeiten verwendet[214]. Insoweit werden bei diesen Ansätzen sowohl objektive als auch subjektive Wahrscheinlichkeiten zu einem dann insgesamt nicht mehr objektiven Risikomaß kombiniert. Bei der Anwendung von Verfahren der statistischen Testtheorie sollen schon vorliegende Prüfungsinformationen ausschließlich bei der Festlegung der Testparameter berücksichtigt werden können[215]. Daher dient die weitere Abgrenzung eher der Analyse der Bereiche des Prüfungsmodells, die bei den entscheidungslogischen Ansätzen nicht oder nur sehr komprimiert berücksichtigt werden können.

---

211 Auch wenn die Anwendung von Schätzverfahren, wie unten in Kapitel A. I. 1. a) im Dritten Teil der Arbeit gezeigt wird, grundsätzlich zur Risikomessung geeignet ist, wird im Schrifttum vielfach davon ausgegangen, daß im Hinblick auf die Urteilsbildung ausschließlich Testverfahren in Betracht kommen. Vgl. *Obermeier*, Abschlussprüfung, 1983, S. 23; *Zimmermann*, Stichprobenprüfung, 1992, Sp. 1875.

212 Siehe oben insbesondere Kapitel C. 2. im Ersten Teil der Arbeit.

213 Siehe oben Kapitel B. I. 1. im Ersten Teil der Arbeit.

214 Vgl. *Wittmann*, Systemprüfung, 1980, S. 251.

215 Vgl. *Obermeier*, Abschlussprüfung, 1983, S. 96f.; *IDW*, HFA 1/1988, 1988, S. 243.

Wie schon bei der Darstellung des hier verwendeten Prüfungsmodells angeführt wurde[216], verwenden die meisten im Schrifttum diskutierten Vorschläge zur Anwendung von mathematisch-statistischen Stichprobenverfahren im Rahmen der Prüfung, soweit sie zur Beurteilung des Fehlerausmaßes für eine Prüfungsgesamtheit herangezogen werden, das in (1.23) definierte Distanzmaß, das auf einen Gesamtwertvergleich zwischen X und Y abstellt. Da dieses ausschließlich dann mit dem hier verwendeten Maß für das Fehlerausmaß übereinstimmt, wenn Abweichungen nur in einer Richtung auftreten können, oder ausschließlich der Gesamtwert Gegenstand des Prüfungsurteils ist, gilt die oben zu dem Distanzmaß (1.23) vorgebrachte Kritik auch für die entscheidungslogischen Ansätze. Soweit die Beurteilung des Fehleranteils einer Prüfungsgesamtheit anhand von Zufallsstichproben in den hier betrachteten Ansätzen untersucht wird, stimmt die Konzeption der mathematisch-statistischen Verfahren mit der Konzeption des vorliegenden Prüfungsmodells überein. Trotz der teilweise konzeptionellen Unterschiede lassen sich die Ansätze jedoch im Hinblick auf die Vorgehensweise bei der Risikomessung mit dem vorliegenden Prüfungsmodell vergleichen. Unterschiede ergeben sich für die beiden Gruppen entscheidungslogischer Ansätze insbesondere bei der Integration von anderen, nicht aus mathematisch-statistischen Verfahren stammenden Prüfungsinformationen.

2. Ansätze der statistischen Testtheorie

Für die auf der statistischen Testtheorie beruhenden Ansätze könnten zum einen die Parameter α und β, welche die Sicherheit des Tests festlegen, und zum anderen die Werte für die zu konkretisierenden Hypothesen aus den übrigen Prüfungsinformationen abgeleitet werden. Die Festlegung der Hypothesenwerte erfolgt, falls die Stichprobe aus der Prüfungsgesamtheit und nicht aus einer Teilgesamtheit gezogen wird, unabhängig von den bis zur Durchführung der Einzelfallprüfungen gewonnenen Erkenntnissen über das Eintreten eines der Umweltzustände. Sie werden regelmäßig aus den festgelegten Grenzwerten θ^* und Δ^* so abgeleitet, daß in der Nullhypothese die Ordnungsmäßigkeit und in der Alternativhypothese die Nichtordnungsmäßigkeit von X behauptet wird[217]. Die bisher vorliegenden Prüfungsinformationen können demnach ausschließlich bei der Ableitung der Obergrenzen für die zugelassenen Wahr-

[216] Siehe oben Kapitel A. 3. im Ersten Teil der Arbeit.

[217] Für die folgenden Ausführungen wird angenommen, daß dies ohne Probleme möglich ist. Siehe aber dazu auch Kapitel A. I. 1. b) im Dritten Teil der Arbeit.

scheinlichkeiten eines Fehlers 1. und 2. Art berücksichtigt werden. Da die beiden möglichen Fehler auf die verwendete Entscheidungsregel zurückgeführt werden können, soll diese zunächst kurz analysiert werden.

Die Entscheidungsregel bei den auf der statistischen Testtheorie beruhenden Verfahren für bzw. gegen eine der beiden Hypothesen beruht auf einem Vergleich von likelihoods. Wird wiederum ein beliebiger Parameter ξ der Prüfungsgesamtheit, für den $(\xi \geq 0)$ gelten soll, betrachtet[218], dann wird der Möglichkeitsraum Ω (Stichprobenraum) durch die möglichen Werte, welche die zugehörige Stichprobenfunktion $Z(Y_1, Y_2, \ldots, Y_n)$ annehmen kann, gebildet. Mit A_1 wird der Teil des Stichprobenraumes bezeichnet, für den gilt:

(2.115) $$A_1 = \left\{x \middle| x \in \Omega \wedge x \leq z^*\right\}.$$

Die bedingte Wahrscheinlichkeit für das Eintreten von A_1, wenn die Nullhypothese gilt, läßt sich dann, wenn n ausreichend groß gewählt wird[219], unter Berücksichtigung der Obergrenze für einen Fehler 1. Art bestimmen durch:

(2.116) $$pr(A_1|O) = pr(A_1|\xi = \xi_0) = F_Z(z^*, \xi_0, n) = 1 - \alpha.$$

Bei Gültigkeit der Alternativhypothese ergibt sich entsprechend:

(2.117) $$pr(A_1|\overline{O}) = pr(A_1|\xi = \xi_1) = F_Z(z^*, \xi_1, n) = \beta.$$

In Abhängigkeit vom Ergebnis der realisierten Stichprobe wird, falls $z(y_1, y_2, \ldots, y_n) \leq z^*$ gilt, also A_1 eingetreten ist, die Alternativhypothese zugunsten der Nullhypothese abgelehnt, da ansonsten unter der Voraussetzung, daß β genügend klein gegenüber $1 - \alpha$ gewählt wurde, akzeptiert werden müßte, daß bei Gültigkeit der Alternativhypothese ein unwahrscheinliches Ereignis eingetreten ist, während gleichzeitig eine Hypothese zur Verfügung steht, die das eingetretene Ereignis besser stützt[220]. Tritt $\overline{A_1}$ ein, wird mit der entsprechenden Argumentation die Nullhypothese abgelehnt und die Alternativhypothese angenommen. Die Anwendung dieser Entscheidungsregel führt somit für den Fall, daß die Prüfungsgesamtheit ordnungsmäßig ist, höchstens mit der Wahrscheinlichkeit α, für den Fall, daß die Prüfungsgesamtheit nicht ordnungsmäßig ist, höchstens mit der Wahrscheinlichkeit β zu einer Fehlentscheidung. Die vor Durchführung des Testverfahrens vorliegenden Prüfungsinformationen sollen

---

[218] Es werden an dieser Stelle die oben in Kapitel B. I. 1. im Ersten Teil der Arbeit definierten Symbole verwendet.

[219] Siehe unten Kapitel A. I. 1. b) aa) im Dritten Teil der Arbeit.

[220] Vgl. *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 167.

durch E ausgedrückt werden. Soweit aus diesen eine Wahrscheinlichkeitsaussage im Hinblick auf die tatsächliche Ausprägung der Prüfungsgesamtheit in der Form $pr(O|E)$ abgeleitet werden kann, spielt diese Information für die aus dem Testverfahren resultierenden Aussagen keine Rolle. Da die likelihoods schon unter der Voraussetzung bestimmt wurden, daß die jeweilige Hypothese wahr ist, können Wahrscheinlichkeitsaussagen über die Ordnungsmäßigkeit der Prüfungsgesamtheit, also die Wahrheit der Hypothesen, die Entscheidungsregel nicht beeinflussen. Selbst die sichere Erkenntnisse über die Wahrheit der Nullhypothese oder der Alternativhypothese ändert nichts an den aus der Zufallsauswahl der Stichprobenelemente resultierenden Fehlentscheidungen bei der Durchführung des Tests, auch wenn dieser dann nicht mehr notwendig wäre. Insoweit können also Prüfungsinformationen der genannten Form keinen Einfluß auf die Wahl der Werte für α und β haben[221].

Fraglich ist, welche Form von Prüfungsinformationen überhaupt einen Einfluß auf die für die mathematisch-statistischen Testverfahren relevanten Parameter α und β haben können. Geht man davon aus, daß bei Vorliegen der Erkenntnisse E und einem Stichprobenergebnis aus A_1 die obige likelihood Regel für die Beurteilung der Grundgesamtheit verwendet werden soll, so muß für die Annahme der Ordnungsmäßigkeit gelten:

(2.118) $$pr\left(A_1 \cap E|O\right) \geq 1-\alpha' \text{ und}$$

(2.119) $$pr\left(A_1 \cap E|\overline{O}\right) \leq \beta'.$$

Die Entscheidungsregel ist demnach bis auf den Unterschied, daß auch die vor Durchführung des Tests vorliegenden Prüfungsergebnisse einbezogen werden, identisch. Dabei sollen α' und β' das maximale Risiko einer Fehlentscheidung, die jeweils für das Prüfungsurteil zugelassen wird, angeben. Unter der Voraussetzung, daß die vorliegenden Erkenntnisse E und das Stichprobenergebnis A_1 bedingt unabhängig im Kontext O bzw. $\overline{O}$ sind[222], lassen sich gemäß (2.22) die angegebenen Bedingungen umformen zu:

(2.120) $$pr\left(\mathrm{A}_1|O\right) pr\left(E|O\right) \geq 1-\alpha' \text{ und}$$

[221] Dieser Zusammenhang wird anscheinend bei den meisten Vorschlägen zur Verknüpfung von Vorinformationen und Sicherheitsgrad mathematisch-statistischer Verfahren übersehen. Vgl. z. B. *Elliot/Rogers*, Sampling, 1972, S. 49f.; *Leffson*, Wirtschaftsprüfung, 1988, S.256; *Adenauer*, Kontrollsystem, 1989, S. 213f.

[222] Diese Voraussetzung dürfte regelmäßig erfüllt sein, solange die Vorinformationen E_2 nicht aus der gleichen Stichprobe stammen, die für den Test herangezogen wird.

(2.121) $$pr\left(A_1\middle|\overline{O}\right)pr\left(E\middle|\overline{O}\right)\leq\beta'.$$

Für die bei dem mathematisch-statistischen Testverfahren zugelassenen Fehlerwahrscheinlichkeiten läßt sich aus (2.116) und (2.120) sowie (2.117) und (2.121) dann ableiten:

(2.122) $$1-\alpha=\frac{1-\alpha'}{pr\left(E\middle|O\right)}\text{ und}$$

(2.123) $$\beta=\frac{\beta'}{pr\left(E\middle|\overline{O}\right)}.$$

Wie zu erkennen ist, können ausschließlich für den Fall, daß die Wahrscheinlichkeit für die bisherigen Prüfungsergebnisse E im Fall einer ordnungsmäßigen Prüfungsgesamtheit Eins ist, α und α' gleichgesetzt werden. In allen anderen Fällen muß $\alpha<\alpha'$ gelten. Umgekehrt gilt für β, daß es größer gewählt werden kann, wenn die Eintrittswahrscheinlichkeit für die vorliegenden Prüfungsergebnisse E bei einer tatsächlich nicht ordnungsmäßigen Prüfungsgesamtheit geringer als Eins ist. Die entsprechende Argumentation läßt sich auch für den Fall, daß ein Stichprobenergebnis aus $\overline{A}_1$ eintritt und die Alternativhypothese gegen die Nullhypothese angenommen werden soll, anführen.

Eine Berücksichtigung von vorliegenden Prüfungsinformationen bei der Festlegung der zulässigen Irrtumswahrscheinlichkeiten mathematisch-statistischer Testverfahren kann demnach ausschließlich dann erfolgen, wenn diese in bedingte Wahrscheinlichkeiten bei gegebener Ordnungsmäßigkeit bzw. Nichtordnungsmäßigkeit transformiert werden können. Dabei sind zwei Dinge zu beachten. Zum einen lassen sich nur solche Unschärfen berücksichtigen, die grundsätzlich durch Wahrscheinlichkeiten gemessen werden können. Zum anderen erfordert die Abschätzung der Wahrscheinlichkeiten eine bestimmte Form der möglichen Schlußfolgerungen, die aus den Prüfungsinformationen möglich sind.

Angenommen die vorliegenden Prüfungsinformationen beschreiben den Zustand eines geprüften Ersatztatbestandes und wurden somit im Rahmen indirekter Messungen ermittelt. Es soll gelten, daß sich E und die wahre logische Aussage z entsprechen. Damit $\alpha=\alpha'$ gewählt werden kann, muß $pr\left(E\middle|O\right)=1$ sein. Dies entspricht dem Fall, daß die logische Implikation $\mathrm{o}\rightarrow\mathrm{z}$ immer wahr ist. Entsprechend darf, falls eine Erhöhung von β erfolgen soll, die Wahrheit der Implikation $\sim\mathrm{o}\rightarrow\mathrm{z}$ nicht für jeden Fall behauptet werden. Eine solche Form der Schlußfolgerung liegt ausschließlich den analytischen Prüfungshandlungen zugrunde[223]. Im Rahmen der Prüfung des datenerzeugenden Systems wird vom Zustand und der Funktionsweise des Systems auf die Ordnungsmäßigkeit der Prüfungsgesamtheit geschlossen. Es wird demnach, falls E die Erkenntnisse über das datenerzeugende System beinhaltet, ein Wert für $pr(O|E)$ bestimmt. Dieser läßt sich wegen (2.10) ohne Kenntnis der absoluten Wahrscheinlichkeit für die Ordnungsmäßigkeit und der absoluten Wahrscheinlichkeit für das Auftreten des vorliegenden datenerzeugenden Systems nicht in die benötigte Wahrscheinlichkeitsaussage transformieren. Da die angeführten absoluten Wahrscheinlichkeiten in der Regel nicht zur Verfügung stehen werden, ist eine Berücksichtigung der Ergebnisse der Systemprüfung zur Verbesserung der Aussagen von statistischen Tests nicht möglich. Die gleiche Argumentation gilt für alle Prüfungsergebnisse über Sachverhalte, für die zwar der Einfluß auf die Ordnungsmäßigkeit der Prüfungsgesamtheit, nicht aber der Einfluß der Ordnungsmäßigkeit auf die festgestellten Sachverhalte bekannt ist. Insoweit können Informationen über die Fehlerwahrscheinlichkeit in der Prüfungsgesamtheit selbst und alle darauf wirkenden Einflüsse, wie z. B. die Art, Größe und Lage des Unternehmens, keinen Einfluß auf die Sicherheit der mathematisch-statistischen Testverfahren haben.

3. Verwendung des Theorems von Bayes

Im Gegensatz zu den Ansätzen, die auf der statistischen Testtheorie beruhen, wird in einer zweiten Gruppe entscheidungslogischer Ansätze, ähnlich wie beim Audit Risk Ansatz, versucht, Wahrscheinlichkeiten für die Ordnungsmäßigkeit bzw. Nichtordnungsmäßigkeit der Prüfungsgesamtheit zu bestimmen. Dazu wird vorgeschlagen, das Theorem von *Bayes* anzuwenden[224]. Eine Entscheidung über die Annahme oder Ablehnung der Ordnungsmäßigkeit wird dann anhand der a-posteriori-Wahrscheinlichkeiten oder dem a-posteriori-Wahrscheinlichkeitsverhältnis getroffen[225]. Zur Bestimmung der a-posteriori-Wahrscheinlichkeiten werden die vorliegenden Prüfungsergebnisse aus anderen Prüfungshandlungen mit den aus einer Zufallsstichprobe ermittelten likelihoods mittels (2.14) verknüpft. Bevor auf die mit der Anwendung des Theorems von *Bayes* verbundenen Probleme eingegangen werden kann, muß der Möglichkeitsraum, auf dem das resultierende Wahrscheinlichkeitsmaß definiert ist, abgegrenzt werden.

---

[223] Siehe oben Kapitel B. II. 2. im Ersten Teil der Arbeit.

[224] Vgl. *Leffson*, Wirtschaftsprüfung, 1988, S. 256f.; *v. Wysocki*, Grundlagen, 1988, S. 218f.; *Coenenberg/Hanisch*, Stichprobenprüfung, 1992, Sp. 1870; *Quick*, Risiken, 1996, S. 148ff.

[225] Vgl. *Hömberg*, Prüfungen, 1981, S. 39f.

Bisher wurde der Möglichkeitsraum bei der Anwendung mathematisch-statistischer Verfahren durch die Stichprobenergebnisse gebildet, die für die betrachtete Prüfungsgesamtheit beim unterstellten Auswahlverfahren der Stichprobenelemente möglich waren. Es wurde insoweit unterstellt, daß der interessierende Parameter der Prüfungsgesamtheit zwar unbekannt aber eindeutig ist. Die Anwendung des Theorems von *Bayes* setzt allerdings voraus, daß der betrachtete wahre Parameter der Prüfungsgesamtheit als Zufallsgröße angesehen werden kann, da ausschließlich dann a-priori-Wahrscheinlichkeiten für mögliche Werte dieses Parameters angegeben werden können[226]. Für die Anwendung des Theorems von *Bayes* auf die Ergebnisse von Zufallsstichproben muß für den Möglichkeitsraum $\Omega = \Omega_1 \times \Omega_2$ gelten. Dabei bezeichnet Ω_1 alle möglichen Abbildungen der Realität für das betrachtete Unternehmen und Ω_2 alle möglichen Stichprobenergebnisse bei der jeweils betrachteten Abbildung. D. h. der Prüfer muß sich zur Ableitung von Wahrscheinlichkeiten nicht nur alle Abbildungen der Realität, die in bezug auf das zu prüfende Unternehmen möglich sind, vergegenwärtigen können, sondern auch zumindest die für die Verteilung der Stichprobenergebnisse notwendigen Parameter für die einzelnen Abbildungen unterscheiden können. Insoweit ist das zentrale Problem der Anwendung des Theorems von *Bayes* die Verwendung subjektiver a-priori-Wahrscheinlichkeiten.

Dabei kann zum einen in der Verwendung subjektiver Wahrscheinlichkeiten selbst ein Problem gesehen werden, da deren Wert das Risiko nicht unerheblich beeinflussen kann und keine intersubjektiv überprüfbaren Verfahren zu ihrer Festlegung existieren[227]. Zum anderen wird ausschließlich in der Transformation der vorliegenden Informationen in Wahrscheinlichkeiten ein Problem gesehen[228]. Wie die Diskussion der unterschiedlichen Unschärfemaße gezeigt hat[229], ergeben sich für die im Rahmen der Prüfung vorliegenden Formen von Unschärfe nur sehr begrenzte Möglichkeiten, objektive Unschärfemaße zu verwenden, so daß ein genereller Verzicht auf subjektive Maßgrößen bedeuten würde, daß der überwiegende Teil der vorhandenen Prüfungsergebnisse zur Bestimmung des Prüfungsrisikos überhaupt nicht genutzt werden könnte. Zu fragen ist allerdings, ob bei der Verwendung subjektiver Einschätzungen ein Maß gewählt werden sollte, das so hohe Anforderung an die benötigte Struktur

226 Vgl. *Bücker*, Lösungsmöglichkeiten, 1973, S. 26.

227 Vgl. *Mandl*, Anwendungsvoraussetzungen, 1984, S. 149. Siehe auch oben Kapitel B. I. 2. b) in diesem Teil der Arbeit.

228 Vgl. *Wittmann*, Systemprüfung, 1980, S. 227f.; *Leffson*, Wirtschaftsprüfung, 1988, S. 258;

229 Siehe oben Kapitel B. in diesem Teil der Arbeit.

der vorliegenden Informationen stellt wie das Wahrscheinlichkeitsmaß[230]. Zu bemängeln ist außerdem, daß versucht wird, die gesamten vorliegenden Prüfungsinformationen durch eine einzige a-priori-Wahrscheinlichkeitsbeurteilung des betrachteten Parameters der Prüfungsgesamtheit auszudrücken, ohne daß der Versuch gemacht wird, schon die einzelnen Einflußgrößen quantitativ zu erfassen und mittels eines akzeptierten Verfahrens zu aggregieren. Dies gilt unabhängig davon, ob versucht wird durch die Angabe kumulierter Wahrscheinlichkeiten zu einzelnen Klassen direkt die a-priori-Verteilung des betrachteten Parameters der Prüfungsgesamtheit zu bestimmen[231], oder ob die Parameter einer stetigen Verteilung, welche die gesuchte Verteilung des betrachteten Parameters der Prüfungsgesamtheit hinreichend genau approximiert, ermittelt werden[232]. In beiden Fällen wird versucht, alle vorhandenen Informationen unmittelbar in die benötigten a-priori-Wahrscheinlichkeiten zu transformieren.

Auf die unterschiedlichen Ansätze zur Ableitung der a-priori-Wahrscheinlichkeitsverteilung des betrachteten Parameters bei unterschiedlichen Prüfungsmethoden soll an dieser Stelle nicht eingegangen werden[233]. Allerdings ist festzuhalten, daß die ermittelten a-priori-Wahrscheinlichkeiten regelmäßig keine absoluten Wahrscheinlichkeiten sein werden, da ansonsten die oben angesprochenen Probleme der Anwendung des Theorems von *Bayes* auftreten werden[234], da a priori beim hier unterstellten Möglichkeitsraum die Wahrscheinlichkeit, daß die Prüfungsgesamtheit tatsächlich ordnungsmäßig ist, nahe bei Null sein wird. Insoweit handelt es sich bei den zu ermittelnden a-priori-Wahrscheinlichkeiten um bedingte Wahrscheinlichkeiten. Dann ist zumindest zu prüfen, ob die neuen Erkenntnisse, die durch die Zufallsstichprobe gewonnen werden, und die schon vorliegenden Erkenntnisse bedingt unabhängig im Sinne von (2.22) im Kontext O und $\overline{O}$ sind.

Selbst wenn es gelingt, ausreichend objektivierbare a priori Wahrscheinlichkeiten aus den vorliegenden Prüfungsinformationen abzuleiten, ergeben sich bei der Anwendung des Theorems von *Bayes* im Hinblick auf die Prüfungsplanung Probleme. Diese bestehen nur vordergründig in dem Umstand, daß der Stichprobenumfang ex ante nicht in Abhängigkeit von der zu treffenden Ur-

---

230 Siehe auch oben Kapitel B. II. 2. a) in diesem Teil der Arbeit.

231 Vgl. zu dieser Vorgehensweise insbesondere den Vorschlag von *Bücker*, Lösungsmöglichkeiten, 1973, S. 55.

232 Vgl. *Wittmann*, Systemprüfung, 1980, S. 237-243.

233 Soweit a-priori-Informationen in a-priori-Wahrscheinlichkeiten transformiert werden können und diese Ansätze für das hier vorliegende Prüfungsmodell relevant sind, werden sie im Dritten Teil der Arbeit behandelt.

234 Siehe oben Kapitel B. I. 2. a) in diesem Teil der Arbeit.

teilsqualität bestimmt werden kann[235]. Wie in (2.15) zu erkennen ist, gilt generell, daß die Veränderung der Wahrscheinlichkeitseinschätzung von der Sensitivität und Spezifität des erhaltenen Stichprobenergebnisses abhängt. Insoweit kann weder die Richtung, in die sich die Wahrscheinlichkeitseinschätzung im Hinblick auf die Ordnungsmäßigkeit oder Nichtordnungsmäßigkeit der Prüfungsgesamtheit verändert, noch der Umfang der Veränderung abgeschätzt werden. Kenntnisse über beide Größen sind aber für die Prüfungsplanung erforderlich.

Als letzter Einwand, der gegen die Anwendung des Theorems von *Bayes* spricht, ist vorzubringen, daß ausschließlich Prüfungsinformationen, die in Form sicherer Erkenntnisse vorliegen, verwertbar sind. So läßt sich z. B. die Unsicherheit über die richtige Beurteilung des datenerzeugenden Systems nicht in die Ableitung der a-priori-Wahrscheinlichkeiten einbeziehen. Insoweit erscheint es zweckmäßig, den Ansatz durch die Verwendung der in (2.50) dargestellten „*Dempster's* rule of combination" zu erweitern, zumal diese das Theorem von *Bayes* als Spezialfall einschließt[236]. Dies ist Gegenstand des nächsten Teils der vorliegenden Arbeit.

---

[235] Vgl. *Quick*, Risiken, 1996, S. 154.

[236] Siehe oben Kapitel B. II. 1. c) in diesem Teil der Arbeit.

DOI https://doi.org/10.3790/978-3-428-48909-1

Dritter Teil

Ermittlung des Gesamtrisikos

A. Risikomessung bei ergebnisorientierter Prüfung

Nachdem im vorigen Teil die einzelnen Unschärfemaße und ihre Eignung für die Risikomessung im vorliegenden Prüfungsmodell untersucht wurden, soll nun zunächst für die ergebnisorientierten Prüfungshandlungen gezeigt werden, durch welche der angeführten Risikomaße die jeweils auftretende Unschärfe im Rahmen der einzelnen Verfahren abgebildet werden kann und welche Anwendungsprobleme bei der Risikomessung auftreten. Dabei wird zunächst davon ausgegangen, daß das jeweilige Verfahren allein angewandt wird. Darüber hinaus soll unterstellt werden, daß die vom Prüfer getroffenen Einzelurteile keinerlei Unschärfe beinhalten, also D eine gewöhnliche und keine unscharfe Menge darstellt.

I. Zufallsstichproben

Die Anwendung von mathematisch-statistischen Schätz- und Testverfahren erlaubt es als einzige Prüfungsmethode objektive Wahrscheinlichkeiten zur Risikomessung zu verwenden, soweit die aus der Zufallsauswahl der Stichprobenelemente resultierende Unschärfe betrachtet wird. Dazu wird hier generell unterstellt, daß es sich bei der Auswahl der Stichprobenelemente um eine echte Zufallsauswahl handelt. Da die Schätz- und Testverfahren zwar auf den gleichen wahrscheinlichkeitstheoretischen Grundlagen basieren, sich jedoch durch die der Auswertung von Stichprobenergebnissen zugrundeliegende Fragestellung unterscheiden, soll zunächst getrennt für die Schätz- und Testverfahren untersucht werden, auf welche Art und Weise die Risikomessung in Abhängigkeit von der verwendeten Fragestellung möglich ist. Die folgenden Ausführungen beschränken sich dabei auf Verfahren, die einfache Mittelwertschätzer aus einfachen Zufallsstichproben verwenden, da die Verwendung gebundener Verfahren oder geschichteter Stichproben, sieht man von den bei der Bestimmung der relevanten Grenzverteilung auftretenden Probleme ab, ausschließlich die Effizienz der Verfahren, nicht die möglichen Aussagen bzw. grundlegenden Probleme der Risikomessung beeinflußt.

Darüber hinaus kann Unschärfe bei den erhaltenen Prüfungsergebnissen dadurch entstehen, daß eine oder mehrere der an die Anwendung der Verfahren im Rahmen der Prüfung geknüpften Voraussetzungen, namentlich die Kenntnis der Grenzverteilung für die Stichprobenergebnisse, die Homogenität der Prüfungsgesamtheit und das Vorliegen eines Häufigkeitsfalles[1], nicht oder zumindest nicht vollständig erfüllt werden. Dabei können Risiken, die aus der Ungewißheit über das Zutreffen der beiden zuletzt genannten Voraussetzungen resultieren, im Rahmen der Risikomessung bei der Anwendung von Zufallsstichproben nicht berücksichtigt werden. Es soll jedoch für Test- und Schätzverfahren gemeinsam geprüft werden, inwieweit sich die durch Unkenntnis über die tatsächliche Grenzverteilung der Stichprobenfunktion entstehende Unschärfe explizit im Modell berücksichtigen läßt.

1. Auswertung der Stichprobe

a) Schätzverfahren

Wie oben angeführt wurde[2], sind für die Prüfung ausschließlich Verfahren der Intervallschätzung relevant. Demnach ist, da die Beurteilung der Prüfungsgesamtheit im vorliegenden Prüfungsmodell anhand des Fehleranteils und der wertmäßigen Abweichung zwischen dem Ist- und dem Sollobjekt erfolgen soll, aus der vorliegenden Stichprobe ein Konfidenzintervall für den wahren Parameter θ bzw. $\Delta(X,Y)$ der Prüfungsgesamtheit zu bestimmen. Die Schätzverfahren sollen im folgenden weitestgehend am Beispiel der Schätzung des Fehleranteils in der Prüfungsgesamtheit dargestellt werden. Die Vorgehensweise und die auftretenden Probleme bei der Anwendung der Verfahren zur Beurteilung der wertmäßigen Abweichung sind die gleichen. Unterschiede ergeben sich erst bei der Bestimmung der Grenzverteilung der Stichprobenfunktion.

aa) Konzeption des Verfahrens

Ausgangspunkt der Überlegungen sei eine einfache Stichprobe vom Umfang n aus der Prüfungsgesamtheit X. Diese ist eine Realisation des Zufallsvektors $(Z_1,Z_2,\ldots,Z_n)$, wobei die einzelnen Elemente des Vektors ausschließlich den Wert Null oder Eins annehmen können. Mit z_i $(i=1,\ldots,n)$ wird hier die interessierende Merkmalsausprägung des i-ten Elements in der Stichprobe bezeichnet.

---

1 Siehe oben Kapitel B. I. 1. im Ersten Teil der Arbeit.

2 Siehe oben Kapitel B. I. 1. im Ersten Teil der Arbeit.

DOI https://doi.org/10.3790/978-3-428-48909-1

Ist x_i das zu z_i gehörende Element aus der Prüfungsgesamtheit X und y_i das zu x_i vom Prüfer abgeleitete Sollobjekt, so gilt:

$$z_i = \pi_D(\langle x_i, y_i \rangle). \tag{3.1}$$

Die Stichprobenfunktion für den Fehleranteilschätzer in der Prüfungsgesamtheit ist definiert durch[3]:

$$P = \frac{1}{n}\sum_{i=1}^{n} Z_i. \tag{3.2}$$

Aus der realisierten Stichprobe $z_1, z_2, \ldots, z_n$ erhält man als Schätzwert für den Fehleranteil:

$$p = \frac{1}{n}\sum_{i=1}^{n} z_i. \tag{3.3}$$

Ziel ist es demnach ein Verfahren anzugeben, das die Ableitung eines Zufallsintervalls $[0; P_o]$ aus der Stichprobenfunktion erlaubt, so daß die Wahrscheinlichkeit, ein Intervall $[0; p_o]$ zu erhalten, das θ überdeckt, mindestens $1-\alpha$ beträgt. Da im Rahmen der Prüfung die einzelnen Stichprobenelemente ohne Zurücklegen gezogen werden, sind die Ausprägungen der angegebenen Stichprobenfunktion hypergeometrisch verteilt. Für kleine θ und einen nicht zu großen Auswahlsatz n/N kann die hypergeometrische Verteilung durch die Poissonverteilung approximiert werden[4]. Für die Grenzverteilung der Stichprobenfunktion gilt demnach[5]:

$$F_P(x,\lambda) = \begin{cases} \sum_{j=0}^{x} \frac{e^{-\lambda}\lambda^j}{j!} & \text{für } x \geq 0 \text{ und } \lambda = n\theta, \\ 0 & \text{für } x < 0. \end{cases} \tag{3.4}$$

Dabei gibt $F_P(x,\lambda)$ die Wahrscheinlichkeit an, daß in einer Stichprobe vom Umfang n aus einer Prüfungsgesamtheit, deren wahrer Fehleranteil θ ist, höchstens x fehlerhafte Elemente auftreten. Ist der Fehleranteil $\theta = \theta'$ in der Prüfungsgesamtheit bekannt, dann läßt sich für Stichproben mit vorgegebenem Umfang n eine Obergrenze für die Anzahl der fehlerhaften Elemente in der Stichprobe so bestimmen, daß gilt:

---

3 Vgl. *Cochran*, Sampling, 1977, S. 52.

4 Dies ist regelmäßig der Fall, wenn $n \leq 0{,}05N$, $\theta \leq 0{,}1$, $n\theta \leq 10$ und $n \geq 50$ ist. Vgl. *Bamberg/Bauer*, Statistik, 1996, S. 328; *Mandl*, Anwendungsvoraussetzungen, 1984, S. 125.

5 In der folgenden Formel bezeichnet e nicht die Genauigkeit der Schätzung sondern die Euler'sche Zahl.

(3.5) $$pr\left(nP \le x_0 \middle| \theta = \theta'\right) = F_P\left(x_0, n\theta'\right) < \alpha .$$

Im Prüfungsmodell liegt allerdings die Situation vor, daß der wahre Fehleranteil θ unbekannt ist und eine Stichprobenrealisation mit bekannter Anzahl Fehler np ermittelt wurde. Für die vorliegende Stichprobe kann dann der wahre Fehleranteil p_0 in der Prüfungsgesamtheit bestimmt werden, für den gerade gelten soll:

(3.6) $$pr\left(nP \le np \middle| \theta = p_0\right) = F_P\left(np, np_0\right) < \alpha .$$

Demnach ist die Wahrscheinlichkeit für die gefundene oder eine geringere Fehleranzahl in der Stichprobe kleiner als α, wenn der wahre Fehleranteil in der Prüfungsgesamtheit höher als p_0 ist. Wird (3.6) zur Konstruktion des Konfidenzintervalls verwendet, dann beträgt die Wahrscheinlichkeit, ein Intervall zu erhalten, in dem der wahre Fehleranteil liegt, mindestens $1-\alpha$. D. h., das bei Anwendung der in (3.6) vorgeschlagenen Konstruktionsmethode resultierende Intervall erfüllt die Anforderungen, die an ein Konfidenzintervall für θ gestellt werden[6]. Aus (3.6) läßt sich die Intervallobergrenze zwar nicht analytisch, aber leicht numerisch bestimmen.

bb) Risikomessung

Fraglich ist, wie das ermittelte Konfidenzintervall zur Risikomessung verwendet werden kann. Dabei ist insbesondere zu untersuchen, ob aus dem Stichprobenergebnis die Ableitung einer Wahrscheinlichkeitsaussage über die Ordnungsmäßigkeit der Prüfungsgesamtheit möglich ist. Da gemäß (1.18) für wahre Fehleranteile, die größer als θ^* sind, die Ordnungsmäßigkeit der Prüfungsgesamtheit abgelehnt wird, könnte ein erster Ansatz sein, das Konfidenzniveau für das Intervall $\left[0;\theta^*\right]$ als Risikomaß zu verwenden[7]. Dieses läßt sich aus (3.6) direkt bestimmen mit:

(3.7) $$1-\alpha = 1 - F_P\left(np, n\theta^*\right).$$

Da es sich bei dem ermittelten Konfidenzniveau um ein Wahrscheinlichkeitsmaß handelt, kann die Bedeutung der ermittelten Größe für die Risikomessung ausschließlich dann untersucht werden, wenn der zugehörige Möglichkeitsraum Ω und der verwendete Ereigniskörper $\mathbb{K}$ bekannt sind. Der im Rahmen der hier diskutierten Stichprobenfunktion unterstellte Möglichkeitsraum

6 Vgl. *Cochran*, Sampling, 1977, S. 57.

7 Vgl. *Leffson*, Wirtschaftsprüfung, 1988, S. 176.

enthält als Elementarereignisse alle möglichen Stichproben vom vorgegebenen Umfang n. Da ein Modell ohne Zurücklegen der einmal in die Stichprobe gelangten Elemente unterstellt wird, ist die Anzahl der Elemente in Ω endlich. Für die Definition des Ereigniskörpers werden jeweils die Elementarereignisse aus Ω, die zur gleichen Merkmalsausprägung der Stichprobenfunktion führen, zusammengefaßt. Das Konfidenzniveau gibt die Wahrscheinlichkeit an, daß ein aus der Stichprobe nach der oben angegebenen Konstruktionsvorschrift abgeleitetes Konfidenzintervall den wahren interessierenden Parameter enthält. Es wird demnach eine Wahrscheinlichkeitsaussage über das Stichprobenergebnis, nicht über den wahren Parameter der Prüfungsgesamtheit gemacht. Dies folgt unmittelbar aus den Definitionen von Ω und $\mathbb{K}$, die gerade nicht auf mögliche Werte des interessierenden Parameters abstellen, sondern auf mögliche Stichprobenergebnisse für eine vorliegende Prüfungsgesamtheit. Dieser Prüfungsgesamtheit kann natürlich auch nur ein wahrer, wenn auch unbekannter Fehleranteil zugeordnet werden. Aus einem Konfidenzniveau von $1-\alpha$ für das Intervall $\lfloor 0;\theta^* \rfloor$ kann demnach keine Wahrscheinlichkeitsaussage über den wahren Wert von θ abgeleitet werden, da unterschiedliche wahre Werte für den Fehleranteil im Möglichkeitsraum nicht vorgesehen sind[8]. Insoweit gibt das Konfidenzniveau auch nicht die Wahrscheinlichkeit für die Ordnungsmäßigkeit der Prüfungsgesamtheit an.

Eine zweite im Schrifttum diskutierte Alternative, aufgrund des Stichprobenergebnisses eine Aussage über die Ordnungsmäßigkeit der Prüfungsgesamtheit zu treffen, wird darin gesehen, für einen genügend kleinen, vorgegebenen Wert α das Konfidenzintervall aus der Stichprobe zu bestimmen und die Ordnungsmäßigkeit der Prüfungsgesamtheit anzunehmen, wenn $p_0 < \theta^*$ gilt, bzw. die Ordnungsmäßigkeit abzulehnen, wenn θ^* im Konfidenzintervall enthalten ist[9]. Zwar ist die Aussage richtig, daß bei einem wahren Fehleranteil von mehr als θ^* in der Prüfungsgesamtheit die Wahrscheinlichkeit für eine korrekte Beurteilung der Prüfungsgesamtheit bei der gewählten Entscheidungsregel größer als $1-\alpha$ ist[10], da dies der Definition des Konfidenzintervalls entspricht. Diese Information allein trägt in der vorliegenden Prüfungssituation allerdings aus den folgenden Gründen nicht dazu bei, eine Entscheidung für oder gegen die Ordnungsmäßigkeit der Prüfungsgesamtheit zu stützen.

---

8 Vgl. *Fisz*, Wahrscheinlichkeitsrechnung, 1978, S. 571.

9 Vgl. *Roberts*, Statistical Auditing, 1978, S. 152f.; *Mandl*, Anwendungsvoraussetzungen, 1984, S. 128.

10 Vgl. *Mandl*, Auswahl, 1981, S. 179.

Betrachtet man zunächst den Fall, daß θ^* im Konfidenzintervall enthalten ist und nach der oben angeführten Regel die Ordnungsmäßigkeit der Prüfungsgesamtheit somit abzulehnen wäre. Der Grund für die Ablehnung der Ordnungsmäßigkeit kann nicht in dem Umstand gesehen werden, daß Fehleranteile im Konfidenzintervall enthalten sind, die, wenn sie tatsächlich vorliegen, zur Ablehnung der Ordnungsmäßigkeit führen würden. Aus dem gleichen Grund könnte dann behauptet werden, daß die Ordnungsmäßigkeit der Prüfungsgesamtheit anzunehmen wäre, da sich natürlich auch Fehleranteile im Konfidenzintervall befinden, die kleiner oder gleich θ^* sind und damit zur Annahme der Ordnungsmäßigkeit führen würden. Auch die Aussage, daß mit der angeführten Entscheidungsregel die Wahrscheinlichkeit, eine tatsächlich nicht ordnungsmäßige Prüfungsgesamtheit abzulehnen, größer als $1-\alpha$ ist, kann nicht als Begründung verwendet werden, da die Wahrscheinlichkeit, eine gerade noch ordnungsmäßige Prüfungsgesamtheit durch die Entscheidungsregel abzulehnen, genau $1-\alpha$ ist. Die Wahrscheinlichkeiten für richtige bzw. falsche Entscheidungen sind demnach fast gleich hoch.

Auch für den Fall, daß θ^* außerhalb des ermittelten Konfidenzintervalls liegt, kann die Entscheidung, die Ordnungsmäßigkeit der Prüfungsgesamtheit anzunehmen nicht allein damit begründet werden, daß die Wahrscheinlichkeit für ein Fehlurteil bei der angeführten Entscheidungsregel geringer als α ist[11]. Die Argumentation entspricht zum Teil der oben angeführten likelihood-Regel, die bei mathematisch-statistischen Testverfahren Verwendung findet[12]. Eine Übereinstimmung besteht darin, daß entweder die Tatsache, daß ein Ereignis, dessen Wahrscheinlichkeit geringer als α ist, beobachtet wurde, oder der Umstand, daß der Fehleranteil in der Prüfungsgesamtheit kleiner als θ^* ist, zutrifft. Die Tatsache, daß die Wahrscheinlichkeit für das Eintreten des vorliegenden Stichprobenergebnisses gering ist, wenn die Prüfungsgesamtheit nicht ordnungsmäßig ist, reicht allerdings nicht alleine dazu aus, die Hypothese, die Prüfungsgesamtheit ist nicht ordnungsmäßig, zu verwerfen[13]. Dies resultiert aus dem Umstand, daß auch bei ordnungsmäßiger Prüfungsgesamtheit für Werte, die nahe bei θ^* liegen, die Wahrscheinlichkeit für den Eintritt des beobachteten Stichprobenergebnisses nur wenig größer als α ist. Insoweit ist nicht ersichtlich, warum die Hypothese, die Prüfungsgesamtheit ist nicht ordnungsmäßig, der Hypothese, die Prüfungsgesamtheit ist ordnungsmäßig, vorgezogen werden

---

[11] Vgl. *Mandl*, Anwendungsvoraussetzungen, 1984, S. 128.

[12] Siehe oben Kapitel C. II. 2. im Zweiten Teil der Arbeit.

[13] A. A. für die Intervallschätzung im heterograden Fall *Obermeier*, Abschlussprüfung, 1983, S. 23.

sollte, obwohl sich die likelihood-Wahrscheinlichkeiten nur wenig voneinander unterscheiden. Eine Hypothese kann demnach generell nicht wegen einer geringen likelihood-Wahrscheinlichkeit der erhaltenen Stichprobe abgelehnt werden, solange keine Hypothese zur Verfügung steht, die das beobachtete Ereignis besser erklärt[14]. Da die angeführte Entscheidungsregel insgesamt nicht geeignet erscheint, in der vorliegenden Prüfungssituation zu einer begründeten Annahme oder Ablehnung der Ordnungsmäßigkeit zu gelangen, soll auf Überlegungen zur Risikomessung im Rahmen dieses Ansatzes verzichtet werden.

Wie die bisherigen Ausführungen deutlich gemacht haben, erscheint die Risikomessung mit Hilfe von Wahrscheinlichkeiten, soweit ausschließlich auf die Stichprobenergebnisse abgestellt wird, nicht möglich, da der zugrundeliegende Möglichkeitsraum dies nicht zuläßt. Wenn aus dem Stichprobenergebnis eine Wahrscheinlichkeitsaussage über den wahren Fehleranteil getroffen werden soll, dann gelingt dies erst, wenn der Möglichkeitsraum auf die Kombination von allen möglichen Fehleranteilen und den dabei jeweils möglichen Stichprobenergebnissen erweitert wird. Die Auswertung des Stichprobenergebnisses kann dann ausschließlich über die Anwendung des Theorems von *Bayes* zu einer Wahrscheinlichkeitsaussage über den wahren Wert des betrachteten Parameters und damit über die Ordnungsmäßigkeit der Prüfungsgesamtheit führen. Dabei treten allerdings die oben schon angesprochenen Probleme bei der Ermittlung der notwendigen a-priori-Wahrscheinlichkeiten auf[15].

An dieser Stelle soll wegen der bei der Verwendung des Theorems von *Bayes* auftretenden Probleme ein alternativer Ansatz zur Verwendung der Schätzergebnisse einer Zufallsstichprobe bei der Risikomessung aufgezeigt werden. Ausgangspunkt der Überlegungen ist dabei ein aus der vorliegenden Stichprobe ermitteltes Konfidenzintervall $[0;p_o]$, dessen Konfidenzniveau gemäß (3.6) ermittelt wird. Dabei soll die Obergrenze des Konfidenzintervalls so bestimmt werden, daß zumindest $p_o \leq \theta^*$ gilt. Auch wenn $1-\alpha$ nicht die Wahrscheinlichkeit für die Ordnungsmäßigkeit der Prüfungsgesamtheit ausdrückt, so wird doch mit zunehmendem Konfidenzniveau die Überzeugung des Prüfers, daß die Prüfungsgesamtheit tatsächlich ordnungsmäßig ist, zunehmen. Für den Grenzfall, daß durch einen genügend großen Stichprobenumfang α für das angeführte Konfidenzintervall den Wert Null annimmt, müßte sich der Prüfer sicher sein, daß die Prüfungsgesamtheit ordnungsmäßig im Hinblick auf das Kriterium Fehleranteil ist. Der angeführte Zusammenhang zwischen Konfi-

14 Vgl. dazu auch *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 145-150.

15 Siehe oben Kapitel B. I. 2. a) und Kapitel C. II. 3. im Zweiten Teil der Arbeit.

denzniveau und Überzeugungsgrad läßt sich durch die Verwendung eines Glaubwürdigkeitsmaßes zur Risikomessung abbilden. Dazu wird der Überzeugungsgrad durch ein Glaubwürdigkeitsmaß gemessen und numerisch mit dem Konfidenzniveau gleichgesetzt. Der Umstand, daß eine Wahrscheinlichkeit α dafür besteht, daß die Aussage „Der wahre Fehleranteil liegt im Konfidenzintervall" falsch ist, beeinflußt die Glaubwürdigkeitseinschätzung der Nichtordnungsmäßigkeit der Prüfungsgesamtheit nicht[16].

Der formale Apparat zur Ableitung der grundlegenden Wahrscheinlichkeitszuordnung für einen frame of discernment Θ aus einem zweiten frame of discernment Θ', über dem, wie im vorliegenden Fall, ein Wahrscheinlichkeitsmaß definiert ist, wurde bereits in (2.67) und (2.68) vorgestellt. Für die hier bestehende Prüfungssituation wird bei Verwendung der dort angegebenen Symbole $\Theta' = \Omega$ gesetzt. Dabei bezeichnet Ω den Möglichkeitsraum, der oben für die Stichprobenfunktion definiert wurde. Wird das Intervall $[0;\theta^*]$ mit E bezeichnet, dann gilt $\Theta' = \{E;\overline{E}\}$. Gemäß (3.7) ist auf Θ' ein Wahrscheinlichkeitsmaß definiert. Für die Wahrscheinlichkeitszuordnungen erhält man $pr(E) = 1 - \alpha$ und $pr(\overline{E}) = \alpha$. Der im Rahmen der Risikomessung interessierende frame of discernment Θ soll ausschließlich O und $\overline{O}$ enthalten[17]. Bevor die grundlegende Wahrscheinlichkeitszuordnung abgeleitet werden kann, muß noch die in (2.67) angeführte Kompatibilitätsrelation angegeben werden.

Soll aus dem Stichprobenergebnis ein Glaubwürdigkeitsgrad für die Ordnungsmäßigkeit der Prüfungsgesamtheit gewonnen werden, so muß wegen der oben angeführten Argumentation zum Zusammenhang zwischen Überzeugungsgrad im Hinblick auf die Ordnungsmäßigkeit und Konfidenzniveau gelten $\Gamma_O(E) = O$. Für $\overline{E}$ gilt, daß es sowohl die Fälle, in denen keine Ordnungsmäßigkeit der Prüfungsgesamtheit gegeben ist, als auch die Fälle, in denen das Konfidenzintervall den wahren Fehleranteil in der Prüfungsgesamtheit nicht überdeckt, repräsentiert. Insoweit kann es nicht einem der beiden Ereignisse O oder $\overline{O}$ ausschließlich zugeordnet werden. Da das Eintreten von $\overline{E}$ auch nicht im Widerspruch zu Θ steht, dies würde zu einer Zuweisung auf $\varnothing$ führen, muß gelten $\Gamma_O(\overline{E}) = \Theta$. Als grundlegende Wahrscheinlichkeitszuordnung erhält man gemäß (2.68):

(3.8) $$m(\varnothing) = 0, m(O) = 1 - \alpha, m(\overline{O}) = 0 \text{ und } m(\Theta) = \alpha.$$

---

[16] Vgl. dazu auch *Srivastava/Shafer*, Belief-Function, 1992, S. 497f.

[17] Es wird insoweit an dieser Stelle unterstellt, daß über die Ordnungsmäßigkeit ausschließlich anhand des Fehleranteils entschieden wird. Eine Erweiterung auf die beiden Kriterien erfolgt später in diesem Kapitel.

Der Grad der Glaubwürdigkeit für die Ordnungsmäßigkeit der Prüfungsgesamtheit beträgt demnach $bel(O)=1-\alpha$, der Grad der Glaubwürdigkeit für die Nichtordnungsmäßigkeit der Prüfungsgesamtheit beträgt $bel(\overline{O})=0$. Dabei kann aus der Glaubwürdigkeitszuordnung von Null für die Nichtordnungsmäßigkeit der Prüfungsgesamtheit nicht gefolgert werden, daß der Fall der tatsächlichen Nichtordnungsmäßigkeit der Prüfungsgesamtheit ausgeschlossen wird. Der Umstand, daß die Glaubwürdigkeitszuordnung für $\overline{O}$ Null ist, drückt ausschließlich das Fehlen von Erkenntnissen für die Beurteilung der Glaubwürdigkeit aus. Für die Risikomessung ergibt sich bei Wahl der Handlungsalternative a_1 gemäß (2.42) $pl(s_2)=pl(\overline{O})=\alpha$, bei der Wahl von a_2 erhält man $pl(s_1)=pl(O)=1$. Insoweit stimmt das Ergebnis zwar teilweise mit dem zuvor angeführten Ansatz überein, allerdings besteht der Unterschied zum einen darin, daß die Begründung für die Vorgehensweise, namentlich die Übertragung von Evidenz, offengelegt wird. Zum anderen ist das verwendete Unschärfemaß kein Wahrscheinlichkeitsmaß.

Die bisher angeführte Argumentation zur Auswertung der Stichprobe im Hinblick auf die Glaubwürdigkeitsbeurteilung der Ordnungsmäßigkeit kann auch auf die Beurteilung der Nichtordnungsmäßigkeit der Prüfungsgesamtheit übertragen werden. Das Konfidenzniveau für das Intervall $(\theta^*;1]$ muß wegen (2.6) α betragen. Die Bedeutung von E und $\overline{E}$ soll der oben angeführten Aufteilung von Θ' entsprechen. Für die neue Kompatibilitätsrelation muß demnach gelten $\Gamma_{\overline{O}}(\overline{E})=\overline{O}$ und $\Gamma_{\overline{O}}(E)=\Theta$. Als grundlegende Wahrscheinlichkeitszuordnung erhält man dann:

(3.9) $$m(\varnothing)=0, m(O)=0, m(\overline{O})=\alpha \text{ und } m(\Theta)=1-\alpha.$$

Der Grad der Glaubwürdigkeit für die Ordnungsmäßigkeit der Prüfungsgesamtheit beträgt dabei $bel(O)=0$, der Grad der Glaubwürdigkeit für die Nichtordnungsmäßigkeit der Prüfungsgesamtheit beträgt $bel(\overline{O})=\alpha$. Für die Risikomessung ergibt sich entsprechend bei der Wahl von a_1 $pl(s_2)=pl(\overline{O})=1$, bei der Wahl von a_2 $pl(s_1)=pl(O)=1-\alpha$. Allerdings kann aus einer einzelnen Stichprobe nur eine der beiden Glaubwürdigkeitseinschätzungen verwendet werden. Eine Kombination der beiden Glaubwürdigkeitseinschätzungen nach „*Dempster's* rule of combination“ (2.50) würde voraussetzen, daß diese voneinander unabhängig sind. Diese Voraussetzung ist jedoch nicht erfüllt, da beide Glaubwürdigkeitseinschätzungen aus einer Stichprobe stammen. Welche der beiden Glaubwürdigkeitszuordnungen verwendet wird, liegt grundsätzlich im Ermessen des Prüfers. Allerdings erscheint es vorteilhafter, diejenige zu verwenden, deren Beitrag zur Einschätzung der Glaubwürdigkeit von O oder $\overline{O}$

am größten ist. Es empfiehlt sich daher, für den Fall, daß $1-\alpha$ größer ist als α, das Stichprobenergebnis im Hinblick auf die Ordnungsmäßigkeit der Prüfungsgesamtheit auszuwerten, und für den Fall, daß $1-\alpha$ kleiner ist als α, die Auswertung im Hinblick auf die Nichtordnungsmäßigkeit der Prüfungsgesamtheit durchzuführen.

Probleme ergeben sich bei der Verwendung dieses Ansatzes, wenn im Rahmen der Prüfungsplanung der notwendige Stichprobenumfang bestimmt werden soll. Dies erfordert zum einen die Angabe der gewünschten Sicherheit und Genauigkeit der Schätzung. Zum anderen muß die Varianz der Stichprobenfunktion geeignet geschätzt werden. Die Zusammenhänge lassen sich am einfachsten deutlich machen, wenn unterstellt wird, daß die Stichprobenfunktion annähernd normalverteilt ist[18]. Bei unterstellter Normalverteilung der Stichprobenfunktion erhält man das Konfidenzintervall für den Fehleranteil in der Prüfungsgesamtheit nach folgender Formel:

$$(3.10) \qquad [0; p + z(1-\alpha)\sigma_P].$$

Dabei bezeichnet $z(1-\alpha)$ die Umkehrfunktion der Verteilungsfunktion für die Standardnormalverteilung. Die Standardabweichung der Stichprobenfunktion σ_P ist für die Fehleranteilschätzung und das hier unterstellte Modell ohne Zurücklegen gegeben durch:

$$(3.11) \qquad \sigma_P = \sqrt{\frac{\theta(1-\theta)}{n}\frac{N-n}{N-1}}.$$

Die Genauigkeit der Schätzung wird durch den Stichprobenfehler e gemäß (1.30) folgendermaßen bestimmt:

$$(3.12) \qquad e = p_o - p = z(1-\alpha)\sigma_P.$$

Durch Einsetzen von (3.11) in (3.12) und der Annahme, daß $N \approx N-1$ ist, erhält man den zur Einhaltung der geforderten Genauigkeit und Sicherheit notwendigen Stichprobenumfang n mit:

$$(3.13) \qquad n \geq \frac{N\theta(1-\theta)z(1-\alpha)^2}{Ne^2 + \theta(1-\theta)z(1-\alpha)^2}.$$

[18] Der notwendige Stichprobenumfang kann auch bestimmt werden, wenn die Binomial- oder die Poissonverteilung als Approximation für die hypergeometrische Verteilung gewählt werden. Allerdings ist dabei keine analytische Bestimmung möglich, so daß die hier zu untersuchenden Zusammenhänge nicht so offen zu Tage treten. Zur Ableitung des Stichprobenumfangs bei den angeführten Verteilungen vgl. *Reuter*, Prüfungsumfang, 1975, S. 107-119.

Wie in (3.13) zu erkennen ist, erfordert eine genaue Bestimmung des notwendigen Stichprobenumfangs zur Fehleranteilschätzung die Kenntnis des wahren Parameters θ. Darüber hinaus müssen das Konfidenzniveau und der maximale Stichprobenfehler vorgegeben werden. Für eine risikoorientierte Prüfungsplanung soll zunächst unterstellt werden, daß die für eine ausreichend sichere Beurteilung fehlende Glaubwürdigkeit, die gleichzeitig das notwendige Konfidenzniveau determiniert, grundsätzlich aus den vorliegenden Erkenntnissen abgeleitet werden kann. Für die Abschätzung des zur Bestimmung des Stichprobenumfangs benötigten Fehleranteils in der Prüfungsgesamtheit kommen unterschiedliche Vorgehensweisen in Frage, namentlich die Verwendung der Daten einer Vorstichprobe, Erfahrungswerte des Prüfers oder eine pessimistische Abschätzung mit $\theta = 0{,}5$, da die Varianz der Fehler in der Prüfungsgesamtheit dann maximal ist[19]. Dieses Problem ist somit zumindest prinzipiell lösbar. Schwierigkeiten bereitet dagegen beim hier vorgestellten Ansatz zur Auswertung der Stichprobe die Festlegung des maximal zulässigen Schätzfehlers e. Dieser ergibt sich gemäß (3.12) als Differenz zwischen der Konfidenzintervallobergrenze p_o und dem Schätzwert, der aus der Stichprobe abgeleitet wird. Da für die Auswertung der Stichprobe im Hinblick auf die Ordnungsmäßigkeit zumindest $p_o \leq \theta^*$ gelten soll, ist die erforderliche Genauigkeit der Schätzung vom realisierten Wert der Stichprobenfunktion abhängig, also eine Zufallsvariable. Insoweit kann eine endgültige Entscheidung darüber, ob ein gewählter Stichprobenumfang ausreichend ist, erst nach der Auswertung der Stichprobe getroffen werden. Die sich bei der Anwendung des Ansatzes zur Beurteilung der Prüfungsgesamtheit ergebende Vorgehensweise entspricht somit weitgehend der Konzeption bei sequentiellen Hypothesentests.

b) Testverfahren

Wie schon oben dargestellt wurde, beinhalten die mathematisch-statistischen Testverfahren eine Entscheidungsregel darüber, welche der alternativen Hypothesen über den interessierenden Parameter der Prüfungsgesamtheit abzulehnen ist bzw. welche angenommen[20] werden soll. Dabei ist das Risiko, nämlich die Wahrscheinlichkeit eine Fehlentscheidung (Irrtumswahrscheinlichkeit) zu treffen, vorzugeben. Die der Entscheidungsregel zugrundeliegende Argumentation und die Probleme, die auftreten, wenn andere Prüfungsinformationen bei der

---

[19] Vgl. *Mandl*, Anwendungsvoraussetzungen, 1984, S. 126f.

[20] Zum Unterschied zwischen Annahme einer Hypothese und Wahrheit der darin gemachten Aussage siehe oben Kapitel B. I. 1. im Ersten Teil der Arbeit.

Entscheidung berücksichtigt werden sollen, wurden schon dargestellt[21], diese sollen hier nicht wiederholt werden. Ungeklärt ist bisher insbesondere, wie die Formulierung der verwendeten Hypothesen erfolgen soll. Dabei soll die Argumentation wiederum beispielhaft für den Fehleranteil in der Prüfungsgesamtheit geführt werden. Ebenso wie bei den Schätzverfahren gilt, daß die Vorgehensweise sowie die auftretenden Probleme bei beiden Parametern der Prüfungsgesamtheit identisch sind.

aa) Konzeption des Verfahrens

Für die Untersuchung der den Testverfahren zugrundeliegenden Konzeption wird zunächst davon ausgegangen, daß die relevanten Werte für die Nullhypothese (H_0) und die Alternativhypothese (H_1) vorgegeben sind. Es soll gelten:

$$\begin{matrix} H_0 : \theta \le \theta_0 \\ H_1 : \theta > \theta_1 \end{matrix} \text{ mit } \theta_0 \le \theta_1 \tag{3.14}$$[22].

Die Ablehnung der Nullhypothese erfolgt bei der den Testverfahren zugrundeliegenden Entscheidungsregel, wenn die Wahrscheinlichkeit für das Eintreten des Stichprobenergebnisses, falls die Nullhypothese richtig ist, geringer als α ist und gleichzeitig die Eintrittswahrscheinlichkeit für das Stichprobenergebnis, falls die Alternativhypothese richtig ist, mindestens $1-\beta$ beträgt. Umgekehrt erfolgt die Ablehnung der Alternativhypothese und damit die Annahme der Nullhypothese, wenn die Eintrittswahrscheinlichkeit für das Stichprobenergebnis bei Gültigkeit der Alternativhypothese geringer ist als β und gleichzeitig die Eintrittswahrscheinlichkeit bei Gültigkeit der Nullhypothese größer als $1-\alpha$ ist. Es existieren demnach zwei Ereignismengen $A_0, A_1 \subset \Omega$, für die gilt:

$$A_0 = \{p | p \in \Omega \wedge pr(P > p_0 | \theta = \theta_0) = \alpha \wedge p > p_0\}, \tag{3.15}$$

$$A_1 = \{p | p \in \Omega \wedge pr(P \le p_1 | \theta = \theta_1) = \beta \wedge p \le p_1\}. \tag{3.16}$$

Die Menge A_0 wird als Ablehnungsbereich der Nullhypothese, die Menge A_1 als Ablehnungsbereich der Alternativhypothese bezeichnet. Die Annahme der Nullhypothese gegen die Alternativhypothese erfolgt, wenn ein Stichprobenergebnis $p \in \overline{A}_0 \cap A_1$ realisiert wird. Entsprechend wird die Nullhypothese abgelehnt und gleichzeitig die Alternativhypothese angenommen, wenn ein Stichprobenergebnis $p \in \overline{A}_1 \cap A_0$ aus der Stichprobe ermittelt wird. Der Test entschei-

---

[21] Siehe oben Kapitel C. II. 2. im Zweiten Teil der Arbeit.

[22] Für den Fall, daß $\theta_0 = \theta_1$ gilt, handelt es sich bei dem Test um einen Signifikanztest, sonst um einen Test mit konkretisierter Gegenhypothese.

DOI https://doi.org/10.3790/978-3-428-48909-1

det für jedes mögliche Stichprobenergebnis, wenn $A_0 \cup A_1 = \Omega$ gilt. Ein solcher Test wird auch bester Test genannt[23].

Zur Analyse, unter welchen Voraussetzungen ein bester Test vorliegt, soll wiederum ausschließlich wegen der einfacheren formalen Handhabung unterstellt werden, daß die Stichprobenfunktion normalverteilt ist. Die Ablehnungsbereiche in (3.15) und (3.16) für die beiden Hypothesen sind dann bestimmt durch[24]:

(3.17) $$p_0 = \theta_0 + z(1-\alpha)\sqrt{\frac{\theta_0(1-\theta_0)}{n}\frac{N-n}{N-1}} \text{ und}$$

(3.18) $$p_1 = \theta_1 - z(1-\beta)\sqrt{\frac{\theta_1(1-\theta_1)}{n}\frac{N-n}{N-1}}.$$

Falls ein bester Test vorliegt muß ein Wert p^* im Stichprobenraum existieren, für den gilt:

(3.19) $$p^* = p_0 = p_1 \text{ und}$$

(3.20) $$pr\left(P > p^* \middle| \theta = \theta_0\right) = \alpha \wedge pr\left(P > p^* \middle| \theta = \theta_1\right) = 1 - \beta \text{ bzw.}$$

(3.21) $$pr\left(P \leq p^* \middle| \theta = \theta_0\right) = 1 - \alpha \wedge pr\left(P \leq p^* \middle| \theta = \theta_1\right) = \beta.$$

Treten Stichprobenergebnisse auf, die größer als p^* sind, wird die Nullhypothese abgelehnt. Dabei impliziert die Ablehnung der Nullhypothese gleichzeitig die Annahme der Alternativhypothese. Für Stichprobenergebnisse, die nicht größer als p^* sind, erfolgt die Annahme der Nullhypothese bei gleichzeitiger Ablehnung der Alternativhypothese. Die Zusammenhänge werden in Abb. 3.1 nochmals graphisch dargestellt. Da alle Größen bis auf n vorgegeben sind, liegt ein bester Test immer dann vor, wenn der Stichprobenumfang ausreichend groß gewählt wird. Für den notwendigen Stichprobenumfang erhält man durch Zusammenfassen der Gleichungen (3.17), (3.18) und (3.19) wiederum unter der Annahme, daß $N \approx N-1$ ist:

(3.22) $$n \geq \frac{N\left(z(1-\alpha)\sqrt{\theta_0(1-\theta_0)} + z(1-\beta)\sqrt{\theta_1(1-\theta_1)}\right)^2}{N(\theta_1-\theta_0)^2 + \left(z(1-\alpha)\sqrt{\theta_0(1-\theta_0)} + z(1-\beta)\sqrt{\theta_1(1-\theta_1)}\right)^2}.$$

Wie in (3.22) zu erkennen ist, muß der Abstand zwischen den beiden Hypothesen größer als Null sein, damit sich ein Stichprobenumfang, der geringer

23 Vgl. *Fisz*, Wahrscheinlichkeitsrechnung, 1978, S. 640f.

24 Vgl. auch *Schildbach*, Stichprobenprüfung, 1992, Sp. 1850.

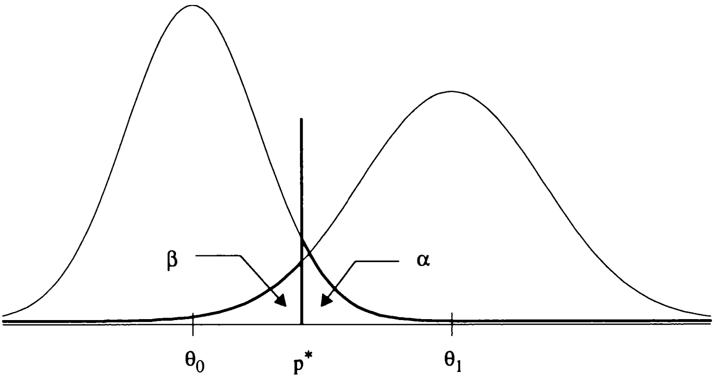

Abb. 3.1: Bester Fehleranteilstest bei unterstellter Normalverteilung der Stichprobenfunktion

als der Umfang der betrachteten Prüfungsgesamtheit ist, ergibt. Ist der Umfang einer vorliegenden Stichprobe kleiner als der aus (3.22) bestimmte notwendige Stichprobenumfang, dann gilt:

(3.23) $$\theta_0 + z(1-\alpha)\sqrt{\frac{\theta_0(1-\theta_0)}{n}\frac{N-n}{N-1}} > \theta_1 - z(1-\beta)\sqrt{\frac{\theta_1(1-\theta_1)}{n}\frac{N-n}{N-1}}.$$

Für diesen Fall sind Stichprobenergebnisse möglich, die in $\overline{A}_1$ aber nicht in A_0 enthalten sind. Tritt eines dieser Stichprobenergebnisse ein, kann keine Entscheidung über die Hypothesen mit den vorgegebenen Irrtumswahrscheinlichkeiten α und β erfolgen. Graphisch wird diese Situation in Abb. 3.2 dargestellt. Zur Beantwortung der Frage, welche Möglichkeiten in einer solchen Indifferenzsituation zur Entscheidungsfindung in Betracht kommen und welche Konsequenzen sich daraus für die Risikomessung ergeben, muß die Bedeutung der für die Nullhypothese und die Alternativhypothese gewählten Werte in der vorliegenden Prüfungssituation geklärt werden.

bb) Bestimmung der Hypothesen

Zur Bestimmung der konkreten Werte für die Hypothesen ist zunächst zu prüfen, in welcher der beiden Hypothesen der kritische Fehleranteil θ^* verwendet werden muß. Die Tatsache, daß die Prüfungsgesamtheit dann als ordnungsmäßig beurteilt werden soll, wenn $\theta \leq \theta^*$ gilt, könnte dafür sprechen, den kritischen Fehleranteil θ^* als Wert für die Nullhypothese zu verwenden. Um zu zeigen, warum diese Vorgehensweise der unterstellten Prüfungssituation nicht

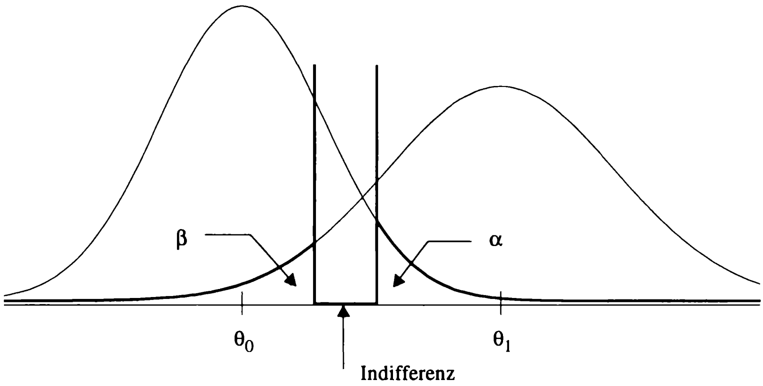

Abb. 3.2: Fehleranteilstest mit Indifferenzbereich

gerecht wird, muß zunächst der Begriff der Operationscharakteristik eines Tests eingeführt werden.

Da grundsätzlich davon ausgegangen werden kann, daß der tatsächliche Fehleranteil in der Prüfungsgesamtheit nicht auf die in den Hypothesen angeführten Wertebereiche beschränkt ist, wird ein bestimmter Test insbesondere durch die Wahrscheinlichkeit charakterisiert, mit der die Nullhypothese angenommen wird, wenn ein bestimmter wahrer Fehleranteil in der Prüfungsgesamtheit als bekannt vorausgesetzt wird. Eine Funktion, die diesen Zusammenhang zwischen alternativen wahren Parametern der betrachteten Gesamtheit und der Wahrscheinlichkeit für die Annahme der Nullhypothese herstellt, wird Operationscharakteristik des Tests genannt[25]. Falls die Stichprobenfunktion normalverteilt ist und ein bester Fehleranteilstest vorliegt, ermittelt man die Operationscharakteristik $L(\theta)$ formal nach:

$$(3.24) \qquad L(\theta)=\begin{cases}\Phi\left(\dfrac{p^*-\theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}}\right), & \text{wenn}\,\theta>0,\\ 1, & \text{wenn}\,\theta=0.\end{cases}$$

Dabei gibt $\Phi(x)$ in (3.24) den Wert der Standardnormalverteilung an der Stelle x an. Nach (3.17), (3.18) und (3.19) sind p^* und n in (3.24) eindeutig bestimmt, wenn die Testparameter α, β, θ_0 und θ_1 festgelegt sind, so daß die Operationscharakteristik ausschließlich vom wahren Fehleranteil θ in der Prüfungs-

[25] Vgl. *Fisz*, Wahrscheinlichkeitsrechnung, 1978, S. 631.

gesamtheit abhängt. In Abb. 3.3 ist beispielhaft der Verlauf der Operationscharakteristik für einen besten Fehleranteilstest bei unterstellter Normalverteilung der Stichprobenfunktion dargestellt. Wie dabei zu erkennen ist, treten Probleme auf, falls der kritische Fehleranteil θ^* zur Formulierung der Nullhypothese verwendet wird und nicht ordnungsmäßige Prüfungsgesamtheiten mit einem Fehleranteil nahe bei θ^* existieren. In diesem Fall beträgt die Wahrscheinlichkeit für die Annahme der Nullhypothese annähernd $1-\alpha$. Insoweit würden auch tatsächlich nicht ordnungsmäßige Prüfungsgesamtheiten mit hoher Wahrscheinlichkeit als ordnungsmäßig angenommen. Diese Vorgehensweise für die Formulierung der Nullhypothese scheidet demnach aus.

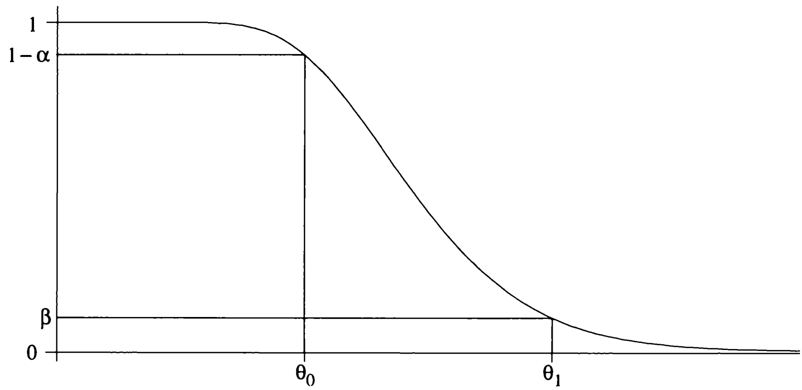

Abb. 3.3: Verlauf der Operationscharakteristik eines besten Fehleranteilstests bei unterstellter Normalverteilung der Stichprobenfunktion

Wie ebenfalls in Abb. 3.3 zu erkennen ist, erfolgt die Annahme der Nullhypothese für Fehleranteile, die größer als θ_1 sind, höchstens mit der Wahrscheinlichkeit β. Wird demnach der kritische Fehleranteil θ^* zur Formulierung der Alternativhypothese herangezogen, erfolgt die Annahme einer tatsächlich nicht ordnungsmäßigen Prüfungsgesamtheit als ordnungsmäßig höchstens mit der durch β vorgegebenen Wahrscheinlichkeit. Insoweit erscheint die Wahl von $\theta_1 = \theta^*$ der vorliegenden Prüfungssituation eher angemessen. Fraglich bleibt, welcher Wert für die Nullhypothese zu verwenden ist. Die Wahl von $\theta_0 = \theta_1 = \theta^*$ führt dazu, daß gemäß (3.22) für einen besten Test ein Stichprobenumfang in Höhe von N benötigt wird. Insoweit könnte nicht mehr von einem Stichprobenverfahren gesprochen werden. Stichprobenumfänge, die geringer sind als der notwendige Stichprobenumfang, führen im angeführten Fall sowie bei Werten für θ_0, die dicht bei θ_1 liegen, zu einem großen Indifferenzbereich

DOI https://doi.org/10.3790/978-3-428-48909-1

für den Test[26]. Insoweit sollten die Hypothesen ausreichend unterscheidbar sein. Auch die Verwendung von $\theta_0 = 0$, womit in der Nullhypothese das Vorliegen einer im Hinblick auf den Fehleranteil optimalen Prüfungsgesamtheit formuliert würde[27], ist nicht zweckmäßig, da zum einen für die Ordnungsmäßigkeit nicht gefordert wird, daß die Prüfungsgesamtheit fehlerfrei ist, zum anderen müßte die Nullhypothese schon dann ablehnt werden, wenn ein einziger Fehler in der Stichprobe gefunden wird.

Fraglich ist somit, ob die zu untersuchende Fragestellung im Rahmen der Prüfung grundsätzlich mit dem den mathematisch-statistischen Testverfahren zugrundeliegenden Konzept übereinstimmt. Dieses verlangt, daß eine Entscheidung zwischen zwei sich ausschließenden, ausreichend unterscheidbaren Hypothesen, von denen genau eine wahr ist, zu treffen ist. Im vorliegenden Prüfungsmodell wird jedoch eine eindeutige Aussage[28] über die Ordnungsmäßigkeit der Prüfungsgesamtheit gefordert. D. h., es sind die beiden Hypothesen

(3.25)
$$\begin{aligned} H_0 &: \theta \leq \theta^* \\ H_1 &: \theta > \theta^* \end{aligned}$$

zu überprüfen. Diese schließen sich zwar gegenseitig aus und es wird unterstellt, daß eine der beiden Hypothesen wahr ist, allerdings sind sie, wie oben schon ausgeführt wurde, nicht ausreichend unterscheidbar, um sie mit genügend geringer Irrtumswahrscheinlichkeit aufgrund einer Stichprobe zu testen. Inwieweit dieser Umstand, daß sich die Konzeption der Testverfahren und die Fragestellung im Hinblick auf das Gesamturteil im vorliegenden Prüfungsmodell nicht exakt entsprechen, bei der Risikomessung berücksichtigt werden kann, wird später untersucht[29]. Vorher soll geprüft werden, ob das den Testverfahren zugrundeliegende Konzept in anderen Bereichen der Prüfung, namentlich bei der Beurteilung von Teilen der Prüfungsgesamtheit (Teilgesamtheiten), verwendet werden kann.

Für das hier vorliegende Prüfungsmodell soll unterstellt werden, daß die Vorgehensweise bei der Beurteilung von Teilgesamtheiten grundsätzlich der Vorgehensweise, wie sie oben für die Beurteilung der Prüfungsgesamtheit be-

---

[26] Vgl. *Schildbach*, Stichprobenprüfung, 1992, Sp. 1849.

[27] Dies würde der Vorgehensweise im Schrifttum im Rahmen der dort verwendeten Ansätze für mathematisch-statistische Testverfahren über das Fehlerausmaß entsprechen. Vgl. z. B. *Sperl*, Prüfungsplanung, 1978, S. 186f.; *Obermeier*, Abschlussprüfung, 1983, S. 101f.

[28] Die Konsequenzen, die sich ergeben, wenn diese Voraussetzung nicht erfüllt wird, werden weiter unten untersucht.

[29] Siehe unten Kapitel A. I. 1. b) cc) in diesem Teil der Arbeit.

schrieben wurde[30], entspricht. Da die eindeutige Aussage des Prüfers zunächst ausschließlich für die Prüfungsgesamtheit gefordert wird, ist es denkbar, daß bei der Beurteilung von Teilgesamtheiten ein Bereich für den Fehleranteil existiert, für den die Beurteilung im Hinblick auf die Ordnungsmäßigkeit nicht eindeutig erfolgen kann. Liegt ein solcher Fall vor, könnte zur Festlegung der Nullhypothese der Fehleranteil gewählt werden, bei dem die Ordnungsmäßigkeit uneingeschränkt angenommen wird. Entsprechend ist zur Bestimmung der Alternativhypothese ein Fehleranteil anzugeben, ab dem die uneingeschränkte Ablehnung der Ordnungsmäßigkeit der betrachteten Teilgesamtheit erfolgt[31]. Wie an der Operationscharakteristik in Abb. 3.3 zu erkennen ist, wird durch das Testverfahren jedoch auch für wahre Fehleranteile, die zwischen den beiden Hypothesen liegen, eine Entscheidung für eine der beiden Hypothesen getroffen. Die Unschärfe in der Beurteilung, die in diesen Fällen an und für sich bestünde, wenn der wahre Fehleranteil bekannt wäre, kommt insoweit nicht zum Ausdruck. Daher sind Testverfahren für Fragestellungen, in denen Unschärfebereiche für die Beurteilung bestehen, nicht geeignet.

cc) Risikomessung

Unter der Voraussetzung, daß die betrachtete Fragestellung in ihrer Struktur mit der Konzeption der mathematisch-statistischen Testverfahren übereinstimmt, kann die Messung des Risikos der durch den Test getroffenen Entscheidung einmal durch die jeweilige Irrtumswahrscheinlichkeit erfolgen. Bei der Annahme der Nullhypothese wäre dies β, bei der Annahme der Alternativhypothese α. Die Bestimmung der konkreten Irrtumswahrscheinlichkeiten für eine vorliegende Stichprobe kann erfolgen, indem in (3.15) und (3.16) p_0 und p_1 jeweils durch den ermittelten Wert p der Stichprobenfunktion ersetzt werden. Die Irrtumswahrscheinlichkeiten repräsentieren dabei nicht die Wahrscheinlichkeit für die jeweils abgelehnte Hypothese, sondern die likelihood des Stichprobenergebnisses bei Gültigkeit der abgelehnten Hypothese. Wie bei den Schätzverfahren ist somit eine direkte Messung der Wahrscheinlichkeiten für O bzw. $\overline{O}$ aus dem Testergebnis nicht möglich. Dieses Ergebnis folgt auch zwingend aus dem Umstand, daß beiden Verfahren der gleiche Möglichkeitsraum und Ereigniskörper zugrundeliegt.

---

30 Siehe oben Kapitel A. im Ersten Teil der Arbeit.

31 Vgl. *Schulte*, Methoden, 1970, S. 118; *Mandl*, Anwendungsvoraussetzungen, 1984, S. 133; *Leffson*, Wirtschaftsprüfung, 1988, S. 187f.

DOI https://doi.org/10.3790/978-3-428-48909-1

Wie oben angeführt wurde, ist die Verwerfung einer Hypothese jedoch ausschließlich dann ausreichend begründet, wenn eine Alternativhypothese zur Verfügung steht, die das Stichprobenergebnis besser erklärt. Dieser Umstand kommt bei der alleinigen Verwendung der Irrtumswahrscheinlichkeiten als Risikomaß nicht zum Ausdruck. Ein für den Test aussagekräftigeres Risikomaß ergibt sich dann, wenn das Verhältnis der likelihoods des realisierten Stichprobenergebnisses verwendet wird. Liegt ein bester Test vor, und wird ein Stichprobenergebnis aus A_0 realisiert, dann gilt wegen (3.15) und (3.16):

(3.26) $$pr(A_0|\theta \leq \theta_0) \leq \alpha \text{ und } pr(\overline{A}_1|\theta > \theta_1) = pr(A_0|\theta > \theta_1) > 1-\beta.$$

Als Risikomaß γ erhält man dann[32]:

(3.27) $$\gamma = \frac{pr(A_0|\theta \leq \theta_0)}{pr(A_0|\theta > \theta_1)} \leq \frac{\alpha}{1-\beta}.$$

Bei dem in (3.27) definierten Risikomaß handelt es sich um kein Unschärfemaß im oben angeführten Sinn, es ist insbesondere nicht automatisch auf das Intervall $[0;1]$ beschränkt. Wird allerdings immer diejenige Hypothese abgelehnt, bei der die likelihood für das beobachtete Stichprobenergebnis am geringsten ist, dann nimmt γ ausschließlich Werte zwischen Null und Eins an, wobei das Risiko bei einem Wert von Eins maximal, bei einem Wert von Null minimal ist. Das Maß eignet sich insbesondere für Fälle, in denen keine direkten Wahrscheinlichkeitsaussagen über die Ordnungsmäßigkeit bzw. die Nichtordnungsmäßigkeit der Prüfungsgesamtheit möglich sind. Allerdings erfordert es, wenn bereits vorliegende Prüfungsinformationen in die Beurteilung einbezogen werden sollen, daß diese ebenfalls in Form von likelihoods vorliegen. Dieser Umstand führt, wie schon gezeigt wurde, bei den vorhandenen Prüfungsmethoden zu Problemen[33].

Darüber hinaus wurde bei der Diskussion zur Bestimmung der Hypothesen deutlich, daß die vorliegende Prüfungssituation mit der den Testverfahren zugrundeliegenden Konzeption hinsichtlich der Art der zu testenden Hypothesen nicht übereinstimmt. Dabei ist das zentrale Problem, wie am Verlauf der Operationscharakteristik in Abb. 3.3 zu erkennen ist, daß die Annahmewahrscheinlichkeit für die Nullhypothese bei Prüfungsgesamtheiten, in denen der wahre Fehleranteil zwischen θ_0 und $\theta_1 = \theta^*$ liegt, weniger als $1-\alpha$, in der Nähe von θ_1

---

[32] Vgl. *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 167f. Zur Risikomessung wird hier allerdings der Kehrwert des von *Stegmüller* vorgeschlagenen Wertes für die Schärfe eines Tests verwendet.

[33] Siehe oben Kapitel C. II. 2. im Zweiten Teil der Arbeit.

nur wenig mehr als β beträgt, obwohl nach den getroffenen Annahmen die genannten Prüfungsgesamtheiten ordnungsmäßig sein sollen. Soll demnach die Ablehnung der Nullhypothese auch zur Ablehnung der Ordnungsmäßigkeit der Prüfungsgesamtheit führen, dann ist γ zur Messung des Risikos dieser Entscheidung nicht geeignet, wenn wahre Fehleranteile im Intervall $[\theta_0;\theta_1]$ auftreten können.

Damit ergeben sich für die Interpretation des Testergebnisses zwei Konsequenzen. Einmal kann die Wahrscheinlichkeit für einen Fehler 1. Art, wenn damit jede Ablehnung der Ordnungsmäßigkeit einer tatsächlich ordnungsmäßigen Prüfungsgesamtheit bezeichnet werden soll, nicht mehr auf α beschränkt werden. Als Obergrenze für die Wahrscheinlichkeit einer solchen Fehlentscheidung erhält man, wie Abb. 3.4 verdeutlicht, unabhängig von der verwendeten Nullhypothese und dem realisierten Stichprobenergebnis immer $1-\beta$. Die Festlegung eines Wertes für die Nullhypothese und für α dient dann im vorliegenden Prüfungsmodell ausschließlich dazu, den Verlauf der Operationscharakteristik für den Test so festzulegen, daß dieser dem Prüfer angemessen erscheint, ohne daß explizit deutlich gemacht wird, wie er zu dieser Festlegung kommt.

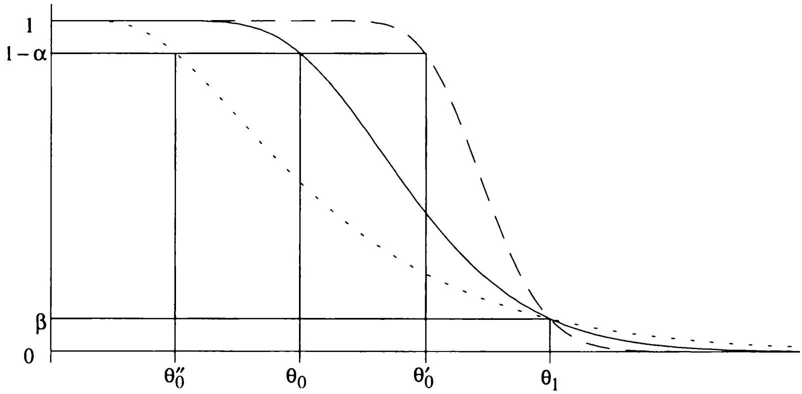

Abb. 3.4: Verlauf der Operationscharakteristik bei alternativen Werten für die Nullhypothese

Als zweite Konsequenz folgt, daß aus der Ablehnung der Nullhypothese nicht ohne weiteres auf die Nichtordnungsmäßigkeit der Prüfungsgesamtheit geschlossen werden kann. Dies kann erst dann erfolgen, wenn zusätzliche Informationen über den Fehleranteil in der Prüfungsgesamtheit vorliegen. Im folgenden soll davon ausgegangen werden, daß die vorliegenden Informationen nicht ausreichen, um einen Bereich von Fehleranteilen $[\theta';\theta^*]$ und $\theta'>0$ mit Sicherheit für die Prüfungsgesamtheit auszuschließen, da es dann durch die Wahl von $\theta_0 = \theta'$ möglich wäre, die Fragestellung an die den Testverfahren zugrundeliegende Konzeption anzupassen. Im besten Fall kann aus den vorliegenden Informationen ein wahrscheinlicher Wert für die Obergrenze des Fehleranteils in der Prüfungsgesamtheit abgeleitet werden. Falls diese kleiner ist als der kritische Fehleranteil θ^*, kann die vermutete Obergrenze zur Formulierung der Nullhypothese verwendet werden. Allerdings folgt dann aus einer Ablehnung der Nullhypothese ausschließlich, daß entweder die aus den Informationen abgeleitete Einschätzung über den Fehleranteil falsch ist, oder ein Stichprobenergebnis mit einer geringeren Wahrscheinlichkeit als α eingetreten ist. Soweit eine Wahrscheinlichkeit für die in der Nullhypothese verwendete Obergrenze angegeben werden kann, lassen sich die beiden Wahrscheinlichkeitseinschätzungen mittels des Theorems von *Bayes* zusammenfassen, wenn auch die absolute Wahrscheinlichkeit für Fehleranteile, die größer als θ_1 sind, bekannt ist. Es gilt dann:

(3.28)
$$pr(\theta \le \theta_0 | A_0) = \frac{\alpha\, pr(\theta \le \theta_0)}{\alpha\, pr(\theta \le \theta_0) + (1-\beta) pr(\theta > \theta_1)}.$$

Insbesondere für die Ableitung von $pr(\theta \le \theta_0)$ und $pr(\theta > \theta_1)$ sind allerdings die oben gegen die Anwendung des Theorems von *Bayes* vorgebrachten Einwendungen zu beachten[34]. Im Schrifttum zur Jahresabschlußprüfung wird darüber hinaus vorgeschlagen, die Abschätzung des wahrscheinlichen Wertes für die Obergrenze des Fehleranteils in der Prüfungsgesamtheit durch das zu prüfende Unternehmen selbst vornehmen zu lassen. Damit wird zwar dem Unternehmen die Möglichkeit eröffnet, über die Qualität der erstellten Daten die Prüfungskosten zu beeinflussen, da c. p. mit zunehmendem Abstand zwischen den Hypothesen der notwendige Stichprobenumfang sinkt[35], allerdings kann der Abschlußprüfer dann die Wahrscheinlichkeitseinschätzung nicht ohne weiteres für die Ableitung des Risikomaßes verwenden, da er gemäß § 43 Abs. 1 WPO zur eigenverantwortlichen Prüfung verpflichtet ist.

Neben der Anwendung des Theorems von *Bayes* kommt bei erfolgter Ablehnung der Nullhypothese durch den Test die Ausdehnung des Prüfungsumfangs in Betracht[36]. Falls zur Beurteilung der Prüfungsgesamtheit ein bester Test herangezogen wurde, führt, wie in Abb. 3.5 deutlich wird, eine Erhöhung des Stichprobenumfangs unter der Voraussetzung, daß die Irrtumswahrschein-

[34] Siehe oben Kapitel C. II. 3. im Zweiten Teil der Arbeit.

[35] Vgl. *Schildbach*, Stichprobenprüfung, 1992, Sp. 1849.

[36] Vgl. *IDW*, HFA 1/1988, 1988, S. 244; *AICPA*, SAS No. 39, 1994, AU § 350.13.

DOI https://doi.org/10.3790/978-3-428-48909-1

lichkeiten zunächst nicht verändert werden, zur Überschneidung der Ablehnungsbereiche für die Hypothesen. Falls das Stichprobenergebnis in den Überschneidungsbereich fällt, ist fraglich, welche der Testparameter anzupassen sind, um wiederum einen besten Test, jetzt allerdings mit höherer Sicherheit oder Genauigkeit zu erhalten.

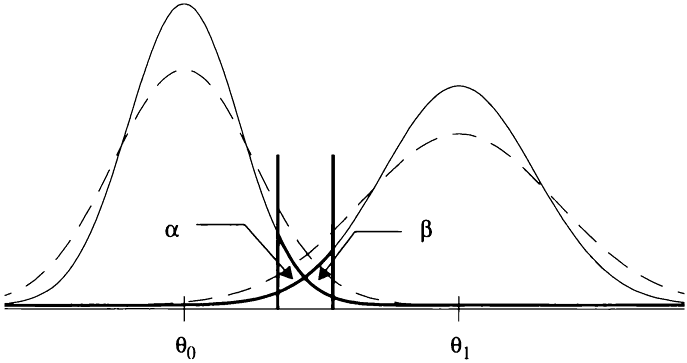

Abb. 3.5: Wirkung der Erhöhung des Stichprobenumfangs bei einem Fehleranteilstest

Eine Alternative besteht darin, die Höhe der Irrtumswahrscheinlichkeiten im gleichen Verhältnis zu reduzieren. Dies führt dazu, daß der für die Entscheidungsfindung relevante Wert p^*, der sich nach der Erhöhung des Stichprobenumfangs ergibt, mit dem Wert vor der Erhöhung des Stichprobenumfangs identisch ist[37]. Insoweit kommt der Test ausschließlich dann zu einem anderen Ergebnis, wenn sich der Schätzwert für den Fehleranteil in der größeren Stichprobe gegenüber dem vorherigen Zustand verringert. Eine Verringerung der Irrtumswahrscheinlichkeit für wahre Fehleranteile zwischen den Hypothesen erfolgt nicht. Es erscheint daher der vorliegenden Prüfungssituation eher angemessen, wenn unter Beibehaltung von β, das bei der ursprünglichen Konzeption des Tests für ausreichend gehalten wurde, die Wahrscheinlichkeit für einen Fehler 1. Art soweit verringert wird, daß die Nullhypothese angenommen werden kann, falls wie oben unterstellt wurde, das Testergebnis im Überschneidungsbereich liegt.

Zum gleichen Ergebnis kommt man, wenn nicht die Irrtumswahrscheinlichkeit angepaßt wird, sondern die Hypothesen verändert werden. Da die Alternativhypothese so gewählt wurde, daß sie mit dem vorgegebenen kritischen Fehleranteil θ^* übereinstimmt, ist eine Veränderung von θ_1 nicht möglich, ohne

[37] Vgl. *Obermeier*, Abschlussprüfung, 1983, S. 158.

neue Kriterien für das Gesamturteil über die Prüfungsgesamtheit festzulegen. Es bleibt demnach ausschließlich die Möglichkeit θ_0 soweit zu erhöhen, daß mit der vorgegebenen Wahrscheinlichkeit für einen Fehler 1. Art die Annahme der Nullhypothese erfolgen kann. Wobei wiederum davon ausgegangen wurde, daß das Stichprobenergebnis im Überschneidungsbereich lag. Dieser Vorgehensweise liegt die Darstellung in Abb. 3.6 zugrunde. Falls das Testergebnis wiederum zur Ablehnung der Nullhypothese führt, beträgt zwar die Wahrscheinlichkeit für einen Fehler 1. Art immer noch maximal $1-\beta$, allerdings wird der Bereich von Fehleranteilen, bei denen die Irrtumswahrscheinlichkeit geringer als α ist, größer.

Insgesamt ist festzuhalten, daß wegen der Relation $\theta_1 = \theta^*$ und der dazu festgelegten Obergrenze für die Wahrscheinlichkeit, einen Fehler 2. Art zu begehen, ausschließlich die Nullhypothese oder die Wahrscheinlichkeit für einen Fehler 1. Art verändert werden können, ohne die Urteilskriterien für die Prüfungsgesamtheit neu zu bestimmen. Insoweit gilt auch für den oben angesprochenen Fall, daß, falls bei nicht ausreichendem Stichprobenumfang das Stichprobenergebnis in den Indifferenzbereich fällt[38], ausschließlich die Möglichkeit besteht, die Ordnungsmäßigkeit der Prüfungsgesamtheit abzulehnen und die über die Vorgabe hinausgehende Wahrscheinlichkeit für einen Fehler 1. Art in Kauf zu nehmen, wenn nicht der Stichprobenumfang erhöht wird.

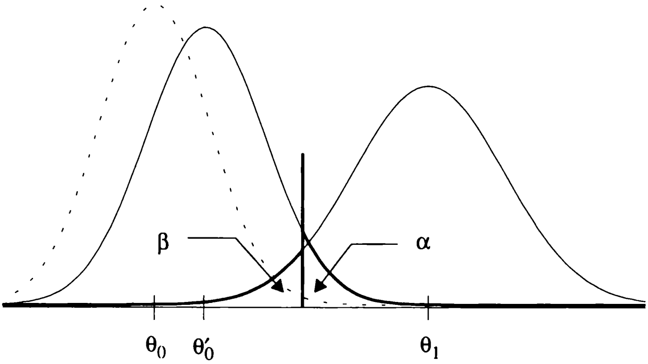

Abb. 3.6: Wirkung der Erhöhung der Nullhypothese bei einem Fehleranteilstest

Für die Prüfungsplanung erscheint es zunächst so, daß sich bei der Verwendung von Testverfahren Vorteile gegenüber den Schätzverfahren ergeben, da zum einen bei der Bestimmung des notwendigen Stichprobenumfangs gemäß

[38] Siehe dazu oben Abb. 3.2.

(3.22) keine Annahmen über den wahren Fehleranteil in der Prüfungsgesamtheit gemacht werden müssen. Zum anderen kann der notwendige Stichprobenumfang so bestimmt werden, daß der Test bei jedem möglichen Stichprobenergebnis entscheidet. Wie die vorstehenden Ausführungen zur Risikomessung allerdings gezeigt haben, sind zum einen für die Formulierung der Nullhypothese Annahmen über den wahren Fehleranteil in der Prüfungsgesamtheit zu treffen, zum anderen ist zumindest dann, wenn das Stichprobenergebnis zur Ablehnung der Nullhypothese führt, auch eine Erhöhung des Stichprobenumfangs notwendig, wenn das gemessene Risiko als zu hoch erscheint. Die Planungsprobleme für beide Auswertungsmethoden sind demnach weitgehend identisch.

2. *Unkenntnis der Grenzverteilung*

Bisher wurde davon ausgegangen, daß die Grenzverteilung der verwendeten Stichprobenfunktion bekannt ist. Dies trifft im vorliegenden Prüfungsmodell ausschließlich für Fehleranteilschätzungen oder -tests zu. In diesem Fall ist die in (3.2) angegebene Stichprobenfunktion hypergeometrisch verteilt. Für die Verteilungsfunktion gilt dann:

(3.29) $$F_H(x,\theta,n,N)=\frac{\sum_{j=0}^{x}\binom{\theta N}{j}\binom{(1-\theta)N}{n-j}}{\binom{N}{n}} \text{ für } x=0,1,\ldots,n.$$

Die Bestimmung von Werten der hypergeometrischen Verteilung bereitet beim Einsatz von EDV und geeigneter Software keine technischen Probleme[39], insoweit muß in diesem Fall keine zusätzliche Unschärfe auftreten.

Probleme bei der Bestimmung der Grenzverteilung ergeben sich, wenn im Rahmen der Stichprobenverfahren Aussagen über das Fehlerausmaß in der Prüfungsgesamtheit getroffen werden sollen. Die Approximation der hypergeometrischen Verteilung als Grenzverteilung durch die Normalverteilung ist bei geeigneter Wahl des Stichprobenumfangs und bekannter Varianz der Fehler in der Prüfungsgesamtheit grundsätzlich möglich. Fraglich ist, wie groß der Stichprobenumfang sein muß, damit die Stichprobenergebnisse näherungsweise normalverteilt sind. Dabei ist entscheidend, wie schnell die Grenzverteilung der Stichprobenfunktion mit zunehmendem Stichprobenumfang gegen die Normal-

39 Vgl. *Mandl*, Anwendungsvoraussetzungen, 1984, S. 123f.; *v. Wysocki*, Grundlagen, 1988, S. 201-203.

verteilung konvergiert. Da die Normalverteilung symmetrisch um den Mittelwert ist, wird angenommen, daß die Konvergenzgeschwindigkeit insbesondere von der Schiefe der Verteilung der betrachteten Werte in der Prüfungsgesamtheit abhängt. Bezeichnet G_1 Fisher's Maß der Schiefe, wird vorgeschlagen, bei einfacher Mittelwertschätzung aus einer einfachen Stichprobe den Stichprobenumfang so zu wählen, daß gilt[40]:

$$n > 25 G_1^2. \tag{3.30}$$

Für die Risikomessung interessiert darüber hinaus, wie gut die Annäherung an die Normalverteilung bei einem gegebenen Stichprobenumfang ist bzw. welche Auswirkungen sich auf die Sicherheit der Schätzung ergeben. Diese Angaben können aus (3.30) nicht abgeleitet werden. Für den Zusammenhang zwischen Approximationsgüte ϕ bei einem bestimmten Konfidenzniveau $z(1-\alpha)$ und der Schiefe der Verteilung G_1 der Werte in der Prüfungsgesamtheit erhält man in erster Näherung[41]:

$$\phi = \frac{\left|1 - z(1-\alpha)^2\right|}{6} \frac{G_1}{\sqrt{n}} \frac{1}{\pi\sqrt{2}} e^{-\frac{z(1-\alpha)^2}{2}}. \tag{3.31}$$

Ist G_1 bekannt so ergibt sich für einen gegebenen Stichprobenumfang als Konfidenzniveau unter Berücksichtigung der Approximationsgüte $1-\alpha-\phi$.

Probleme bei der Abschätzung der Unschärfe, die sich aus der Unkenntnis der genauen Grenzverteilung ergeben, entstehen, wenn die Varianz bzw. die Schiefe der Verteilung der Fehler in der Prüfungsgesamtheit unbekannt sind. Es besteht zwar grundsätzlich die Möglichkeit die Parameter aus der Stichprobe zu schätzen, allerdings sind die dabei möglichen Verzerrungen, wenn die Werte in der betrachteten Gesamtheit nicht normalverteilt sind, nur schwer abzuschätzen[42]. Darüber hinaus treten Probleme auf, wenn keine Fehler in der Stichprobe gefunden wurden. Die Ableitung der benötigten Parameter aus den Werten der Prüfungsgesamtheit ist bei dem vorliegenden Modell nicht möglich, da angenommen werden muß, daß das Ausmaß der einzelnen Fehler vom Wert des jeweiligen Elementes unabhängig ist.

---

[40] Vgl. *Cochran*, Sampling, 1977, S. 40-42; *Obermeier*, Abschlussprüfung, 1983, S. 113f.

[41] Vgl. *Bühler*, Stichprobenumfang, 1984, S. 707f.

[42] Vgl. *Cochran*, Stichprobenverfahren, 1972, S. 62f.; *Obermeier*, Abschlussprüfung, 1983, S. 111.

Im Schrifttum wird vorgeschlagen, in solchen Fällen eine konservative Abschätzung des Fehlerausmaßes mit Hilfe des aus der Stichprobe ermittelten Konfidenzintervalls für den Fehleranteil vorzunehmen, dazu soll gelten[43]:

$$\Delta_o = N p_o d_{max} \tag{3.32}$$

Dabei bezeichnet Δ_o die Konfidenzintervallobergrenze für das Fehlerausmaß und d_{max} den maximal möglichen Fehlerbetrag, der in der Prüfungsgesamtheit auftreten kann. Das Konfidenzniveau für das nach (3.32) ermittelte Konfidenzintervall entspricht dem Konfidenzniveau, das für den zugrundegelegten Fehleranteil p_o ermittelt wurde. Mit dieser Vorgehensweise sind im vorliegenden Prüfungsmodell zwei Probleme verbunden. Zum einen wird, selbst wenn der maximale Fehlerbetrag bekannt sein sollte, in realen Prüfungssituationen nur eine sehr ungenaue Schätzung möglich sein[44]. Zum anderen widerspricht die Vorgehensweise der im vorliegenden Prüfungsmodell für die Beurteilung der Prüfungsgesamtheit getroffenen Annahme, daß die Anzahl der Fehler und das relative Gewicht, das ihnen beizumessen ist, voneinander unabhängig sind und daher in zwei getrennten Kriterien berücksichtigt werden müssen. Insgesamt bleibt demnach festzustellen, daß die Verwendung von Zufallsstichprobenverfahren zur Schätzung des Fehlerausmaßes in der Prüfungsgesamtheit nur eingeschränkt möglich sein wird.

II. Stichproben mit bewußter Auswahl

Wie schon mehrfach ausgeführt wurde, beruht die Unschärfe von Prüfungsinformationen aus Stichproben, die bewußt ausgewählt wurden[45], auf der Ungewißheit darüber, ob ein der vorliegenden Prüfungssituation angemessenes Auswahlkriterium verwendet wurde. Im Unterschied zu den auf Zufallsstichproben beruhenden Verfahren variiert das Stichprobenergebnis für eine bestimmte Prüfungsgesamtheit nicht, wenn das gleiche Auswahlkriterium mehrfach angewandt wird. Falls ein Risikomaß für die Ergebnisse aus der Stichprobe abgeleitet werden soll, muß demnach ein Weg gefunden werden, die Angemessenheit eines Auswahlkriteriums zu messen. Zur Diskussion der Alternativen, die für die Risikomessung zur Verfügung stehen, soll zunächst die dem Verfahren zugrundeliegende Konzeption dargestellt werden. Anschließend wird dann für die unterschiedlichen im Schrifttum zur Jahresabschlußprüfung

---

43 Vgl. *Sperl*, Prüfungsplanung, 1978, S. 173f.

44 Vgl. *Bühler*, Stichprobenumfang, 1984, S. 702.

45 Wenn im folgenden von Stichproben die Rede ist, sollen immer Stichproben, die auf bewußter Auswahl der Stichprobenelemente beruhen, gemeint sein.

genannten Auswahlkriterien geprüft, welche Maße für die Risikomessung bei deren Anwendung zur Verfügung stehen.

1. Konzeption

Für die Vorgehensweise bei der Prüfung wird unterstellt, daß zunächst mit Hilfe eines bestimmten Auswahlkriteriums eine Teilmenge X' der Prüfungsgesamtheit ausgewählt wird. Nach Durchführung der Prüfungshandlungen, d. h. nach Ableitung der zugehörigen Menge der Sollobjekte Y', liegt eine Menge D' vor die folgendermaßen definiert ist:

(3.33) $$D' = \{\langle x_i, y_i \rangle | x_i \in X' \wedge y_i \in Y' \wedge \pi_D(\langle x_i, y_i \rangle) = 1\}.$$

Das Risiko bei der hier unterstellten stichprobenweisen Prüfung resultiert aus dem Umstand, daß die Beurteilung der Prüfungsgesamtheit nicht anhand der Menge D, wie sie in (1.16) definiert wurde, sondern anhand von D' erfolgen soll. Unproblematisch im Hinblick auf die Risikomessung ist der Fall, daß die in D' enthaltenen Fehler ausreichen, um die Ordnungsmäßigkeit abzulehnen. Für diesen Fall ist das Risiko bei der Wahl von Handlungsalternative a_1 Eins, bei der Wahl von a_2 Null, da die tatsächliche Ausprägung der Prüfungsgesamtheit bekannt ist. Insoweit wird im folgenden, wenn von Risikomessung die Rede ist, davon ausgegangen, daß aus der vorliegenden Stichprobe kein eindeutiges Prüfungsergebnis abgeleitet werden kann.

Für den Fall, daß die in D' enthaltenen Fehler nicht ausreichen, um die Ordnungsmäßigkeit abzulehnen, hängt das Risiko der möglichen Entscheidungen demnach von Annahmen über die Identität der Mengen D und D' ab. Im für die Risikomessung einfachsten Fall kann der Prüfer die (subjektive) Wahrscheinlichkeit für die Übereinstimmung der beiden Mengen quantifizieren. Die Risikomessung erfolgt dann analog den mathematisch-statistischen Schätzverfahren mittels eines Glaubwürdigkeitsmaßes[46]. Wird mit $pr(D = D')$ die Wahrscheinlichkeit für die Übereinstimmung der beiden Mengen bezeichnet, dann erhält man, da aus der Nichtübereinstimmung der beiden Mengen nicht auf Nichtordnungsmäßigkeit der Prüfungsgesamtheit geschlossen werden darf, $bel(s_1) = pr(D = D')$ und $bel(s_2) = 0$ und damit als Risiko bei der Wahl von Handlungsalternative a_1 $pl(s_2) = 1 - pr(D = D')$ sowie $pl(s_1) = 1$ bei der Wahl von a_2.

Nachteilig an dieser Vorgehensweise ist, daß nicht transparent wird, wie die vom Prüfer vorgenommene Glaubwürdigkeitseinschätzung zustande kommt.

---

46 Vgl. *Shafer* et al., Auditor's Assistant, 1988, S. 71.

Es wird daher im weiteren unterstellt, daß für das verwendete Auswahlkriterium Voraussetzungen angeben werden können, bei deren Vorliegen die Verwendung des Kriteriums angemessen, d. h., die Identität der beiden Mengen D und D' garantiert ist. Die logische Form der Schlußfolgerung, die dabei verwendet wird, entspricht der Vorgehensweise im Rahmen des indirekten Messens. Es muß demnach eine Majorprämisse, die eine Gesetzmäßigkeit zwischen bestimmten Eigenschaften der Prüfungsgesamtheit und der Angemessenheit eines Auswahlkriteriums herstellt, bekannt sein, die zusammen mit der Minorprämisse, in der das Vorliegen der Voraussetzungen festgestellt wird, zu einer Schlußfolgerung über die Angemessenheit des Auswahlkriteriums führt[47]. Für den Fall, daß keinerlei Unschärfe auftritt, läßt sich die Schlußfolgerung formal beschreiben durch:

$$((\mathrm{a} \rightarrow D = D') \wedge \mathrm{a}) \rightarrow D = D'. \tag{3.34}$$

Dabei bezeichnet a die logisch wahre Aussage, daß die Voraussetzungen für die Angemessenheit des verwendeten Auswahlkriteriums auf die Prüfungsgesamtheit zutreffen. Damit die Voraussetzungen überprüft werden können, müßten sie Aussagen über meßbare Eigenschaften der Prüfungsgesamtheit enthalten. Dabei genügt für den hier betrachteten Fall, wenn die Eigenschaft auf einer Nominalskala gemessen werden kann. Auch wenn bei der in (3.34) verwendeten Formulierung keinerlei Unschärfe auftritt, ist für den Fall, daß die Voraussetzungen nicht oder nicht vollständig erfüllt werden, keine eindeutige Aussage anhand des Stichprobenergebnisses möglich, da die Mengen D und D' übereinstimmen können oder auch nicht, es sei denn, daß eine Äquivalenzrelation zwischen den Voraussetzungen und der Identität der beiden Mengen konstruiert werden kann.

Eine erste Möglichkeit, Unschärfe bei der Formulierung des Zusammenhangs zwischen Voraussetzungen und Identität der beiden Mengen D und D' zu berücksichtigen, besteht darin, daß bei vorliegenden Voraussetzungen die Identität der beiden Mengen nicht generell behauptet wird. Zur Modellierung dieser Form von Unschärfe bei der Majorprämisse bestehen grundsätzlich zwei Möglichkeiten. Zum einen kann, falls eine Abschätzung der relativen Häufigkeit für die Gültigkeit der Implikation möglich ist, die relative Häufigkeit als Abschätzung der bedingten Wahrscheinlichkeit für die Identität der beiden Mengen bei vorliegenden Voraussetzungen $pr(D = D' | A)$ verwendet werden. Der relevante Möglichkeitsraum über dem das Wahrscheinlichkeitsmaß definiert ist, setzt sich dabei aus den möglichen Abbildungen der Realität für das be-

47 Vgl. *Hagest*, Urteilsstichprobe, 1976, S. 125.

trachtete Unternehmen, von denen eine in X realisiert ist, zusammen. Der relevante Ereigniskörper enthält zumindest die Menge der Abbildungen A, bei denen die angegebenen Voraussetzungen für die Verwendung des betrachteten Auswahlkriteriums erfüllt sind, die Menge der Abbildungen, die beim betrachteten Auswahlkriterium zu identischen Mengen D und D' führen, und die gemäß (2.1) bis (2.3) benötigten Schnittmengen sowie die Komplemente der angeführten Mengen. Für die Risikomessung im Hinblick auf die Beurteilung der Ordnungsmäßigkeit resultiert dann ein Glaubwürdigkeitsmaß mit der gleichen Argumentation, die auch oben für den Fall angeführt wurden, daß die absolute Wahrscheinlichkeit für die Identität der beiden Mengen angegeben werden kann.

Zum anderen besteht die Möglichkeit, die Unschärfe durch eine linguistische Variable HÄUFIGKEIT, deren Ausprägung im vorliegenden Fall z. B. „meistens" oder „fast immer" sein könnte, zu modellieren. Liegen die Voraussetzungen der Implikation vor, dann kann die resultierende Ausprägung der linguistischen Variable als Möglichkeitsverteilung betrachtet werden, und zur Ableitung eines Notwendigkeitsmaßes für die Übereinstimmung der Mengen dienen[48]. Den beiden Alternativen liegen insoweit unterschiedliche Interpretationen der betrachteten Unschärfe zugrunde. Während die Wahrscheinlichkeitsaussage direkt die Ungewißheit über die Identität der beiden Mengen ausdrückt, kann durch die Verwendung einer linguistischen Variable auch die Unschärfe, die in einem bestimmten Kontext, wie z. B. einer Gruppe von Prüfern, über die Gültigkeit der Implikation besteht, ausgedrückt werden.

Bisher wurde davon ausgegangen, daß die Minorprämisse eine eindeutige Aussage über das Vorliegen der Voraussetzungen enthält. Liegt hier ausschließlich eine Wahrscheinlichkeits- oder Glaubwürdigkeitsbeurteilung vor, dann kann die oben angeführte bedingte Wahrscheinlichkeit nicht verwendet werden, da dies voraussetzen würde, daß das Ereignis A tatsächlich eingetreten ist. Werden sowohl die für die Majorprämisse als auch die für die Minorprämisse vorliegenden Wahrscheinlichkeits- oder Glaubwürdigkeitsbeurteilungen als Evidenzen aufgefaßt, die sich mit Hilfe von „Dempster's rule of combination" nach (2.50) kombinieren lassen, dann kann eine gemeinsame Glaubwürdigkeitsbeurteilung für die Schlußfolgerung ermittelt werden[49]. Diese kann zur Risikomessung herangezogen werden. Für die Ableitung des Risikomaßes tritt

48 Siehe oben Kapitel B. III. 3. im Zweiten Teil der Arbeit.

49 Zur Ermittlung der gemeinsamen Glaubwürdigkeitsbeurteilung siehe unten Kapitel C. I. 2. in diesem Teil der Arbeit.

DOI https://doi.org/10.3790/978-3-428-48909-1

die Glaubwürdigkeitsbeurteilung an die Stelle der bedingten Wahrscheinlichkeit.

Voraussetzung für die Anwendung von (2.50) ist zum einen, daß die Evidenzen unabhängig voneinander sind, zum anderen dürfen sie wegen (2.51) nicht vollständig widersprüchlich sein. Die Unabhängigkeit der Erkenntnisse ist dann gewährleistet, wenn sie nicht aus einer einzigen Quelle gewonnen wurden. Da es sich bei der Majorprämisse um eine allgemeine Gesetzmäßigkeit handeln soll und die Gültigkeit der Minorprämisse für die betrachtete Prüfungsgesamtheit bestimmt wird, ist die Unabhängigkeit der Erkenntnisse gewährleistet. Da die Unschärfe bei der Majorprämisse in Form einer bedingten Wahrscheinlichkeitsaussage bei vorliegenden Voraussetzungen formuliert wurde, kann kein Widerspruch zu der Erkenntnis, daß die Voraussetzungen mit einer bestimmten Wahrscheinlichkeit oder Glaubwürdigkeit vorliegen bestehen. Inwieweit eine Risikomessung möglich ist, falls die Unschärfe bei der Majorprämisse mit Hilfe eines Notwendigkeitsmaßes formuliert wird, soll an dieser Stelle nicht geprüft werden. Ansätze und Voraussetzungen dazu werden allerdings weiter unten diskutiert[50].

Den bisherigen Ansätzen zur Modellierung des Zusammenhangs zwischen Voraussetzungen für ein Auswahlkriterium und Übereinstimmung der beiden Mengen D und D' ist gemeinsam, daß zum einen die Risikomessung ausschließlich dann erfolgen kann, wenn die Minorprämisse und die in der Majorprämisse angeführten Voraussetzungen genau übereinstimmen. Zum anderen konnte aus den bisher angesprochenen Risikomaßen nur ein Maß für die Ungewißheit darüber abgeleitet werden, ob eine Abweichung zwischen den beiden Mengen besteht. Es beinhaltet insoweit keinerlei Aussage über den Umfang der Abweichung, falls die Voraussetzungen nicht oder nicht vollständig vorliegen. Beide Einschränkungen können bei bestimmten Annahmen aufgehoben werden. Dazu ist allerdings notwendig, daß die Eigenschaft der Prüfungsgesamtheit, deren Ausprägungen bei der Festlegung der Voraussetzungen verwendet werden, und die zu ermittelnde Abweichung zumindest grundsätzlich auf metrischen Skalen gemessen werden können. Für die zu ermittelnde Abweichung zwischen D und D' liegt Meßbarkeit auf einer metrischen Skala vor, da ausschließlich die Auswirkungen im Hinblick auf die Größen θ und $\Delta(X,Y)$ interessieren.

Für die folgende Darstellung soll vereinfachend nur die Abweichung zischen D und D' im Hinblick auf den Fehleranteil betrachtet werden. Für diese Abwei-

50 Siehe unten Kapitel C. I. 3. in diesem Teil der Arbeit.

chung wird unterstellt, daß ihr Umfang U durch die Ausprägungen einer geeignet definierten linguistischen Variable ausgedrückt werden kann. Darüber hinaus soll als Voraussetzung für die Angemessenheit eines bestimmten Auswahlkriteriums ausschließlich eine Eigenschaft A der Prüfungsgesamtheit, die ebenfalls durch eine linguistische Variable beschrieben werden kann, betrachtet werden. Der unscharfe Zusammenhang zwischen der Voraussetzung und der Abweichung zwischen den Mengen D und D' kann dann folgendermaßen formuliert werden[51]:

$$\left(\left(\mathrm{A} = \tilde{A} \rightarrow \mathrm{U} = \tilde{U}\right) \wedge \mathrm{A} = \tilde{A}'\right) \rightarrow \mathrm{U} = \tilde{U}'. \tag{3.35}$$

Die in (3.35) verwendete Schlußform wird als „plausibles Schließen" bezeichnet[52]. Im Unterschied zu der in (3.34) angegebenen Schlußform sind zum einen die einzelnen Aussagen unscharf formuliert, zum anderen wird in (3.35) nicht verlangt, daß die Minorprämisse exakt mit dem Vordersatz der als Majorprämisse verwendeten Implikation übereinstimmt[53]. Die Anwendung der Schlußform auf eine konkrete Situation, in der die Zugehörigkeitsfunktionen für $\tilde{A}$, $\tilde{U}$ und $\tilde{A}'$ bekannt sein müssen, liefert dann eine Zugehörigkeitsfunktion für die Ausprägung $\tilde{U}'$ der linguistischen Variable, welche die Abweichung zwischen D und D' im Hinblick auf den Fehleranteil beschreibt. Aus dieser kann dann ein Notwendigkeitsmaß sowohl für die Ordnungsmäßigkeit als auch für die Nichtordnungsmäßigkeit der Prüfungsgesamtheit zumindest im Hinblick auf den Fehleranteil abgeleitet werden, da mit der durch $\pi_{\tilde{U}'}$ repräsentierten grundlegenden Möglichkeitszuordnung auch Informationen über mögliche Fehleranteile, die über die in D' enthaltenen Fehler hinausgehen, vorliegen. Der Ursprung für die Informationen über die nicht in D' enthaltenen Fehler muß in den verwendeten Voraussetzungen zumindest implizit enthalten sein. Damit die Anwendung der unscharfen Schlußform in (3.35) möglich ist, muß zumindest ein unscharfer funktionaler Zusammenhang zwischen den Voraussetzungen und den Fehlern in der Prüfungsgesamtheit bestehen[54].

Allerdings erfordert die Anwendung der in (3.35) dargestellten Schlußform neben der Kenntnis der oben angeführten Zugehörigkeitsfunktionen auch

---

[51] Die im folgenden Ausdruck verwendeten unscharfen logischen Operatoren entsprechen in ihrer Bedeutung den jeweiligen, aus der zweiwertigen Logik bekannten Operatoren. Auf die Einführung neuer Symbole wurde verzichtet, da aus dem Kontext hervorgeht, welche Klasse von Operatoren gemeint ist.

[52] Vgl. *Werners*, Inferenz , 1994, S. 253.

[53] Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 128.

[54] Vgl. hierzu auch den Ansatz für die Generalisierung des modus ponens in *Zadeh*, Concept, Part 3, 1976, S. 57f.

Festlegungen im Hinblick auf die zu verwendenden Operatoren, namentlich den Operator für die unscharfe Konjunktion, einen Operator für die zulässigen Modifikationen von $\tilde{A}$ nach $\tilde{A}'$ sowie den Implikationsoperator. Da die Konjunktion von logischen Ausdrücken und die Durchschnittsbildung von Mengen isomorph sind, können für die unscharfe Konjunktion die oben angeführten t-Normen oder Durchschnittsoperatoren[55] verwendet werden. Für die Operatoren, die der Modifikation von Ausprägungen linguistischer Variablen dienen, wie z. B. „sehr“ oder „mehr oder weniger“, wurde oben schon ausgeführt, daß der Anwendungsbereich dieser Operatoren bisher theoretisch wenig untersucht ist[56]. Die Anwendbarkeit einzelner im Schrifttum vorgeschlagener Operatoren muß demnach im Einzelfall geprüft werden, und ihre Verwendung im Hinblick auf die damit erzielten Ergebnisse begründet werden[57]. Auch für die unscharfe Implikation stehen unterschiedliche Klassen von Operatoren zur Verfügung[58], die jeweils zu unterschiedlichen Ergebnissen bei der Schlußfolgerung führen können[59]. Wie für die Modifikationsoperatoren gilt auch für die unscharfe Implikation, daß die Interpretation der einzelnen Operatoren theoretisch noch nicht ausreichend untersucht ist[60], so daß auch hier die Anwendung einzelner Operatoren nicht theoretisch, sondern ausschließlich für den Einzelfall im Hinblick auf die damit erzielten Ergebnisse begründet werden kann.

2. Auswahlkriterien

Nachdem bisher mögliche Konzeptionen für die Risikomessung bei der Verwendung von Stichproben mit bewußter Auswahl dargestellt wurden, soll im folgenden für die im Schrifttum zur Jahresabschlußprüfung diskutierten Auswahlkriterien die Anwendbarkeit der Konzepte geprüft werden. Die Auswahl typischer Fälle weicht dabei grundsätzlich von der oben für die Stichprobenverfahren mit bewußter Auswahl unterstellten Konzeption ab, da nicht angestrebt wird, alle Fehler zu finden, sondern eine für die jeweils betrachtete Äquivalenzklasse repräsentative Stichprobe zu erhalten. Dabei tritt zum einen das Problem auf, daß nicht unmittelbar erkennbar ist, wodurch sich typische Fälle auszeichnen. Stellt man darauf ab, daß es sich dabei um Elemente der

---

[55] Siehe oben Kapitel B. III. 2. b) im Zweiten Teil der Arbeit.

[56] Siehe oben Kapitel B. III. 3. b) im Zweiten Teil der Arbeit.

[57] So wohl auch *Zadeh*, Theory, 1979, S. 164.

[58] Eine Übersicht sowie die Untersuchung, inwieweit einzelne Operatoren bestimmte wünschenswerte Axiome erfüllen, findet sich bei *Dubois/Prade*, Approximate Reasoning, 1991, S. 155-157.

[59] Vgl. die Übersicht bei *Tilli*, Fuzzy-Logik, 1992, S. 170-178.

[60] Vgl. *Dubois/Prade*, Approximate Reasoning, 1991, S. 158f.

Prüfungsgesamtheit handelt, die den gleichen Verarbeitungsabläufen innerhalb des datenerzeugenden Systems unterliegen, dann sind die gewonnenen Aussagen im Rahmen der Systemprüfung zu verarbeiten[61]. Zum anderen bleibt unklar, wie anhand der Stichprobe eine direkte Beurteilung der Prüfungsgesamtheit erfolgen soll, da keine begründeten Hochrechungsverfahren für das Stichprobenergebnis angegeben werden. Insbesondere ist fraglich, inwieweit die Fehler für den Fall, daß es sich bei der Prüfungsgesamtheit um eine Buchführung handelt, tatsächlich bei den typischen und nicht bei den atypischen Vorgängen auftreten. Insoweit soll die Auswahl typischer Elemente hier nicht weiter untersucht werden.

Bei der Konzentrationsauswahl sollen die Elemente nach ihrer Bedeutung für das Prüfungsurteil ausgewählt werden[62]. Im vorliegenden Prüfungsmodell kann grundsätzlich jeder Fehler Bedeutung für die Beurteilung haben, da bei den verwendeten Kriterien θ und $\Delta(X,Y)$ die Gesamtheit der Fehler betrachtet wird, und damit ein einzelner Fehler, der dazu führt, daß der kritische Fehleranteil θ^* überschritten wird, unabhängig von seinem Betrag erhebliche Bedeutung für die Beurteilung hat. Allerdings haben nicht alle Fehler a priori die gleiche Bedeutung. Da große Fehlerbeträge c. p. zu einem höheren Fehlerausmaß und damit auch eher zur Nichtordnungsmäßigkeit der Prüfungsgesamtheit führen, ist ihre Bedeutung vor der Stichprobenziehung größer. Die Auswahl nach dem Konzentrationsprinzip stellt daher regelmäßig darauf ab, die betragsmäßig großen Fehler in der Prüfungsgesamtheit zu entdecken. Sie ist demnach ausschließlich zur Beurteilung des Fehlerausmaßes geeignet. Insoweit sollen die folgenden Ausführungen zu diesem Auswahlkriterium auf die Ermittlung des Fehlerausmaßes und die dabei mögliche Risikomessung beschränkt werden. Für die Ermittlung des Fehleranteils müssen demnach andere Prüfungsverfahren verwendet werden[63].

Damit eine Auswahl der Elemente aus X nach ihrer Bedeutung für das Prüfungsurteil, hier also nach der Höhe der tatsächlich vorliegenden Fehler, erfolgen kann, müßte ein Merkmal gefunden werden, dessen Ausprägungen vor Durchführung der Prüfungshandlungen bekannt und eng mit der Fehlerhöhe korreliert sind. Ein solches Merkmal existiert zumindest im Rahmen der Jahresabschlußprüfung regelmäßig nicht, insbesondere muß der Fehlerbetrag nicht mit dem Wert der einzelnen Elemente zusammenhängen. Falls die Beurteilung

---

61 Vgl. *Lanfermann*, Stichprobenprüfung, 1992, Sp. 1858.

62 Siehe oben Kapitel B. I. 2. im Ersten Teil der Arbeit.

63 Zu Möglichkeiten, die Risikomaße zu kombinieren, siehe unten Kapitel C. I. 1. in diesem Teil der Arbeit.

der Prüfungsgesamtheit auf die Fehler begrenzt werden kann, bei denen gemäß (1.32) der Sollwert jeweils kleiner als der Istwert ist, kann das Fehlerausmaß zur Beurteilung der Prüfungsgesamtheit ausreichend genau bestimmt werden[64]. Voraussetzung ist dabei, daß andere Fehler ausgeschlossen werden können oder für die Beurteilung nicht relevant sind. Die Majorprämisse enthält in diesem Fall keinerlei Unschärfe, es gilt also $pr(\Delta(X,Y)=\Delta \rceil A)=1$. Der Umstand, daß in realen Prüfungssituationen die Voraussetzungen regelmäßig nicht mit Sicherheit vorliegen werden, kann berücksichtigt werden, wenn zumindest die Wahrscheinlichkeit für die Gültigkeit der Prämissen angegeben werden kann. Die Risikomessung kann, wie oben gezeigt wurde, dann durch ein Glaubwürdigkeitsmaß erfolgen.

Erfolgt eine dedektive Auswahl[65] der Stichprobenelemente, wird unabhängig von der Bedeutung einzelner Fehler versucht, alle fehlerhaften Elemente in der Prüfungsgesamtheit zu finden. Die durch das Auswahlkriterium gewonnene Stichprobe ist demnach grundsätzlich zur Beurteilung der Prüfungsgesamtheit sowohl hinsichtlich des Fehleranteils als auch hinsichtlich des Fehlerausmaßes geeignet. Die Anwendung der dedektiven Auswahl setzt voraus, daß dem Prüfer a-priori-Informationen zur Verfügung stehen, die zur Auswahl der Stichprobenelemente verwendet werden können. Dabei sind grundsätzlich zwei Ansätze denkbar. Zum einen können die Informationen positive Hinweise auf die Merkmalsausprägungen, die fehlerhafte Elemente auszeichnen, geben, wie z. B. die schwierig zu erfassenden Sachverhalte[66], die einfache Verwertbarkeit von Vermögensgegenständen[67] oder auch Elemente, die von fehlerhaften Verarbeitungsfunktionen erzeugt wurden. Zum anderen können Informationen darüber vorliegen, welche Teile der Prüfungsgesamtheit sicher oder mit hoher Wahrscheinlichkeit fehlerfrei sind, während über die Fehlerhaftigkeit der verbleibenden Bereiche keine Aussage getroffen werden kann. Dieser Fall kann z. B. dann eintreten, wenn Bereiche im datenerzeugenden System mit nicht ausreichender Kontrolldichte identifiziert wurden. Es kann dann zwar für die Bereiche mit zuverlässigem Kontrollsystem angenommen werden, daß der Fehleranteil gering ist, aus den fehlenden Kontrollen darf jedoch nicht zwingend auf eine hohe Anzahl von Verarbeitungsfehlern in dem betroffenen Be-

---

64 Zur Vorgehensweise siehe oben Kapitel B. I. 2. im Ersten Teil der Arbeit.

65 Der Begriff Fehlerwahrscheinlichkeit soll hier vermieden werden, da er schon durch den Audit Risk Ansatz mit anderer Bedeutung belegt ist.

66 Vgl. *Hagest*, Urteilsstichprobe, 1976, S. 121.

67 Vgl. *IDW*, HFA 1/1988, 1988, S. 241.

reich geschlossen werden[68]. Es handelt sich insoweit um eine negative Abgrenzung fehlerhafter Bereiche.

Für die Majorprämisse wird im ersten Fall unterstellt, daß beim Vorliegen der positiven Hinweise für eine bestimmte Prüfungsgesamtheit die beiden Mengen D und D' dann übereinstimmen, wenn alle Elemente, für die Hinweise auf die Fehlerhaftigkeit bestanden haben, in die Stichprobe aufgenommen wurden. In Abhängigkeit davon, wie eindeutig die Majorprämisse im Hinblick auf den Zusammenhang zwischen Merkmal und Fehlerhaftigkeit formuliert ist, ergeben sich Probleme, wenn nicht alle oder der überwiegende Teil der Elemente in der Stichprobe fehlerhaft sind, da dann unterstellt werden muß, daß die in der Majorprämisse verwendete Gesetzmäßigkeit für die vorliegende Prüfungsgesamtheit nicht anwendbar ist[69]. Insoweit ist die Stichprobe für die Beurteilung der Prüfungsgesamtheit und damit auch die Risikomessung ausschließlich dann geeignet, wenn die Majorprämisse durch das Stichprobenergebnis nicht widerlegt wird. Die genannten Probleme treten bei negativer Abgrenzung der potentiell fehlerhaften Elemente nicht auf, da hier keine Hypothese über die Fehlerhaftigkeit der Elemente, die in der Stichprobe enthalten sind, aufgestellt wird, sondern eine Hypothese über den nicht in der Stichprobe enthaltenen Teil der Prüfungsgesamtheit.

Fraglich ist, welche von den oben angeführten unscharfen Schlußfolgerungen der dedektiven Auswahl zugrundegelegt werden kann. Soweit positive Hinweise auf die Fehlerhaftigkeit von Elementen der Prüfungsgesamtheit verwendet werden, kommt eine unscharfe Formulierung sowohl der Majorprämisse als auch der Minorprämisse in Betracht. Damit eine unscharfe Formulierung in der in (3.35) angegebenen Form möglich wäre, müßte ein Zusammenhang zwischen den Voraussetzungen und den nicht in die Stichprobe gelangten Elementen der Prüfungsgesamtheit hergestellt werden können. Da die unterstellte Gesetzmäßigkeit ausschließlich auf die fehlerhaften Elemente abstellt, ist diese Voraussetzung nicht erfüllt. Insoweit kommt insbesondere die Risikomessung durch ein Glaubwürdigkeitsmaß in Betracht, wenn sowohl die in der Majorprämisse benötigte bedingte Wahrscheinlichkeit, als auch die Wahrscheinlichkeit für das Vorliegen der Voraussetzungen bekannt sind. Erfolgt die negative Abgrenzung der Stichprobenelemente, dann ist grundsätzlich, da die Voraussetzungen eine Hypothese über die nicht in die Stichprobe gelangten Elemente der Prüfungsgesamtheit enthalten, auch eine unscharfe Schlußfolgerung, wie

68 Vgl. *Wittmann*, Systemprüfung, 1980, S. 149f.

69 Vgl. *Wittmann*, Systemprüfung, 1980, S. 154f.

sie in (3.35) dargestellt ist, anwendbar. Dabei können in den Voraussetzungen unscharfe Formulierungen wie z. B.:

„Wenn bestimmte Bereiche der Prüfungsgesamtheit durch zuverlässige Teile des Datenverarbeitungssystem erzeugt wurden und die übrigen Bereiche in die Stichprobe aufgenommen wurden, dann ist die Abweichung zwischen D und D' gering.“

verwendet werden. In diesem Fall könnten Zusammenhänge zwischen Modifikationen bei der Zuverlässigkeit mit entsprechenden Modifikationen beim Umfang der Abweichung zwischen D und D' begründet werden. Auf die Fragen, wie die Messung der Zuverlässigkeit des Datenverarbeitungssystems erfolgt und wie die Zusammenhänge mit dem Fehleranteil abgeleitet werden können, soll an dieser Stelle nicht eingegangen werden, sie werden im nächsten Kapitel behandelt. Fraglich ist darüber hinaus, welcher Implikations- und Konjunktionsoperator zur Risikomessung zu verwenden wäre. Dies führt zu einem Problem, welches die Risikomessung bei bewußter Auswahl der Stichprobenelemente generell betrifft.

Wie oben angesprochen wurde, kann die Wahl bestimmter Operatoren für die Verknüpfung von unscharfen Mengen regelmäßig nicht theoretisch begründet werden, sondern wird mit Blick auf die Brauchbarkeit der Ergebnisse, die mit den einzelnen Verknüpfungsoperatoren erzielt wurden, begründet. Insoweit besteht insbesondere für die zuletzt angeführte Alternative zur Risikomessung die Notwendigkeit, die Ergebnisse, die mit verschiedenen Implikations- bzw. Konjunktionsoperatoren erzielt wurden, mit der „richtigen Schlußfolgerung“ im Hinblick auf die beiden Mengen D und D' zu vergleichen. Diese Möglichkeit besteht allerdings in realen Prüfungssituationen regelmäßig nicht, da dort auch ex post die für die Beurteilung der Prüfungsgesamtheit relevante Menge D nicht bekannt wird. Bevor demnach eine Anwendung von Schlußfolgerungen in der Form von (3.35) möglich wird, muß die Verwendung bestimmter Verknüpfungsoperatoren entweder theoretisch oder empirisch begründet werden.

Der Umstand, daß die Fehler, die bei der Verwendung bestimmter Auswahlkriterien gemacht werden, im Rahmen realer Prüfungen regelmäßig nicht bekannt werden, führt auch zu Problemen bei der Ableitung objektivierter Wahrscheinlichkeitszuordnungen für die im einzelnen zu treffenden Aussagen. Auch für die Verwendung subjektiver Wahrscheinlichkeiten ergeben sich Bedenken, da zumindest ein Lernen aus Erfahrung nicht stattfinden kann[70]. Insoweit be-

70 Siehe oben Kapitel B. I. 2. b) im Zweiten Teil der Arbeit.

dürfen die im Rahmen der Majorprämisse verwendeten Gesetzmäßigkeiten ebenfalls noch einer theoretischen oder zumindest einer gesicherten empirischen Begründung. Allerdings erscheint für den Fall, daß überhaupt eine quantitative Risikomessung erfolgen soll, die Angabe von Wahrscheinlichkeiten zu den einzelnen Voraussetzungen und die Offenlegung der verwendeten Voraussetzungen selbst besser geeignet, das ermittelte Gesamtrisikomaß im Hinblick auf seine Plausibilität zu überprüfen, als eine direkte Abschätzung des Gesamtrisikos für einzelne Ergebnisse von Prüfungshandlungen.

B. Risikobeiträge indirekter Messungen

Wie schon bei der Darstellung des Prüfungsmodells deutlich wurde[71], sind die Ergebnisse, welche die indirekten Messungen im Rahmen der Prüfung liefern können, regelmäßig nicht ausreichend, um für sich allein die Beurteilung der Prüfungsgesamtheit zu ermöglichen, wenn davon abgesehen wird, daß die Ablehnung der Ordnungsmäßigkeit unter Umständen schon nach der Durchführung von analytischen Prüfungshandlungen möglich sein kann. Allerdings werden die Ergebnisse der Prüfung des datenerzeugenden Systems im Rahmen der dedektiven Auswahl von Stichprobenelementen benötigt[72]. Die Ergebnisse analytischer Prüfungshandlungen können unter bestimmten Voraussetzungen die Aussagefähigkeit von mathematisch-statistischen Testverfahren verbessern[73]. Insoweit soll im folgenden untersucht werden, welche grundsätzlichen Möglichkeiten bestehen, die bei den Prüfungsergebnissen aus den indirekten Messungen auftretende Unschärfe abzuschätzen.

I. Prüfung des datenerzeugenden Systems

Neben der Verwendung der Ergebnisse einer Prüfung des datenerzeugenden Systems bei der bewußten Auswahl von Stichprobenelementen, die voraussetzt, daß aus der Systembeurteilung auf die Fehlerhaftigkeit der Prüfungsgesamtheit geschlossen werden kann, müssen durch die Systemprüfung auch Aussagen über die Vertrauenswürdigkeit der im Unternehmen vorliegenden Dokumentation der abgebildeten realen Sachverhalte gewonnen werden können[74], soweit die Ableitung von Sollobjekten im Rahmen der Prüfung auf diese internen

---

[71] Siehe oben Kapitel B. II im Ersten Teil der Arbeit.

[72] Siehe oben Kapitel A. II. 2. in diesem Teil der Arbeit.

[73] Vgl. insbesondere oben Kapitel C. II. 2. im Zweiten Teil der Arbeit.

[74] Vgl. *Konrath*, Auditing, 1993, S. 117.

Unterlagen des Unternehmens angewiesen ist. Die beiden Bereiche sollen im folgenden getrennt untersucht werden.

1. Aussagen über Fehler in der Prüfungsgesamtheit

Die der Auswertung von Ergebnissen der Systemprüfung im Hinblick auf die Fehlerhaftigkeit der Prüfungsgesamtheit zugrundeliegende Schlußfolgerung wurde schon in (1.35) dargestellt. Um sowohl Aussagen über den Fehleranteil als auch über das Fehlerausmaß ableiten zu können, muß neben einer Beschreibung des Systemzustandes auch der Systeminput, d. h. im vorliegenden Prüfungsmodell U', bekannt sein. Die zuletzt genannte Voraussetzung wird in realen Prüfungssituationen regelmäßig nicht erfüllt werden können, da ansonsten zumindest theoretisch für den Prüfer die Möglichkeit bestünde, das Sollobjekt Y direkt aus U' durch die eigene Verarbeitung der Daten abzuleiten. Im folgenden soll davon ausgegangen werden, daß zumindest die Abschätzung der Fehlerhäufigkeit eines datenerzeugenden Systems auch ohne die genaue Kenntnis von U' möglich ist. Zur Abschätzung des Fehlerausmaßes müßten allerdings neben den Informationen über den Systeminput auch Informationen über die Höhe der einzelnen Fehler, die im Rahmen der Verarbeitungsfunktionen auftreten können, vorliegen. Diese Informationen werden regelmäßig erst im Rahmen der Einzelfallprüfung gewonnen. Insoweit wird der Prüfer zumindest im Hinblick auf das Fehlerausmaß nur sehr eingeschränkte Aussagen aus der Systemprüfung ableiten können. Die folgenden Ausführungen beschränken sich daher auf die Einflüsse der aus der Systemprüfung gewonnenen Ergebnisse auf die Risikomessung bei der Beurteilung des Fehleranteils in der Prüfungsgesamtheit.

Wie schon angeführt wurde, stehen für die Ableitung der in (1.35) verwendeten Majorprämisse grundsätzlich zwei Ansätze, namentlich die Analyse eines Modells des datenerzeugenden Systems[75] und die ordinale Bewertung der Systemzuverlässigkeit, zur Verfügung. Diese werden zunächst untersucht, bevor geprüft wird, welche Probleme bei direkter Abschätzung des Fehleranteils aus den Ergebnissen der Systemprüfung durch den Prüfer auftreten.

---

[75] Zu den analytischen Ansätzen werden hier insbesondere der auf Markovketten basierende Ansatz von *Yu* und der auf der Zuverlässigkeitstheorie aufbauende Ansatz von *Cushing* gezählt. Vgl. *Yu*, Model, 1972; *Cushing*, Mathematical Approach, 1974, S. 24-41. Die Ansätze sollen wegen der bei der Informationsbeschaffung auftretenden Probleme hier nicht dargestellt werden. Siehe oben Kapitel B. II. 1. im Ersten Teil der Arbeit.

a) Analyse des Systemverhaltens

Den Ansätzen, die Aussagen über die Zuverlässigkeit des datenerzeugenden Systems anhand der Analyse von Modellen ableiten wollen, ist gemeinsam, daß bestimmte Wahrscheinlichkeiten für das Verhalten einzelner Verarbeitungs- und Kontrollfunktionen als Modellinput verfügbar sein müssen, diese werden dann in eine Wahrscheinlichkeitsaussage über den Fehleranteil im Systemoutput des datenerzeugenden Systems transformiert. Dabei treten neben den Problemen bei der Beschaffung der benötigten Informationen für die Modellierung eine Reihe von Problemen im Hinblick auf die Interpretation der als Modellinput verwendeten Wahrscheinlichkeiten auf.

Eine objektive Interpretation der Wahrscheinlichkeiten wird regelmäßig nicht möglich sein, da dies voraussetzen würde, daß es sich bei den Verarbeitungs- und Kontrollprozessen um Versuchsanordnungen handelt, denen die Fehlerwahrscheinlichkeit als theoretische Eigenschaft zugerechnet werden kann. Darüber hinaus ist auch bei subjektiver Interpretation der Wahrscheinlichkeiten fraglich, ob diese ein geeignetes Unschärfemaß für die Fehlerverteilung bei den betrachteten Datenverarbeitungsfunktionen darstellen, da zumindest die Unschärfe bei den von maschinellen Aufgabenträgern ausgeführten Funktionen aus der fehlenden Kenntnis über die tatsächlich ausgeführten Funktionen resultiert, nicht aus dem Verhalten der Aufgabenträger[76]. Sollen subjektive Wahrscheinlichkeiten als Unschärfemaß in diesen Modellen verwendet werden, so müßte die Abgrenzung der auf unterster Ebene im Modell betrachteten Funktionen generell in der Weise erfolgen, daß die Funktionen von maschinellen Aufgabenträgern mit den Funktionen der personellen Aufgabenträger an den Eingabeschnittstellen zusammengefaßt werden. Damit würde erreicht, daß die Fehler der betrachteten Funktion zufällig entstehen. Fraglich ist bei dieser Vorgehensweise allerdings, wie die Ergebnisse aus einer Prüfung des datenerzeugenden Systems bei der Ermittlung der Fehlerwahrscheinlichkeiten zu berücksichtigen wären. Insoweit sind die Ansätze, die anhand der Analyse des Systemverhaltens mit Hilfe von Modellen versuchen zu Aussagen über die Fehlerverteilung in der Prüfungsgesamtheit zu kommen, nicht in der Lage, die auftretenden Formen von Unschärfe adäquat abzubilden.

---

[76] Vgl. auch *Cooley/Hicks*, Fuzzy Set Approach, 1983, S. 318.

DOI https://doi.org/10.3790/978-3-428-48909-1

b) Ableitung aus der Zuverlässigkeit des datenerzeugenden Systems

Soll im Rahmen der Stichprobenverfahren mit bewußter Auswahl die in (3.35) dargestellte Schlußform verwendet werden, muß als Majorprämisse eine allgemein gültige Gesetzmäßigkeit zur Verfügung stehen, die einen Zusammenhang zwischen der Zuverlässigkeit des datenerzeugenden Systems und dem Fehleranteil in der erzeugten Prüfungsgesamtheit herstellt. Die Frage, wie die Zuverlässigkeit des Systems gemessen werden soll, und welches Skalenniveau das resultierende Maß hat, ist allerdings noch offen. Dabei kann der tatsächliche bzw. erwartete Fehleranteil im Systemoutput nicht zur Messung der Zuverlässigkeit verwendet werden, da dieser gerade aus dem ermittelten Maß für das vorliegende datenerzeugende System abgeleitet werden soll. Wäre der Fehleranteil bekannt, würde sich die Anwendung der unterstellten Gesetzmäßigkeit erübrigen. Für das Skalenniveau des Zuverlässigkeitsmaßes soll unterstellt werden, daß die Messungen zumindest ordinal erfolgen können.

Die Messung der Zuverlässigkeit setzt zunächst voraus, daß das Istobjekt der Systemprüfung, also das tatsächlich im Unternehmen ausgeführte Verarbeitungs- und Kontrollsystem bekannt ist. Bei der Systemerfassung auftretende Unschärfen sollen zunächst unberücksichtigt bleiben. Unabhängig davon, welche Systemeigenschaften zur Messung der Zuverlässigkeit herangezogen werden, ergibt sich als Konsequenz für die Ableitung des Sollobjektes der Prüfung, daß der Prüfer für die Systemeigenschaften, die zur Ableitung des Zuverlässigkeitsmaßes herangezogen werden sollen, angeben kann, unter welchen Bedingungen diese zur maximalen Zuverlässigkeit des betrachteten datenerzeugenden Systems führen. Insoweit müssen auch einige dieser Eigenschaften zumindest ordinal skaliert sein. Die Systemeigenschaften, soweit sie hier für die Messung der Zuverlässigkeit relevant sind, lassen sich generell drei unterschiedlichen Kategorien zuordnen.

Zum einen kann als Eigenschaft eines Systems die Einhaltung bestimmter Grundsätze, die für die Gestaltung des Systems gelten sollen, wie z. B. der Grundsatz der Organisation von Arbeitsabläufen[77] oder die Verwendung angemessener Methoden bei der Systemerstellung, insbesondere bei integrierten EDV-Systemen[78], gesehen werden. Die angeführten Eigenschaften sind regelmäßig nicht ordinal meßbar. Dieser Umstand führt auch dazu, daß sich die Auswirkungen auf die Zuverlässigkeit des Systems für den Fall, daß die angesprochenen Grundsätze nicht beachtet werden, nur schwer abschätzen lassen.

---

[77] Vgl. *IDW*, WP-Handbuch, 1996, P Rz. 142.
[78] Vgl. *Göbel*, Prüfung, 1990, S. 46-63.

Eine zweite Kategorie betrifft Eigenschaften der tatsächlich vorliegenden Struktur des datenerzeugenden Systems, wie z. B. die Kontrolldichte oder die Einhaltung von Funktionstrennungsgrundsätzen[79]. Diese dürften zumindest einer ordinalen Messung zugänglich sein. Als letzte Kategorie von Eigenschaften wären schließlich solche zu nennen, die sich auf die Funktionsweise einzelner Verarbeitungs- oder Kontrollfunktionen beziehen.

Ansätze zur Prüfung des datenerzeugenden Systems, die es erlauben, ein Maß für die Zuverlässigkeit unter Berücksichtigung von Eigenschaften aus allen genannten Kategorien zu bestimmen, existieren bisher nicht. Allerdings existiert ein Vorschlag von *Cooley/Hicks*[80] für die Aggregation der Zuverlässigkeit einzelner Elemente des Kontrollsystems. Das betrachtete Kontrollsystem wird dabei durch die Verknüpfung von drei unterschiedlichen Komponenten abgebildet. Diese sind das Kontrollziel, die möglichen Risiken, auf die sich die Kontrolle bezieht, und der eigentliche Kontrollprozeß. Jede der Komponenten hat genau eine für die Betrachtung relevante Eigenschaft. Für das Kontrollziel wird dessen Wichtigkeit, für die Risiken deren Eintrittswahrscheinlichkeit und für den Kontrollprozeß seine Zuverlässigkeit betrachtet. Bei der Abbildung eines Kontrollsystems werden alle Kontrollziele sowie die zugehörigen Kontrollprozesse und Risiken als voneinander unabhängig betrachtet. Die einzige Möglichkeit, Interdependenzen zwischen Kontrollen abzubilden, besteht darin, diese zusammen als eine einzige Kontrolle zu betrachten. Zur Messung der Eigenschaften aller Elemente wird eine linguistischen Variable verwendet, deren Ausprägungen aus einer Grundmenge mit den Elementen {„stark“, „mittel“, „schwach“} stammen. Darüber hinaus wird eine Modifikation der Ausdrücke durch den Begriff „sehr“ definiert[81], so daß neben den Elementen der Grundmenge noch „sehr schwach“ und „sehr stark“ als mögliche Ausprägungen zugelassen werden. Als Grundbereich für die Ausprägungen der linguistischen Variable wird eine Skala von Eins bis Neun gewählt. Nach der Bewertung aller Komponenten eines betrachteten Kontrollsystems werden die Zuverlässigkeitsgrade der einzelnen Kontrollprozesse durch einen geeigneten Operator aggregiert. Hierzu können unterschiedliche Operatoren verwendet werden, je nachdem, ob die Eintrittswahrscheinlichkeiten der Risiken und die Wichtigkeit des Kontrollziels als Gewichte in die Betrachtung einbezogen werden sollen oder nicht. Eine Darstellung der im einzelnen untersuchten Aggre-

79 Vgl. *Pougin*, Kontrollsystem, 1959, S. 50f.

80 Vgl. zu der folgenden Darstellung *Cooley/Hicks*, Fuzzy Set Approach, 1983, S. 319-328.

81 Zur Modifikation wird der in Abb. 2.3 dargestellte Algorithmus verwendet.

gationsoperatoren soll unterbleiben, da die bisherigen Ausführungen zur Diskussion der grundlegenden Probleme eines solchen Ansatzes ausreichen.

Zunächst ist festzustellen, daß die dargestellte Vorgehensweise als Ergebnis ein Maß für die Zuverlässigkeit des betrachteten Kontrollsystems liefert, das unabhängig vom Fehleranteil im Systemoutput ermittelt werden kann. Insoweit könnte das in dieser Weise ermittelte Zuverlässigkeitsmaß im Rahmen einer noch abzuleitenden Gesetzmäßigkeit, die den Zusammenhang zwischen einzelnen Zuverlässigkeitsgraden eines Systems und dem Fehleranteil im Systemoutput beschreibt, verwendet werden. Der Ansatz läßt allerdings offen, was unter der Zuverlässigkeit selbst zu verstehen ist, da er ausschließlich die Zuverlässigkeitsgrade einzelner Kontrollprozesse aggregiert. Dieser Mangel wird auch in der Wahl des Grundbereichs für die linguistische Variable deutlich, da z. B. eine Interpretation der Aussage: „Die Zuverlässigkeit des Kontrollprozesses ist Neun" nicht möglich ist. Voraussetzung für die Verwendung von linguistischen Variablen ist aber, daß die zu bewertende Eigenschaft zumindest grundsätzlich auf der zugrundegelegten Skala gemessen werden könnte, den Meßwerten also auch eine Bedeutung zugeordnet werden kann. Darüber hinaus sollte, damit aus den gemessenen Zuverlässigkeitsgraden eine allgemeingültige Gesetzmäßigkeit abgeleitet werden kann, die Bedeutung, die einzelnen Werten auf der verwendeten Skala zugeordnet wird, für unterschiedliche Prüfer weitgehend identisch sein.

Ist dies nicht der Fall, ergeben sich gegenüber der direkten Angabe eines Zuverlässigkeitsgrades für das betrachtete Gesamtsystem durch den Prüfer nur geringfügige Verbesserungen im Hinblick auf die Aussagekraft der ermittelten Größe. Zwar wird einerseits durch die Offenlegung des zur Aggregation der einzelnen Zuverlässigkeitsgrade verwendeten Verfahrens die intersubjektive Überprüfbarkeit des Ergebnisses verbessert, auf der anderen Seite kann aber, solange nicht deutlich wird, was im einzelnen gemessen wird, die Wahl eines bestimmten Aggregationsoperators weder theoretisch noch empirisch begründet werden. Dieser Nachteil dürfte den genannten Vorteil annähernd aufheben. Darüber hinaus besteht zumindest bei einem der untersuchten Aggregationsoperatoren das Problem, daß die Unschärfe des Ergebnisses regelmäßig größer ist als die Unschärfe der zugrundeliegenden Messungen[82]. Es kann demnach bei großen Systemen und entsprechend großer Anzahl von Elementen die Situation entstehen, daß das ermittelte Ergebnis zu unscharf und damit nicht ver-

82 Vgl. *Cooley/Hicks*, Fuzzy Set Approach, 1983, S. 329.

wertbar ist. Das Problem ist allerdings nicht spezifisch für den hier vorgestellten Ansatz.

Neben den angesprochenen grundlegenden Problemen fehlt in dem angeführten Ansatz die Möglichkeit, Auswirkungen, die sich aus unterschiedlichen Anordnungen von Kontrollen ergeben können, in die Messung einzubeziehen. Außerdem können unterschiedliche Eigenschaften der Verarbeitungsfunktion ausschließlich über die Wichtigkeit des Kontrollziels berücksichtigt werden. Der Ansatz erfüllt demnach die für die Ableitung einer allgemeinen Gesetzmäßigkeit zwischen Zuverlässigkeit des datenerzeugenden Systems und Fehleranteil in der Prüfungsgesamtheit erforderlichen Anforderungen nicht. Insgesamt ist demnach festzustellen, daß eine indirekte Messung der Fehler in der Prüfungsgesamtheit über die Prüfung des datenerzeugenden Systems[83] scheitert, solange die Zuverlässigkeit bzw. jede andere Systemeigenschaft, für die ein Zusammenhang mit der Fehlerhaftigkeit der Prüfungsgesamtheit konstruiert werden soll, nicht intersubjektiv überprüfbar und unabhängig vom Fehleranteil selbst gemessen werden kann. Da bisher keine Prüfungsmethoden für die Systemprüfung verfügbar sind, die eine indirekte Messung des Fehleranteils zulassen, bleibt ausschließlich eine direkte Abschätzung des Fehleranteils. Fraglich ist, wie die bei dieser Abschätzung bestehende Unschärfe gemessen werden kann.

c) Direkte Abschätzung des Fehleranteils

Für die direkte Abschätzung des in der Prüfungsgesamtheit vermuteten Fehleranteils bestehen grundsätzlich drei Alternativen, namentlich die Angabe eines konkreten Fehleranteils auf der zugrundegelegten Skala zwischen Null und Eins, die Verwendung einer linguistischen Variable FEHLERANTEIL oder die Angabe, ob die Prüfungsgesamtheit im Hinblick auf den Fehleranteil als ordnungsmäßig oder nichtordnungsmäßig einzustufen ist. Insoweit kann, soweit Ungewißheit hinsichtlich des genauen Wertes für den Fehleranteil besteht, die dabei auftretende Unschärfe durch die Verwendung einer linguistischen Variable, wie sie z. B. in Abb. 2.2 dargestellt wurde, abgebildet werden. Darüber hinaus kann unabhängig von der Form, in der die Abschätzung des Fehleranteils erfolgt, die getroffene Aussage selbst unsicher sein.

Dabei kommen für die Unsicherheit insbesondere zwei Ursachen in Betracht. Zum einen kann Unsicherheit darüber bestehen, ob das Istobjekt der

[83] A. A. *v. Wysocki*, Grundlagen, 1988, S. 165; *Hagest*, Logik, 1975, S. 107f.

Prüfung, also das tatsächlich im Unternehmen eingesetzte datenerzeugende System, richtig erfaßt wurde. Zum anderen kann Unsicherheit darüber bestehen, welche genaue Wirkung auf den Fehleranteil von den festgestellten Fehlern und Schwachstellen im betrachteten System ausgeht. Demnach ist zu prüfen, wie sich die Möglichkeit, daß sich der Prüfer bei der Systembeurteilung irrt, auf die von ihm getroffene Aussage auswirkt. Es soll davon ausgegangen werden, daß er die Wahrscheinlichkeit für eine falsche Systembeurteilung subjektiv mit α angeben kann. Für die getroffene Aussage ergibt sich dann die Glaubwürdigkeit mit $1-\alpha$. Die Zuordnung einer Wahrscheinlichkeit zu der getroffenen Aussage kann nicht erfolgen, da aus einem Irrtum bei der Systembeurteilung nicht zwingend gefolgert werden kann, daß die getroffene Aussage falsch ist. Das ermittelte Glaubwürdigkeitsmaß ist wie die zugrundegelegte Wahrscheinlichkeitsbeurteilung rein subjektiv, da, wie oben gezeigt wurde, bisher keine Systemprüfungsmethoden zur Verfügung stehen, die es erlauben, objektivierte Größen abzuleiten.

2. *Vertrauenswürdigkeit der Dokumentation*

Zur Ableitung der einzelnen Sollobjekte wird gemäß (1.8) bis (1.10) grundsätzlich vorausgesetzt, daß der Prüfer die realen Sachverhalte tatsächlich kennt. Zumindest für die Jahresabschlußprüfung ist diese Voraussetzung regelmäßig nicht für alle abgebildeten Sachverhalte erfüllt. Insoweit ist der Prüfer bei der Bestimmung seiner Sollobjekte auf unterschiedlich vertrauenswürdige Informationsquellen angewiesen. Soweit die Informationen vom Prüfer selbst durch Beobachtungen oder von Unternehmensexternen ermittelt werden, muß die Vertrauenswürdigkeit der verwendeten Informationen durch den Prüfer direkt beurteilt werden. Soweit die durch das datenerzeugende System gespeicherte Dokumentation Informationsquelle für die Ableitung von Sollobjekten ist, soll im folgenden geprüft werden, inwieweit Ergebnisse der Systemprüfung zu Aussagen über die Vertrauenswürdigkeit der Dokumentation herangezogen werden können. Im vorliegenden Modell besteht die Dokumentation, die dem Prüfer als Informationsquelle dienen kann, aus den in X enthaltenen Merkmalen, die aus Messungen tatsächlicher Sachverhalte resultieren. Es soll gelten, daß die Vertrauenswürdigkeit dieser Dokumentation dann uneingeschränkt gewährleistet ist, wenn sie richtig und vollständig ist.

Zur Ableitung eines Maßes der Vertrauenswürdigkeit müssen zwei Problembereiche unterschieden werden. Zum einen ist der relevante Teil des datenerzeugenden Systems abzugrenzen. Zum anderen ist festzulegen, wie aus der Systembeurteilung das Maß für die Vertrauenswürdigkeit abzuleiten ist. Werden die Komponenten des datenerzeugenden Systems, die ihren Input nicht oder nicht ausschließlich von anderen Komponenten des betrachteten Systems erhalten, sowie die zu diesen Komponenten existierenden Kontrollfunktionen als Erfassungssystem bezeichnet, dann ist ausschließlich die Zuverlässigkeit dieses Erfassungssystems für die Vertrauenswürdigkeit der Dokumentation verantwortlich. Dabei nimmt die Vertrauenswürdigkeit mit zunehmender Zuverlässigkeit zu und mit abnehmender Zuverlässigkeit ab. Notwendig ist daher zumindest eine ordinale Messung der Zuverlässigkeit des Erfassungssystems. Soll allerdings das Maß für die ermittelte Vertrauenswürdigkeit im Sinne eines Glaubwürdigkeitsmaßes, das die Überzeugung des Prüfers für die Korrektheit seines Urteils im Hinblick auf die zugrundegelegten Informationen ausdrückt, verwendet werden, muß auch das Zuverlässigkeitsmaß ein Glaubwürdigkeitsmaß sein[84].

Für die Messung der Zuverlässigkeit treten, auch wenn ausschließlich ein Teilsystem beurteilt werden soll, grundsätzlich die gleichen Probleme auf, die auch bei der Beurteilung der Zuverlässigkeit des gesamten datenerzeugenden Systems bestehen[85]. Allerdings stehen zumindest im Rahmen der Jahresabschlußprüfung für die Beurteilung der Zuverlässigkeit des Erfassungssystems einzelne Beurteilungskriterien zur Verfügung, die aus den in X enthaltenen Merkmalen abgeleitet werden können. Wie schon oben angeführt wurde[86], kommt insbesondere die Zeitgerechtigkeit der Erfassung als Merkmal für die Zuverlässigkeit des Erfassungssystems und damit der Vertrauenswürdigkeit der Dokumentation in Betracht. Die Zeitgerechtigkeit der Erfassung könnte dabei sowohl anhand des Anteils von Elementen in X, die nicht zeitgerecht erfaßt wurden, als auch anhand der durchschnittlichen Zeitspanne, die zwischen einem Ereignis und seiner Erfassung liegt, gemessen werden. Insoweit ist die Situation gegenüber der Beurteilung des Gesamtsystems verbessert, da die Messung der Zuverlässigkeit des Erfassungssystems anhand eindeutig meßbarer Kriterien erfolgen könnte.

---

84 Die Möglichkeit, aus ordinal skalierten Werten für die Zuverlässigkeit des Erfassungssystems auf die Werte einer linguistischen Variable für die Vertrauenswürdigkeit der Dokumentation zu schließen, soll zunächst ausgeschlossen werden, da weder eine theoretische noch eine empirische Begründung für die Handhabung des entstehenden Modells der Risikomessung möglich erscheint. Vgl. *Bandemer/Gottwald*, Einführung, 1993, S. 138f.

85 Siehe oben Kapitel B. I. 1. b) in diesem Teil der Arbeit.

86 Siehe oben Kapitel A. II. im Ersten Teil der Arbeit.

DOI https://doi.org/10.3790/978-3-428-48909-1

Durch das Abstellen auf die Zeitgerechtigkeit der Erfassung werden allerdings nicht alle Probleme der Messung der Zuverlässigkeit behoben. Zum einen folgt aus einer zeitgerechten Erfassung noch nicht die Vollständigkeit der erfaßten Sachverhalte. Zum anderen ist sie zwar notwendig, aber nicht hinreichend für eine vertrauenswürdige Dokumentation. Die Vertrauenswürdigkeit der betrachteten Dokumentation wird darüber hinaus von den integrierten Kontrollen und der Anordnung der Erfassungsfunktionen beeinflußt[87]. Insoweit wird auch zur Beurteilung der Vertrauenswürdigkeit der verwendeten Dokumentation ausschließlich eine direkte Zuordnung eines Glaubwürdigkeitsmaßes in Betracht kommen, bei dessen Ableitung zwar die aus der Prüfung des Erfassungssystems erhaltenen Ergebnisse durch den Prüfer verwendet werden, die Ableitung selbst aber nicht intersubjektiv überprüfbar ist.

II. Analytische Prüfungshandlungen

Grundlage für die Verwendung der aus analytischen Prüfungshandlungen stammenden Ergebnisse für die Urteilsbildung ist der in (1.36) formulierte Zusammenhang zwischen der Ordnungsmäßigkeit der Prüfungsgesamtheit und bestimmten Informationen über reale Sachverhalte auf der einen Seite und einer Prognose im Hinblick auf die daraus zu erwartenden Ergebnisse der Prüfungshandlungen auf der anderen Seite. Da alle Verfahren auf wertmäßige Zusammenhänge abstellen, wird ausschließlich die Ordnungsmäßigkeit im Hinblick auf das Fehlerausmaß betrachtet. Der unterstellte Zusammenhang läßt sich unter Verwendung von Wahrscheinlichkeiten auch unscharf formulieren. Wird das Ereignis, daß die Prognose für die verwendete Kenngröße zutrifft, mit E bezeichnet und wird unterstellt, daß die Irrtumswahrscheinlichkeit der Gesetzmäßigkeit, die der Prognose zugrundegelegt wird, α beträgt, dann gilt:

$$(3.36) \qquad pr(E|O \cap U) = 1 - \alpha \text{ und}$$

$$(3.37) \qquad pr(\overline{E}|O \cap U) = \alpha .$$

Bevor die möglichen Ansätze zur Risikomessung untersucht werden, soll geprüft werden, welche Alternativen zur Bestimmung dieser Wahrscheinlichkeiten verfügbar sind, und welche Probleme bei der Abgrenzung von E auftreten. Die Bestimmung der benötigten Wahrscheinlichkeit ist abhängig davon, welches der am Anfang der Arbeit angeführten Modelle zugrundegelegt wird[88]

87 Vgl. *Wittmann*, Systemprüfung, 1980, S. 130f.

88 Siehe oben Kapitel B. II. 2. im Ersten Teil der Arbeit.

und welche Prognosemethoden für das gewählte Modell zur Verfügung stehen bzw. angewandt werden. Sieht man von den Modellen ab, die eine streng funktionale Abhängigkeit unterstellen, da hier keine Unschärfe auftritt, kommt eine Risikomessung insbesondere bei der Verwendung von Trendmodellen oder von Erklärungsmodellen in Betracht. Dabei existieren in beiden Fällen Verfahren, die eine Intervallprognose und damit auch die Angabe eines Konfidenzniveaus erlauben. Damit kann zumindest die dem Prognoseverfahren immanente Unschärfe objektiv ermittelt werden. Allerdings besteht bei diesen Verfahren auch Unschärfe, die aus der Ungewißheit über das Vorliegen der jeweiligen Voraussetzungen, die an die Anwendung der Verfahren gestellt werden, resultiert[89]. Diese Unschärfe kann ausschließlich durch eine subjektive Wahrscheinlichkeitsbeurteilung erfaßt werden. Für die übrigen Prognoseverfahren muß die benötigte Wahrscheinlichkeit insgesamt subjektiv festgelegt werden. Dabei hängt unabhängig vom verwendeten Prognoseverfahren die konkrete Festlegung der Wahrscheinlichkeit auch mit der verwendeten Definition für E zusammen.

Bisher wurde zur Abgrenzung von E ausschließlich festgelegt, daß es aus der Menge der Ereignisse besteht, bei denen die abgeleitete Prognose zutrifft. Wie insbesondere bei den Prognoseverfahren, die ein Prognoseintervall liefern, deutlich wird, muß nicht jede Abweichung zwischen einem einzelnen prognostizierten Wert und dem tatsächlich ermittelten Wert bedeuten, daß die der Prognose zugrundegelegten Annahmen falsch sind. Dabei wird für die meisten verwendeten Gesetzmäßigkeiten gelten, daß c. p. die Irrtumswahrscheinlichkeit mit größerem E sinkt. Bevor jedoch die Auswirkungen auf die Aussagekraft der Ergebnisse untersucht werden können, muß geprüft werden, welche Schlußfolgerungen für den hier unscharf formulierten Zusammenhang zwischen Ordnungsmäßigkeit der Prüfungsgesamtheit und Prognosewert ausreichend begründet sind.

Für die Beurteilung der Prüfungsgesamtheit gilt, daß, auch wenn α genügend klein ist, die Ordnungsmäßigkeit der Prüfungsgesamtheit nicht schon aus dem Grund angenommen werden darf, weil E eingetreten ist, da dieses Ergebnis bei Gültigkeit der Hypothese $O \cap U$ zu erwarten war. Falls $\overline{E}$ eintritt, es sich also nach Durchführung der Prüfungshandlungen herausstellt, daß die Prognose nicht zutrifft, kann dies darauf zurückzuführen sein, daß die Informationen über die realen Sachverhalte nicht zutreffen. Allerdings darf, auch wenn letzteres ausgeschlossen werden kann, die Ordnungsmäßigkeit der Prüfungsgesamtheit nicht ohne weiteres abgelehnt werden, da für das tatsächlich eingetretene

89 Vgl. *Kinney/Haynes*, Analytical Procedure, 1990, S. 94.

Ereignis eine Wahrscheinlichkeit in Höhe von α verbleibt. Insoweit besteht kein logischer Widerspruch zwischen dem ermittelten Ergebnis und der Ordnungsmäßigkeit der Prüfungsgesamtheit. Eine Ablehnung der angeführten Hypothese $O \cap U$ kann erst dann erfolgen, wenn eine andere Hypothese zu Verfügung steht, die das eingetretene Ergebnis $\overline{E}$ sehr viel besser erklärt. Es muß demnach, bevor aus den Ergebnissen analytischer Prüfungshandlungen eine Entscheidung über die Ordnungsmäßigkeit oder Nichtordnungsmäßigkeit der Prüfungsgesamtheit getroffen werden kann, eine der beiden folgenden bedingten Wahrscheinlichkeiten bekannt sein:

(3.38) $$pr\left(\overline{E} \middle| \overline{O} \cap U\right) = 1 - \beta \text{ und}$$

(3.39) $$pr\left(E \middle| \overline{O} \cap U\right) = \beta.$$

Wird weiterhin unterstellt, daß $\overline{E}$ eingetreten ist, muß, damit die Ordnungsmäßigkeit der Prüfungsgesamtheit abgelehnt werden kann, $1-\beta$ ausreichend groß sein. Entsprechend gilt für den Fall des Eintretens von E, daß β ausreichend klein sein muß. Es kann somit analog der Vorgehensweise bei den mathematisch-statistischen Testverfahren ein Risikomaß γ abgeleitet werden, für das je nachdem, welches Ergebnis im Rahmen der analytischen Prüfungshandlungen ermittelt wird, gilt:

(3.40) $$\gamma = \frac{pr\left(\overline{E} \middle| O \cap U\right)}{pr\left(\overline{E} \middle| \overline{O} \cap U\right)} = \frac{\alpha}{1-\beta} \text{ oder}$$

(3.41) $$\gamma = \frac{pr\left(E \middle| \overline{O} \cap U\right)}{pr\left(E \middle| O \cap U\right)} = \frac{\beta}{1-\alpha}.$$

Dabei mißt der nach (3.41) ermittelte Wert das Risiko bei der Wahl von Handlungsalternative a_1, also der Annahme der Ordnungsmäßigkeit, der nach (3.40) ermittelte Wert das Risiko bei der Ablehnung der Ordnungsmäßigkeit der Prüfungsgesamtheit, also bei Wahl der Handlungsalternative a_2. Die Aussagekraft der aus dem Ergebnis der analytischen Prüfungshandlungen abgeleiteten Schlußfolgerungen ist um so größer je kleiner γ ist. Die Auswirkungen auf die Aussagekraft des Ergebnisses die aus einer Erweiterung von E resultieren, können nicht generell abgeschätzt werden. Steigt allerdings β in gleichem Maß, in dem die Irrtumswahrscheinlichkeit sinkt, dann wird γ größer, und damit die Aussagekraft geringer.

Es erscheint allerdings mehr als fraglich, ob die der Prognose zugrundeliegende Gesetzmäßigkeit geeignet ist, um die Wahrscheinlichkeit für den prognostizierten Wert unter der Voraussetzung abzuleiten, daß die Prüfungsgesamtheit tatsächlich nicht ordnungsmäßig ist, da die verwendete Gesetzmäßigkeit regelmäßig einen funktionalen oder doch zumindest kausalen Zusammenhang zwischen der Ordnungsmäßigkeit und E herstellt. Damit kann aber über den Eintritt von E, wenn gänzlich andere Ausgangsgrößen vorliegen, nichts ausgesagt werden. Darüber hinaus dürfte regelmäßig auch keine eigenständige Gesetzmäßigkeit bekannt sein, aus der sich die Wahrscheinlichkeit von E für den Fall einer nicht ordnungsmäßigen Prüfungsgesamtheit ableiten läßt, da die Eigenschaft der Nichtordnungsmäßigkeit einer Prüfungsgesamtheit sehr viel diffusere Sachverhalte umfaßt als die Eigenschaft der Ordnungsmäßigkeit. Insoweit läßt sich β ausschließlich durch subjektive Schätzung festlegen. Allerdings dürfte bei der Verwendung globaler Kennzahlen, die es erlauben, daß sich einzelne Fehler ausgleichen, β nicht viel kleiner sein als $1-\alpha$. Damit ist aber die Aussagekraft der Ergebnisse aus analytischen Prüfungshandlungen regelmäßig sehr gering[90].

Fraglich ist, ob aus dem Ergebnis analytischer Prüfungshandlungen Glaubwürdigkeitseinschätzungen für die Ordnungsmäßigkeit oder Nichtordnungsmäßigkeit der Prüfungsgesamtheit abgeleitet werden können[91]. Da es die zugrundeliegende Gesetzmäßigkeit im Unterschied zu den mathematisch-statistischen Verfahren regelmäßig nicht zulassen wird, aus dem ermittelten Ergebnis oder aus der Abweichung zwischen prognostiziertem und tatsächlichem Wert eine Obergrenze für das Fehlerausmaß in der Prüfungsgesamtheit abzuleiten[92], besteht auch keine Möglichkeit aus dem Ergebnis eine begründete Glaubwürdigkeitseinschätzung zu gewinnen. Soweit demnach aus analytischen Prüfungshandlungen Glaubwürdigkeitseinschätzungen zur Risikomessung verwendet werden, sind diese regelmäßig nicht objektivierbar.

C. Verknüpfung der Einzelrisiken

Wie die bisherigen Ausführungen zur Risikomessung bei den einzelnen Prüfungsmethoden gezeigt haben, existieren neben unterschiedlichen Unschärfemaßen insbesondere zwei unterschiedliche Konzeptionen zur Entscheidungsfindung für den Prüfer. Zum einen kann versucht werden, die Einschätzungen des Prüfers über den tatsächlichen Zustand der Prüfungsgesamtheit durch ein Unschärfemaß auszudrücken, um daraus ein Risikomaß für einzelne Hand-

---

90 A. A. *Akresh*, Response, 1990, S. 107; *Dörner*, Audit Risk, 1992, Sp. 89.

91 So anscheinend *Srivastava/Shafer*, Belief-Function, 1992, S. 257.

92 Vgl. *Kinney/Haynes*, Analytical Procedure, 1990, S. 90.

lungsalternativen zu bestimmen. Dazu wird unterstellt, daß sich die Einschätzungen des Prüfers über die Prüfungsgesamtheit aus den Ergebnissen der angewandten Prüfungsmethoden ableiten lassen. Zum anderen kann argumentiert werden, daß der Prüfer ausschließlich Hypothesen über die tatsächliche Ausprägung der Prüfungsgesamtheit aufstellen kann. Wird davon ausgegangen, daß die ermittelten Prüfungsergebnisse von den beiden Hypothesen unterschiedlich gut erklärt, und somit im Umkehrschluß die Hypothesen unterschiedlich gut durch die Prüfungsergebnisse gestützt werden, dann kann das Risiko der Entscheidung für eine der beiden Hypothesen durch das Verhältnis des Stützungsgrades der abgelehnten zur angenommenen Hypothese gemessen werden. Die zuletzt genannte Konzeption ist zwar wissenschaftstheoretisch besser gestützt[93], setzt allerdings voraus, daß die unterschiedlichen Hypothesen als konkurrierende Theorien, welche die Prüfungsergebnisse besser oder schlechter erklären, aufgefaßt werden können. Diese Voraussetzung ist bei der Prüfung des datenerzeugenden Systems und den auf die dabei ermittelten Ergebnisse aufbauenden Prüfungsverfahren nicht erfüllt, da die Prüfungsgesamtheit das Ergebnis der ausgeführten Datenverarbeitungsfunktionen, also Wirkung, nicht Ursache des datenerzeugenden Systems ist. Daher kann auch die Ordnungsmäßigkeit als Eigenschaft der Prüfungsgesamtheit die Zuverlässigkeit des datenerzeugenden Systems nicht erklären. Da die Prüfungsergebnisse aus Stichproben mit bewußter Auswahl und aus der Systemprüfung bei dieser Konzeption nicht in die Risikomessung einbezogen werden können, soll die angeführte Konzeption im Hinblick auf die Verknüpfung von Einzelrisiken zu einem Gesamtrisikomaß nicht weiter untersucht werden[94].

Für die oben zuerst genannte Konzeption muß noch gezeigt werden, wie die mit Hilfe von einzelnen Prüfungsmethoden gewonnenen Einschätzungen des Prüfers zu einem Gesamtrisikomaß zu verknüpfen sind. Dabei können grundsätzlich zwei Formen der Verknüpfung von Risiken unterschieden werden. Zum einen kann Verknüpfung im Sinne von Kombination unterschiedlicher Einschätzungen des Prüfers, welche die Prüfungsgesamtheit betreffen, zum anderen im Sinne von Aggregation der Einschätzungen, die zu Teilen der Prüfungsgesamtheit vorliegen, verstanden werden. Beide Formen der Verknüpfung von Einzelrisiken werden im folgenden getrennt untersucht.

---

[93] Vgl. die Argumentation bei *Popper*, Logik, 1994, S. 340-343; *Stegmüller*, Wahrscheinlichkeit, 2. Halbband, 1973, S. 102-104.

[94] Soweit ausschließlich Prüfungsmethoden, die Ergebnisse der angeführten Art liefern, eingesetzt werden, wurden die Ansätze zur Risikomessung im Kapitel A. I. 1. b) cc) in diesem Teil der Arbeit und die Aggregation der Prüfungsergebnisse in Kapitel C. II. 2. im Zweiten Teil der Arbeit dargestellt.

I. Kombination von Einzelrisiken

Bei der Kombination von Glaubwürdigkeitseinschätzungen, die aus der Anwendung einzelner Prüfungsmethoden resultieren, sind, soweit ausschließlich Wahrscheinlichkeits- und Glaubwürdigkeitseinschätzungen betrachtet werden, grundsätzlich zwei Fälle zu unterscheiden. Zum einen können alternative Erkenntnisse zu einer einzelnen in irgendeiner Form benötigten Aussage vorliegen. Zum anderen kann der Fall auftreten, daß einzelne Aussagen und die zugehörigen Glaubwürdigkeitseinschätzungen zu einer kombinierten Aussage verknüpft werden müssen. Letzterer liegt insbesondere der Ableitung einer Aussage über die Ordnungsmäßigkeit der Prüfungsgesamtheit zugrunde, die sowohl eine Aussage über den Fehleranteil als auch über das Fehlerausmaß und eine Aussage über die Vollständigkeit der Prüfungsgesamtheit sowie die Vertrauenswürdigkeit der verwendeten Informationen, soweit diese nicht im Rahmen der ergebnisorientierten Prüfung beurteilt werden konnten, erforderlich macht. Da zumindest im Rahmen der Stichprobenverfahren mit bewußter Auswahl darüber hinaus in einzelnen Fällen Unschärfe ausschließlich durch ein Notwendigkeitsmaß gemessen werden kann, soll im Anschluß noch untersucht werden, unter welchen Voraussetzungen Glaubwürdigkeits- und Notwendigkeitsmaße kombiniert werden können.

1. Alternative Erkenntnisse zu einer einzelnen Aussage

Einzelne Aussagen, für die eine Kombination alternativer Erkenntnisse in Betracht kommt, sind insbesondere Aussagen über die für die Beurteilung der Prüfungsgesamtheit relevanten Kriterien, also eine Aussage über den Fehleranteil oder eine Aussage über das Fehlerausmaß, sowie die im Rahmen der bewußten Auswahl benötigten Aussagen über das Vorliegen der benötigten Prämissen. Es soll zunächst vereinfachend der Fall betrachtet werden, daß zwei voneinander unabhängige alternative Glaubwürdigkeitseinschätzungen zu einer benötigten Aussage vorliegen. Im vorliegenden Prüfungsmodell kommen hier die Ergebnisse zweier Zufallsstichproben oder die Ergebnisse aus einer Stichprobe mit bewußter Auswahl und einer Zufallsstichprobe in Betracht. Die Erweiterung auf den Fall, daß mehr als zwei unabhängige Glaubwürdigkeitseinschätzungen vorliegen, erfolgt durch wiederholte Anwendung der im folgenden angeführten Vorgehensweise.

Die vorliegenden Erkenntnisse werden im angeführten Fall durch zwei unterschiedliche grundlegende Wahrscheinlichkeitszuordnungen m_1 und m_2 zu einem frame of discernment Θ abgebildet. Es soll $\Theta = \{q;\overline{q}\}$ gelten, wobei q für die

$A_i, B_j \in \Theta$	$m_1(A_i)$	$m_2(B_j)$
$\varnothing$	0	0
$\{q\}$	$1-\alpha_1$	$1-\alpha_2$
$\{\bar{q}\}$	0	0
Θ	α_1	α_2

Abb. 3.7: Grundlegende Wahrscheinlichkeitszuordnungen bei gleichgerichteten Glaubwürdigkeitseinschätzungen

Wahrheit der betrachteten Aussage steht, also z. B., daß $\theta \leq \theta^*$ gilt, und entsprechend $\bar{q}$ für die Falschheit der betrachteten Aussage stehen soll.

Die Kombination der Erkenntnisse kann gemäß (2.50) erfolgen. Für die Risikomessung sind insbesondere zwei Fälle interessant. Zum einen können sich beide Erkenntnisse auf eines der beiden Elemente aus Θ beziehen, die Erkenntnisse sind demnach gleichgerichtet. Zum anderen ist der Fall denkbar, daß sowohl für die Glaubwürdigkeit von q als auch für die Glaubwürdigkeit von $\bar{q}$ Erkenntnisse vorhanden sind, diese also zum Teil widersprüchlich sind. Es sollen zunächst die Auswirkungen auf das Gesamtrisiko bei Vorliegen des ersten Falles gezeigt werden. Dazu wird angenommen, daß die in Abb. 3.7 dargestellten grundlegenden Wahrscheinlichkeitzuordnungen ermittelt wurden. Für α_1 und α_2 soll generell gelten:

(3.42) $$0 < \alpha_1 < 1-\alpha_1 \text{ und } 0 < \alpha_2 < 1-\alpha_2.$$

$A_i \in \Theta$	$B_j \in \Theta$	$A = A_i \cap B_j$	$m_1(A_i)m_2(B_j)$
$\{q\}$	$\{q\}$	$\{q\}$	$(1-\alpha_1)(1-\alpha_2)$
$\{q\}$	Θ	$\{q\}$	$(1-\alpha_1)\alpha_2$
Θ	$\{q\}$	$\{q\}$	$\alpha_1(1-\alpha_2)$
Θ	Θ	Θ	$\alpha_1\alpha_2$

Abb. 3.8: Bestimmung der gemeinsamen grundlegenden Wahrscheinlichkeitszuordnungen bei gleichgerichteten Glaubwürdigkeitseinschätzungen

Die Verknüpfung der einzelnen grundlegenden Wahrscheinlichkeitszuordnungen m_1 und m_2 ist in Abb. 3.8 dargestellt. Nach (2.50) erhält man als gemeinsame grundlegende Wahrscheinlichkeitszuordnung $m(\{q\}) = 1-\alpha_1\alpha_2$ und

$A_i, B_j \in \Theta$	$m_1(A_i)$	$m_2(B_j)$
$\varnothing$	0	0
$\{q\}$	$1-\alpha_1$	0
$\{\overline{q}\}$	0	$1-\alpha_2$
Θ	α_1	α_2

Abb. 3.9: Grundlegende Wahrscheinlichkeitszuordnungen bei widersprüchlichen Glaubwürdigkeitseinschätzungen

$m(\Theta)=\alpha_1\alpha_2$. Als Glaubwürdigkeitsmaß erhält man $bel(\{q\})=1-\alpha_1\alpha_2$ und $bel(\{\overline{q}\})=0$. Die gemeinsame Glaubwürdigkeit der durch q repräsentierten Aussage ist demnach in jedem Fall größer als jede der beiden ursprünglichen Glaubwürdigkeitseinschätzungen. Demnach wird erwartungsgemäß das Gesamtrisiko für die betrachtete Aussage geringer, wenn sämtliche vorliegenden Erkenntnisse für die betrachtete Aussage sprechen.

Die grundlegenden Wahrscheinlichkeitszuordnungen für den Fall, daß die vorliegenden Erkenntnisse im Hinblick auf die Glaubwürdigkeit von q und $\overline{q}$ widersprüchlich sind, ist in Abb. 3.9 dargestellt. Wie die Verknüpfung der einzelnen grundlegenden Wahrscheinlichkeitszuordnungen m_1 und m_2 in Abb. 3.10 zeigt, wirkt sich der Widerspruch der Erkenntnisse durch die Zuweisung eines Teils der grundlegenden Wahrscheinlichkeiten an $\varnothing$ aus. Diese ist nach (2.47) nicht zulässig, so daß sich nach Normierung der gemeinsamen grundlegenden Wahrscheinlichkeitszuordnung die folgenden Glaubwürdigkeitseinschätzungen für q und $\overline{q}$ ergeben:

(3.43) $$bel(\{q\})=(1-\alpha_1)\frac{\alpha_2}{1-(1-\alpha_1)(1-\alpha_2)} \text{ und}$$

(3.44) $$bel(\{\overline{q}\})=(1-\alpha_2)\frac{\alpha_1}{1-(1-\alpha_1)(1-\alpha_2)}.$$

Wie zu erkennen ist, nimmt die Glaubwürdigkeitseinschätzung der beiden möglichen Aussagen für den Fall, daß entgegengesetzte Erkenntnisse vorliegen, generell ab. Insoweit entsprechen auch in diesem Fall die durch die Kombination von Glaubwürdigkeitseinschätzungen gewonnenen Ergebnisse den intuitiv zu erwartenden Ergebnissen.

DOI https://doi.org/10.3790/978-3-428-48909-1

$A_i \in \Theta$	$B_j \in \Theta$	$A = A_i \cap B_j$	$m_1(A_i)m_2(B_j)$
$\{q\}$	$\{\bar{q}\}$	$\varnothing$	$(1-\alpha_1)(1-\alpha_2)$
$\{q\}$	Θ	$\{q\}$	$(1-\alpha_1)\alpha_2$
Θ	$\{\bar{q}\}$	$\{\bar{q}\}$	$\alpha_1(1-\alpha_2)$
Θ	Θ	Θ	$\alpha_1\alpha_2$

Abb. 3.10: Bestimmung der gemeinsamen grundlegenden Wahrscheinlichkeitszuordnungen bei widersprüchlichen Glaubwürdigkeitseinschätzungen

2. *Risikomessung bei kombinierten Aussagen*

Bei der Zusammenfassung der Glaubwürdigkeitseinschätzungen für kombinierte Aussagen ergeben sich Besonderheiten ausschließlich dann, wenn die zu kombinierenden Aussagen logisch durch eine Konjunktion verknüpft sind. Dieser Fall kann außer bei der Ableitung des Gesamturteils über die Prüfungsgesamtheit insbesondere auch im Rahmen der Stichprobenverfahren mit bewußter Auswahl dann auftreten, wenn mehrere Voraussetzungen für die Anwendung eines Auswahlkriteriums erfüllt sein müssen oder wenn sowohl für die Majorprämisse als auch für die Minorprämisse nicht angenommen werden kann, daß diese mit Sicherheit gelten, sondern ausschließlich Glaubwürdigkeitseinschätzungen für ihre Gültigkeit vorliegen. Soweit die logische Verknüpfung der zu kombinierenden Aussagen durch eine Disjunktion erfolgt, können die Glaubwürdigkeitseinschätzungen wie alternative Erkenntnisse zur benötigten kombinierten Aussage behandelt werden. Es gelten somit hinsichtlich der dabei auftretenden Ergebnisse die im vorigen Abschnitt gemachten Ausführungen.

$A_i \in \Theta_1$	$m_1(A_i)$
$\varnothing$	0
$\{q_1\}$	$1-\alpha_1$
$\{\bar{q}_1\}$	0
Θ_1	α_1

$B_j \in \Theta_2$	$m_2(B_j)$
$\varnothing$	0
$\{q_2\}$	$1-\alpha_2$
$\{\bar{q}_2\}$	0
Θ_2	α_2

Abb. 3.11: Grundlegende Wahrscheinlichkeitszuordnungen bei gleichgerichteten Glaubwürdigkeitseinschätzungen zu einer kombinierten Aussage

DOI https://doi.org/10.3790/978-3-428-48909-1

Für die folgende Darstellung soll davon ausgegangen werden, daß Erkenntnisse zu zwei unterschiedlichen Aussagen, die logisch durch „und“ verknüpft werden sollen, vorliegen. Es existieren demnach zwei grundlegende Wahrscheinlichkeitszuordnungen m_1 und m_2, die allerdings im Unterschied zu den oben betrachteten Fällen über unterschiedlichen frames of discernment Θ_1 und Θ_2 definiert sind. Für die beiden frames of discernment soll gelten $\Theta_1 = \{q_1;\overline{q}_1\}$ und $\Theta_2 = \{q_2;\overline{q}_2\}$. Für die kombinierte Aussage existiert dann der folgende verfeinerte frame of discernment:

(3.45) $$\Theta = \Theta_1 \times \Theta_2 = \{\langle q_1;q_2\rangle;\langle q_1;\overline{q}_2\rangle;\langle \overline{q}_1;q_2\rangle;\langle \overline{q}_1;\overline{q}_2\rangle\}.$$

Für die beiden zu kombinierenden Aussagen gilt wiederum, daß die erste Aussage wahr ist, wenn q_1 eintritt, und die zweite, wenn q_2 eintritt. Demnach ist die kombinierte Aussage ausschließlich dann wahr, wenn sowohl q_1 als auch q_2 eintreten. Es wird zunächst wiederum der Fall untersucht, daß die vorhandenen Erkenntnisse für die betrachtete Aussage sprechen. Dazu sollen die in Abb. 3.11 dargestellten grundlegenden Wahrscheinlichkeitszuordnungen ermittelt worden sein. Wie aus Abb. 3.12 abgelesen werden kann, erhält man als Glaubwürdigkeit für die Wahrheit der kombinierten Aussage bzw. ihrer Negation:

(3.46) $$bel(\{\langle q_1;q_2\rangle\}) = (1-\alpha_1)(1-\alpha_2),$$

(3.47) $$bel(\{\langle q_1;\overline{q}_2\rangle;\langle \overline{q}_1;q_2\rangle;\langle \overline{q}_1;\overline{q}_2\rangle\}) = 0.$$

Es soll auch hier abschließend der Fall betrachtet werden, daß die vorliegenden Erkenntnisse sich zum Teil widersprechen. Die grundlegenden Wahrscheinlichkeitszuordnungen sind in Abb. 3.13 dargestellt. Aus Abb. 3.14 ist zu erkennen, daß die Glaubwürdigkeit für die Wahrheit der kombinierten Aussage jetzt Null ist, da auf $\{\langle q_1;q_2\rangle\}$ keine grundlegende Wahrscheinlichkeit zugewie-

$A_i \in \Theta_1$	$B_j \in \Theta_2$	$A = A_i \cap B_j$	$m_1(A_i)m_2(B_j)$
$\{q_1\}$	$\{q_2\}$	$\{\langle q_1;q_2\rangle\}$	$(1-\alpha_1)(1-\alpha_2)$
$\{q_1\}$	Θ_2	$\{\langle q_1;q_2\rangle;\langle q_1;\overline{q}_2\rangle\}$	$(1-\alpha_1)\alpha_2$
Θ_1	$\{q_2\}$	$\{\langle q_1;q_2\rangle;\langle \overline{q}_1;q_2\rangle\}$	$\alpha_1(1-\alpha_2)$
Θ_1	Θ_2	Θ	$\alpha_1\alpha_2$

Abb. 3.12: Bestimmung der gemeinsamen grundlegenden Wahrscheinlichkeitszuordnungen bei gleichgerichteten Glaubwürdigkeitseinschätzungen zu einer kombinierten Aussage

DOI https://doi.org/10.3790/978-3-428-48909-1

sen wird. Allerdings ist die Glaubwürdigkeit für die Falschheit der betrachteten Aussage jetzt größer als Null. Sie ergibt sich nach Normalisierung der in Abb. 3.14 ermittelten grundlegenden Wahrscheinlichkeitszuordnung mit:

$$bel\left(\left\{\left\langle\overline{q_1;q_2}\right\rangle\right\}\right)=\frac{\alpha_1(1-\alpha_2)}{1-(1-\alpha_1)(1-\alpha_2)}. \tag{3.48}$$

Neben den bisher diskutierten Fällen von vollständig gleichgerichteten und widersprüchlichen Glaubwürdigkeitseinschätzungen sind auch Kombinationen denkbar, bei denen die Glaubwürdigkeitseinschätzungen zum Teil gleichgerichtet und zum Teil widersprüchlich sind. Diese werden insbesondere aus der Verknüpfung alternativer Erkenntnisse entstehen, da die einzelnen Prüfungsmethoden regelmäßig nur einfache Glaubwürdigkeitszuordnungen, die entweder für oder gegen eine Aussage sprechen, erzeugen[95]. Zu beachten ist, daß, wie die Ergebnisse in Abb. 3.12 gezeigt haben, bei der Ermittlung des Gesamtrisikos selbst im bestmöglichen Fall vollständig gleichgerichteter Erkenntnisse, die Glaubwürdigkeit für die Wahrheit der kombinierten Aussage, schließt man den Fall sicherer Erkenntnisse aus, immer kleiner ist als die kleinste Glaubwürdigkeit der miteinander zu verknüpfenden Aussagen.

$A_i \in \Theta_1$	$m_1(A_i)$
$\varnothing$	0
$\{q_1\}$	$1-\alpha_1$
$\{\overline{q}_1\}$	0
Θ_1	α_1

$B_j \in \Theta_2$	$m_2(B_j)$
$\varnothing$	0
$\{q_2\}$	0
$\{\overline{q}_2\}$	$1-\alpha_2$
Θ_2	α_2

Abb. 3.13: Grundlegende Wahrscheinlichkeitszuordnungen bei widersprüchlichen Glaubwürdigkeitseinschätzungen zu einer kombinierten Aussage

[95] Siehe oben Kapitel A. und B. in diesem Teil der Arbeit. Sowie *Shafer* et al., Auditor's Assistant, 1988, S. 65.

$A_i \in \Theta_1$	$B_j \in \Theta_2$	$A = A_i \cap B_j$	$m_1(A_i)m_2(B_j)$
$\{q_1\}$	$\{\bar{q}_2\}$	$\varnothing$	$(1-\alpha_1)(1-\alpha_2)$
$\{q_1\}$	Θ_2	$\{\langle q_1;q_2\rangle;\langle q_1;\bar{q}_2\rangle\}$	$(1-\alpha_1)\alpha_2$
Θ_1	$\{\bar{q}_2\}$	$\{\langle q_1;\bar{q}_2\rangle;\langle \bar{q}_1;\bar{q}_2\rangle\}$	$\alpha_1(1-\alpha_2)$
Θ_1	Θ_2	Θ	$\alpha_1\alpha_2$

Abb. 3.14: Bestimmung der gemeinsamen grundlegenden Wahrscheinlichkeitszuordnungen bei widersprüchlichen Glaubwürdigkeitseinschätzungen zu einer kombinierten Aussage

3. Kombination von Glaubwürdigkeits- und Notwendigkeitsmaß

Wie oben ausgeführt wurde[96], kommt bei der Anwendung von Stichprobenverfahren mit dedektiver Auswahl der Stichprobenelemente die Verwendung von linguistischen Variablen im Rahmen des plausiblen Schließens in Betracht. Aus dem Stichprobenergebnis selbst und der unscharfen Aussage über den möglichen Fehleranteil für die Elemente der Prüfungsgesamtheit, die nicht durch die Stichprobe erfaßt werden, erhält man insgesamt ein Möglichkeits- oder Notwendigkeitsmaß über den gesamten Fehleranteil. Dabei interessiert im vorliegenden Modell zur Risikomessung ausschließlich ein Notwendigkeitsmaß über den Fehleranteil[97]. Da aus den anderen Prüfungshandlungen im wesentlichen Glaubwürdigkeitseinschätzungen über den Fehleranteil in der Prüfungsgesamtheit vorliegen, ist fraglich, unter welchen Voraussetzungen eine Kombination der beiden Unschärfemaße zu einem Gesamtrisikomaß möglich ist. Dabei gilt zunächst, daß die in (2.39) bis (2.41) festgelegten Axiome, die ein Glaubwürdigkeitsmaß erfüllen muß, von einem Notwendigkeitsmaß nicht generell erfüllt werden, so daß die Zusammenfassung der beiden Maße nicht unter allen Umständen garantiert ist.

Da es sich bei dem Notwendigkeitsmaß um ein allgemeines Unschärfemaß im Sinne der Voraussetzungen nach (2.69) bis (2.71) handelt und ein Glaubwürdigkeitsmaß dann ein allgemeines Unschärfemaß ist, wenn die grundlegende Wahrscheinlichkeitszuordnung konsonant ist[98], kann in diesem Fall das ermittelte Notwendigkeitsmaß zumindest formal wie ein Glaubwürdigkeitsmaß

96 Siehe oben Kapitel A. II. 2. in diesem Teil der Arbeit.

97 Siehe oben Kapitel B. III. 3. a) im Zweiten Teil der Arbeit.

98 Siehe oben Kapitel B. III. 1. im Zweiten Teil der Arbeit.

behandelt werden. Dazu muß allerdings unterstellt werden, daß die gemäß (2.106) zu dem Notwendigkeitsmaß gehörende grundlegende Möglichkeitszuordnung aus konsistenten und konsonanten Erkenntnissen abgeleitet wurde[99]. Bei den oben angeführten Anwendungsfällen wird die grundlegende Möglichkeitszuordnung durch die aus der plausiblen Schlußfolgerung resultierende Ausprägung einer linguistischen Variable bestimmt. Werden die Beispieldaten aus Abb. 2.4 verwendet, und führt eine Schlußfolgerung zu dem Ergebnis, daß der FEHLERANTEIL in einem betrachteten Bereich der Prüfungsgesamtheit „nicht sehr hoch" ist, so erhält man formal nach (2.101), wenn angenommen wird, daß der Zugehörigkeitsfunktion eine konsistente und konsonante Glaubwürdigkeitseinschätzung zugrundeliegt, den Grad der Notwendigkeit und der Glaubwürdigkeit für einen Fehleranteil von nicht mehr als 15 % mit:

$$nec(\{\theta|\theta \le 0{,}15\}) = bel(\{\theta|\theta \le 0{,}15\}) = 0{,}1. \tag{3.49}$$

Aus dem Umstand, daß die beiden Maße in bestimmten Fällen die gleichen Axiome erfüllen, kann allerdings nicht zwingend gefolgert werden, daß sie auch inhaltlich das Gleiche messen[100]. Es muß daher im Einzelfall geprüft werden, ob die den beiden Maßen zugrundeliegenden Erkenntnisse die vorgenommene Gleichsetzung erlauben. Die dabei auftretenden Probleme bzw. zu treffenden Annahmen sollen anhand des angeführten Beispiels erläutert werden.

Dazu soll zunächst gezeigt werden, wie die zu der in (3.49) vorgenommenen Glaubwürdigkeitsbeurteilung gehörende grundlegende Wahrscheinlichkeitszuordnung definiert ist. Ausgangspunkt müssen die für die Ableitung der Zugehörigkeitsfunktion erhobenen Daten sein. Diese können zunächst ausschließlich eine Glaubwürdigkeitseinschätzung für einen sehr hohen Fehleranteil liefern. Insoweit muß unterstellt werden können, daß bei veränderter Fragestellung nach Fehleranteilen, die nicht sehr hoch sind, die in Abb. 2.4 dargestellten relativen Häufigkeiten gemessen worden wären. Die Unterstellung ist nicht spezifisch notwendig, um das Notwendigkeits- und Glaubwürdigkeitsmaß gleichzusetzen. Sie wird implizit auch dann gemacht, wenn der in (2.83) definierte Operator für die Bildung des Komplements einer unscharfen Menge zur Ableitung der Zugehörigkeitsfunktion für die Negation einer bekannten Ausprägung einer linguistischen Variable verwendet wird. Auf diese Weise wurde im vorliegenden Fall die zur Bestimmung des Notwendigkeitsmaßes benötigte Zugehörigkeitsfunktion für die Ausprägung „nicht sehr hoch" aus der bekannten Zugehörigkeitsfunktion für die Ausprägung „sehr hoch" abgeleitet.

---

99 Vgl. *Kruse* et al., Fuzzy-Systeme, 1993, S. 86.

100 Vgl. *Smets*, Belief Functions, 1988, S. 264.

i	„nicht sehr hoher" FEHLERANTEIL $c_i \in \mathbb{C}_{\text{nicht sehr hoch}}$	relative Häufigkeit $pr(c_i)$
1	$<15\%$	0,1
2	$<17{,}5\%$	0,25
3	$<20\%$	0,5
4	$<25\%$	0,15

Abb. 3.15: Fortsetzung des Beispiels aus Abb. 2.4
zur Ableitung einer Zugehörigkeitsfunktion für die Ausprägung „nicht sehr hoch"
der linguistischen Variable FEHLERANTEIL

Zur Ableitung einer grundlegenden Wahrscheinlichkeitszuordnung muß zunächst ein geeigneter frame of discernment Θ definiert werden. Es soll $\Theta = [0;1]$ angenommen werden. Aus den Angaben in Abb. 2.4 ergibt sich die folgende grundlegende Wahrscheinlichkeitszuordnung $\langle \mathbb{E}; m \rangle$ mit $\mathbb{E} = \{c_1; c_2; c_3; c_4\}$ und:

(3.50)
$$m(c_1) = pr(c_1) = 0{,}1,\ m(c_2) = pr(c_2) = 0{,}25,$$
$$m(c_3) = pr(c_3) = 0{,}5,\ m(c_4) = pr(c_4) = 0{,}15.$$

Wie aus (3.50) zu erkennen ist, erhält man für $bel(\{\theta | \theta \leq 0{,}15\}) = 0{,}1$, da außer c_1 keine der Mengen, denen eine grundlegende Wahrscheinlichkeit zugeordnet wurde, in der betrachteten Menge $\{\theta | \theta \leq 0{,}15\}$ enthalten ist. Insoweit ist die oben vorgenommene Gleichsetzung von Notwendigkeit und Glaubwürdigkeit berechtigt. Generell gilt, daß eine grundlegende Wahrscheinlichkeitszuordnung $\langle \mathbb{E}; m \rangle$ immer dann konsonant ist, wenn, wie im vorliegenden Fall alle Elemente aus $\mathbb{E}$ verschachtelt sind, also $c_1 \subset c_2 \subset c_3 \subset c_4$ gilt[101].

Die Gleichsetzung von Notwendigkeits- und Glaubwürdigkeitsmaß stellt also zum einen bestimmte Anforderungen an die den verwendeten Zugehörigkeitsfunktionen zugrundezulegenden Daten. Zum anderen muß auch der Kontext, aus dem diese Daten stammen, für die benötigte Glaubwürdigkeitsbeurteilung relevant sein. Dies ist im vorliegenden Fall ausschließlich dann gewährleistet, wenn die befragten Prüfer aus der Gruppe stammen, aus der auch die Zugehörigkeitsfunktionen für die Ausprägungen der übrigen im Rahmen der verwendeten Schlußform benötigten linguistischen Variablen abgefragt wurde. Darüber hinaus ergibt sich zumindest im vorliegenden Fall eine geänderte Interpretation für die abgeleitete Glaubwürdigkeit des ermittelten Fehleranteils, da sie nicht mehr den Grad der Glaubwürdigkeit, den der Prüfer dem Vorliegen eines bestimmten Fehleranteils in der betrachteten Prüfungsgesamt-

---

[101] Vgl. *Shafer*, Theory, 1976, S. 219f; *Dubois/Prade*, Fuzzy Sets, 1989, S. 137.

DOI https://doi.org/10.3790/978-3-428-48909-1

heit beimißt, ausdrückt, sondern den Grad, mit dem die Gruppe von Prüfern einen bestimmten Fehleranteil bei den vorliegenden Prämissen generell für glaubwürdig hält, darstellt. Zumindest im Rahmen der Jahresabschlußprüfung muß der Prüfer entscheiden, inwieweit er ein solches Glaubwürdigkeitsmaß auf die vorliegende Prüfungssituation übertragen will, da er gemäß § 43 Abs. 1 WPO zur eigenverantwortlichen Prüfung verpflichtet ist.

II. Aggregation von Risiken

Zumindest bei der Prüfung größerer oder komplexer Prüfungsgesamtheiten werden die im einzelnen angeführten Prüfungsmethoden regelmäßig nur auf Teile der Prüfungsgesamtheit angewandt. Insbesondere im Rahmen der Jahresabschlußprüfung wird die Aufteilung des gesamten Prüfungsgebiets in Prüffelder und Prüffeldgruppen gefordert[102]. Allerdings stehen keine widerspruchsfreien Kriterien zur Verfügung, wie diese Aufteilung zu erfolgen hat[103]. Für die folgenden Ausführungen soll eine Aufteilung in Teilgesamtheiten ausschließlich anhand der im einzelnen angewandten Prüfungsmethoden erfolgen. D. h., es werden Teilgesamtheiten betrachtet, zu denen entweder Ergebnisse aus der Anwendung von Zufallsstichprobenverfahren oder von Stichprobenverfahren mit bewußter Auswahl der Stichprobenelemente vorliegen. Dabei ist nicht ausgeschlossen, daß zur Bildung der einzelnen Teilgesamtheiten andere Kriterien, wie die Vertrauenswürdigkeit der verwendeten Informationen oder die Verfügbarkeit von Aussagen aus der Prüfung des datenerzeugenden Systems, herangezogen wurden. Die Beschränkung auf diese beiden Typen von Teilgesamtheiten stellt allerdings keine wesentliche Einschränkung des verwendeten Prüfungsmodells dar, da die Prüfungsinformationen aus den indirekten Messungen zu einzelnen Teilgesamtheiten nach den im vorherigen Abschnitt behandelten Grundsätzen zur Kombination von Einzelrisiken mit den Stichprobenergebnissen zusammengefaßt werden können. Sie finden insoweit bei der Bestimmung des Gesamtrisikos Berücksichtigung.

Für die Aggregation der Ergebnisse zu den einzelnen Teilgesamtheiten sind zunächst zwei grundlegend unterschiedliche Vorgehensweisen denkbar. Zum einen kann die einzelne Teilgesamtheit für sich betrachtet werden und ein Urteil über ihre Ordnungsmäßigkeit sowie das damit verbundene Risiko nach den bisher erläuterten Grundsätzen ermittelt werden. Für die Beurteilung der Prü-

[102] Vgl. *IDW*, FG 1/1988, 1989, S. 12.

[103] Vgl. *Leffson*, Wirtschaftsprüfung, 1988, S. 162f.; *v. Wysocki*, Grundlagen, 1988, S. 270f.

fungsgesamtheit wären dann diese Teilurteile geeignet zu aggregieren[104]. Zum anderen kann versucht werden, die Prüfungsergebnisse getrennt nach den beiden Urteilskriterien direkt zu aggregieren und erst für die Prüfungsgesamtheit eine Beurteilung anhand des Gesamtergebnisses für beide Kriterien θ und $\Delta(X,Y)$ vorzunehmen.

Bei der zuerst genannten Vorgehensweise treten eine Reihe von Problemen auf. Zunächst müßten, um eine Beurteilung der Teilgesamtheiten zu ermöglichen, aus den für die Prüfungsgesamtheit vorgegebenen Kriterien θ^* und Δ^* Urteilskriterien für die Teilgesamtheiten abgeleitet werden. Diese müssen, zumindest wenn Stichprobenverfahren mit Auswahl der Stichprobenelemente nach ihrer Bedeutung verwendet werden sollen, schon vor der Prüfungsdurchführung festgelegt werden[105]. Darüber hinaus müßten in Abhängigkeit von den für die Teilgesamtheiten gewählten Urteilskriterien die Urteilskriterien für die Prüfungsgesamtheit neu bestimmt werden, da dann als Elemente der Prüfungsgesamtheit nicht mehr die ursprünglich definierten Elemente x_i fungieren würden sondern die einzelnen Teilgesamtheiten. Kann außerdem unterstellt werden, daß die Unschärfe der Urteile über die einzelnen Teilgesamtheiten bei der Bestimmung des Gesamtrisikos berücksichtigt werden soll, ist weiterhin fraglich, wie die einzelnen Glaubwürdigkeitseinschätzungen der Teilurteile zu gewichten sind. Der Ansatz soll wegen der angeführten Probleme hier nicht weiter untersucht werden, zumal die angeführte Alternative mit weniger Problemen zumindest für das hier unterstellte Prüfungsmodell, in dem ausschließlich ein Gesamturteil gefordert ist, zu einer Lösung führt.

Zur Frage, wie die Aggregation der einzelnen Teilergebnisse erfolgen kann, muß zwischen den Ergebnissen aus mathematisch-statistischen Verfahren und den Stichprobenverfahren mit bewußter Auswahl differenziert werden. Wie oben angeführt wurde, soll die Aufteilung der Prüfungsgesamtheit in Teilgesamtheiten so erfolgen, daß zu einer Teilgesamtheit entweder eine Zufallsstichprobe oder eine Stichprobe mit bewußter Auswahl vorliegt, aber nicht beide gemeinsam. Die zur Prüfungsplanung festgelegten Prüffelder müssen demnach, falls sowohl Zufallsstichproben als auch Stichproben mit bewußter Auswahl aus einem Prüffeld erhoben wurden, zumindest in zwei Teilgesamtheiten aufgeteilt werden. Falls ein Verfahren gefunden wird, alle Ergebnisse aus den Zufallsstichproben einerseits und alle übrigen Ergebnisse andererseits zu aggre-

[104] Vgl. *v. Wysocki*, Grundlagen, 1988, S. 249f.

[105] Zur Festlegung der Materiality-Grenzen für die Teilgesamtheiten siehe z.B. *Quick*, Risiken, 1996, S. 215ff.

gieren, kann das zu lösende Problem auf die Beurteilung einer Prüfungsgesamtheit mit zwei Teilgesamtheiten reduziert werden.

Für die mathematisch-statistischen Verfahren können die Ergebnisse über alle Teilgesamtheiten, für die Zufallsstichproben vorliegen, durch die Anwendung der bei den Verfahren der einfachen Mittelwertschätzung aus geschichteten Zufallsstichproben diskutierten Stichprobenfunktionen aggregiert werden[106]. Dabei sind die im Zusammenhang mit den geschichteten Stichprobenverfahren auftretenden Planungsprobleme, namentlich die Bestimmung der Schichtenanzahl und der Schichtgrenzen sowie die Bestimmung und Aufteilung des Stichprobenumfangs, für die hier beabsichtigte Anwendung nicht relevant, da ausschließlich eine Zusammenfassung der vorliegenden Prüfungsergebnisse und keine Planung der Stichprobenprüfung beabsichtigt ist. Auch der Umstand, daß die Ergebnisse im Hinblick auf die mögliche Sicherheit und Genauigkeit der Schätzung nur zufällig optimal sein werden, da die Schichten und evtl. auch der Stichprobenumfang innerhalb der Schichten nicht nach Effizienzkriterien festgelegt wurden[107], beeinflußt die hier untersuchte Möglichkeit zur Aggregation der Ergebnisse grundsätzlich nicht. Probleme können ausschließlich bei der Bestimmung der relevanten Grenzverteilung für die verwendete Stichprobenfunktion auftreten[108]. Insgesamt erscheint es jedoch möglich, aus den vorliegenden Zufallsstichproben Konfidenzintervalle für die beiden Urteilskriterien abzuleiten, welche die Prüfungsinformationen sämtlicher Teilgesamtheiten, zu denen Zufallsstichproben vorliegen, berücksichtigen.

Die Aggregation der Ergebnisse aus den Stichproben mit bewußter Auswahl sowie den übrigen zu diesen Teilgesamtheiten existierenden Glaubwürdigkeitseinschätzungen erfordert zwei Schritte. Zunächst werden die alternativen Erkenntnisse für jede Teilgesamtheit nach den oben angeführten Regeln aggregiert[109]. Danach können die Stichprobenergebnisse so zu einem Gesamtergebnis D' zusammengefaßt werden, als ob sie aus einer einzelnen Stichprobe stammen. Die Glaubwürdigkeitseinschätzung dieses Gesamtergebnisses ergibt sich durch Kombination der Glaubwürdigkeiten der Einzelergebnisse, wobei unterstellt wird, daß die einzelnen Aussagen durch eine logische Konjunktion verknüpft werden[110].

---

[106] Vgl. *Cochran*, Sampling, 1977, S. 91f.; *Obermeier*, Abschlussprüfung, 1983, S. 137f.; *Mandl*, Anwendungsvoraussetzungen, 1984, S. 130.

[107] Vgl. *Mandl*, Anwendungsvoraussetzungen, 1984, S. 63-67 und 131f.

[108] Vgl. *Obermeier*, Abschlussprüfung, 1983, S. 140.

[109] Siehe oben Kapitel C. I. 1. in diesem Teil der Arbeit.

[110] Siehe oben Kapitel C. I. 2. in diesem Teil der Arbeit.

Insgesamt liegen dann für die Prüfungsgesamtheit zu jedem Urteilskriterium zwei Stichprobenergebnisse vor, die sich jeweils auf einen Teil der Prüfungsgesamtheit beziehen. Der Umfang des Teils der Prüfungsgesamtheit, für den ausschließlich ein Konfidenzintervall über den wahren Parameter θ' bzw. Δ' in der Teilgesamtheit vorliegt, soll mit N' bezeichnet werden. Für die zweite Teilgesamtheit liegen Fehleranteil bzw. Fehlerausmaß als feste Größe vor. Zur Kombination der beiden Ergebnisse wird das Konfidenzniveau für die Intervallobergrenze bestimmt, bei der zusammen mit den Fehlern, die in der Stichprobe mit bewußter Auswahl enthalten sind, die kritischen Werte θ^* bzw. Δ^* genau eingehalten werden. Wird die Vorgehensweise wiederum für den Fehleranteil betrachtet, muß das Konfidenzniveau für ein Intervall mit der Obergrenze p_o abgeleitet werden, die nach folgender Formel zu bestimmen ist:

(3.51) $$p_o = \theta^* - \frac{|D'|}{N - N'}.$$

Zur Beurteilung des Fehlerausmaßes wird das Konfidenzniveau für ein Intervall, dessen Obergrenze analog dem Vorgehen in (3.51) aus Δ^* abgeleitet wird, herangezogen.

Insgesamt kann dann für beide Urteilskriterien ein Glaubwürdigkeitsmaß im Hinblick auf die Beurteilung der Prüfungsgesamtheit als ordnungsmäßig bzw. nicht ordnungsmäßig abgeleitet werden. Dazu werden die beiden vorliegenden Aussagen kombiniert, wobei wiederum die oben angeführten Regeln für die Kombination von Aussagen, die durch eine logische Konjunktion verknüpft sind, Anwendung finden. Abschließend werden nach der gleichen Regel die Glaubwürdigkeitseinschätzungen für die beiden Urteilskriterien zu einem Gesamtrisikomaß zusammengefaßt. Mit Hilfe der hier angeführten Vorgehensweise läßt sich somit eine Glaubwürdigkeitseinschätzung für das Gesamturteil über die Prüfungsgesamtheit auch bei Prüfungsergebnissen, die zunächst ausschließlich für Teilgesamtheiten ermittelt wurden, ableiten.

DOI https://doi.org/10.3790/978-3-428-48909-1

Ergebnisse

A. Risikomessung im Rahmen der Prüfung

Wie die Analyse des Modells einer Prüfung im Ersten Teil der Arbeit gezeigt hat, treten bei der Beurteilung einer Prüfungsgesamtheit unterschiedliche Formen der Unschärfe auf, die zum Teil auf Unschärfe der im Rahmen der Prüfungshandlungen ermittelten Ergebnisse beruhen, zum Teil aber auch aus fehlenden Kenntnissen über die genauen Zusammenhänge zwischen Realität und deren Abbildung in der Prüfungsgesamtheit einerseits sowie dem datenerzeugenden System und den Eigenschaften des Systemoutputs andererseits resultieren[1]. Insoweit erscheint der gewählte Ansatz, die unterschiedlichen Formen der Unschärfe und zumindest nicht zu Anfang einzelne, wie auch immer abgegrenzte Risikokomponenten zum Ausgangspunkt der Untersuchung zu machen, gerechtfertigt. Da der Arbeit ein einfaches Modell realer Prüfungssituationen zugrundeliegt, sind ausschließlich die einschränkenden Ergebnisse ohne weitere Annahmen auf die Risikomessung in realen Prüfungen, wie sie z. B. die Jahresabschlußprüfung darstellt, übertragbar. Die übrigen Ergebnisse sind insbesondere mit Blick auf die Festlegung, das Risiko als eine theoretische Größe zu ermitteln, welche die Konsequenzen des Prüfungsurteils für den Prüfer nicht berücksichtigt, zur Erklärung oder Prognose des Prüferverhaltens in realen Prüfungssituationen nur eingeschränkt verwendbar. Sie bieten allerdings Ansätze für die Qualitätssicherung im Rahmen von Prüfungen, da diese ausschließlich die ebenfalls theoretische Größe Prüfungsqualität zum Gegenstand hat.

I. Meßbarkeit des Prüfungsrisikos

Wie gezeigt wurde, werden im Rahmen der verfügbaren Prüfungsmethoden zwei Konzeptionen zur Risikomessung verwendet. Einmal kann versucht werden, Risiko aus dem Grad der Überzeugung des Prüfers für eine der beiden möglichen Ausprägungen der Prüfungsgesamtheit abzuleiten. Dieser Ansatz führt zu einem Glaubwürdigkeits- oder Notwendigkeitsmaß bezüglich des Grades der Überzeugung. Das Risiko wird dann durch ein Plausibilitäts- bzw.

[1] Siehe oben Kapitel B. im Ersten Teil der Arbeit.

Möglichkeitsmaß für die nicht gewählte Alternative gemessen. Die zweite Konzeption geht davon aus, daß der Prüfer ausschließlich Hypothesen über die tatsächliche Ausprägung der Prüfungsgesamtheit aufstellen kann. Für eine Entscheidung wird dann auf die unterschiedliche Erklärungskraft der Hypothesen im Hinblick auf die ermittelten Prüfungsergebnisse abgestellt. Das Risiko wird dann, soweit die Erklärungskraft der Hypothesen durch bedingte Wahrscheinlichkeiten ausgedrückt werden kann, durch das Verhältnis zwischen der Wahrscheinlichkeit für die Prüfungsergebnisse bei Gültigkeit der abgelehnten Hypothese zur Wahrscheinlichkeit für die Prüfungsergebnisse bei Gültigkeit der angenommenen Hypothese gemessen.

Probleme bei dieser Konzeption treten dann auf, wenn die Ordnungsmäßigkeit der Prüfungsgesamtheit nicht als Ursache für bestimmte Prüfungsergebnisse betrachtet werden kann, die Ergebnisse in diesem Sinn also nicht erklären kann. Letzteres gilt, wie schon festgestellt wurde, zumindest für die im Rahmen der Systemprüfung getroffenen Feststellungen. Darüber hinaus wurde gezeigt[2], daß auch im Rahmen von mathematisch-statistischen Testverfahren Probleme bei der geeigneten Formulierung der Nullhypothese auftreten, so daß aus der Ablehnung der Nullhypothese nicht in jedem Fall auf die Nichtordnungsmäßigkeit der Prüfungsgesamtheit geschlossen werden kann. Für die Verfahren mit bewußter Auswahl der Stichprobenelemente existiert ebenfalls keine Gesetzmäßigkeit, die einen Zusammenhang zwischen Prüfungsergebnis und Ordnungsmäßigkeit der Prüfungsgesamtheit in der Form herstellen kann, die eine Anwendung der zuletzt angeführten Konzeption zur Risikomessung erlauben würde.

Wie die Untersuchung der verfügbaren Unschärfemaße im Hinblick auf ihre Eignung, den Überzeugungsgrad des Prüfers zu messen, ergeben hat[3], sind die Anforderungen, die an das Wissen des Prüfers über die im einzelnen betrachteten Sachverhalte und die bestehenden Zusammenhänge gestellt werden, bei den verschiedenen Unschärfemaßen unterschiedlich streng. Dabei wurde deutlich, daß unabhängig von der zugrundegelegten Interpretation, objektiv oder subjektiv, die Verwendung eines Wahrscheinlichkeitsmaßes für die Risikomessung nicht in Betracht kommt, da die erforderlichen Informationen über den der Wahrscheinlichkeitsbeurteilung zugrundezulegenden Möglichkeitsraum und dem darauf definierten Ereigniskörper dem Prüfer regelmäßig nicht zur Verfügung stehen. Auch die aus der Anwendung von mathematisch-statisti-

---

2 Siehe oben Kapitel A. I. 1. b) im Dritten Teil der Arbeit.

3 Siehe oben Kapitel B. im Zweiten Teil der Arbeit.

schen Verfahren resultierenden objektiven Wahrscheinlichkeiten über das Stichprobenergebnis führen im Hinblick auf die Risikomessung nicht zu einem Wahrscheinlichkeitsmaß sondern, wie gezeigt wurde[4], zu einem Glaubwürdigkeitsmaß.

Im Gegensatz zum Wahrscheinlichkeitsmaß, dessen Verwendung einen sehr fein strukturierten Möglichkeitsraum und die Kenntnis des vollständigen Ereigniskörpers voraussetzt, wird für das Glaubwürdigkeitsmaß zunächst nur eine sehr grobe Struktur für den frame of discernment, der zu den einzelnen vorliegenden Erkenntnissen definiert werden muß, unterstellt. Eine Verfeinerung ist ausschließlich dann notwendig, wenn Erkenntnisse zu einzelnen Aussagen, die durch eine logische Konjunktion verknüpft sind, zusammengefaßt werden sollen. In den übrigen Fällen wird ausschließlich die Kenntnis der jeweiligen Kompatibilitätsrelation vorausgesetzt, um die grundlegenden Wahrscheinlichkeitszuordnungen zu übertragen[5]. Da der Prüfer Erkenntnisse über die Struktur der Sachverhalte, die das Prüfungsurteil determinieren, oftmals erst während der Prüfung erlangt, spiegelt somit die zuletzt angeführte Vorgehensweise die Prüfungssituation besser wider als das bei der Verwendung eines Wahrscheinlichkeitsmaßes unterstellte a-priori-Wissen über die gesamte Problemstruktur. Darüber hinaus wird bei der Verwendung eines Glaubwürdigkeitsmaßes zugelassen, daß Erkenntnisse, die selbst nicht sicher sind, in die Risikomessung einbezogen werden.

Grundsätzlich kann aus dem Prüfungsergebnis sowohl bei den auf Zufallsstichproben als auch bei den auf bewußt ausgewählten Stichprobenelementen beruhenden Prüfungsverfahren ein Glaubwürdigkeitsmaß für die Ordnungsmäßigkeit der Prüfungsgesamtheit abgeleitet werden. Im Rahmen der Stichprobenverfahren mit bewußter Auswahl ist dann allerdings eine Wahrscheinlichkeits- oder Glaubwürdigkeitseinschätzung über die nicht in der Stichprobe enthaltenen Elemente der Prüfungsgesamtheit notwendig. Diese kann grundsätzlich aus Erkenntnissen über Zuverlässigkeit des datenerzeugenden Systems für bestimmte Bereiche der Prüfungsgesamtheit bestimmt werden. Allerdings fehlt es bisher sowohl an einer geeigneten Methode zur Messung der Zuverlässigkeit des datenerzeugenden Systems als auch an empirischen Ergebnissen oder Konventionen, die es erlauben, aus dem Zuverlässigkeitsgrad des betrachteten Systems auf den Fehleranteil in der Prüfungsgesamtheit zu schließen. Insoweit muß der Prüfer die für die Prämissen bei den Verfahren der bewußten Auswahl

---

4 Siehe oben Kapitel A. I. 1. a) im Dritten Teil der Arbeit.

5 Siehe oben Kapitel C. im Dritten Teil der Arbeit.

DOI https://doi.org/10.3790/978-3-428-48909-1

benötigten Glaubwürdigkeitsbeurteilungen im eigenen Ermessen festlegen, solange keine geeigneteren Systemprüfungsmethoden zur Verfügung stehen.

II. Überprüfbarkeit der Risikomessung

Im Hinblick auf die intersubjektive Überprüfbarkeit der gemessenen Risiken sind zwei Fragen zu unterscheiden. Zum einen die Frage, inwieweit die Unschärfe der einzelnen Prüfungsergebnisse objektiv gemessen werden kann, zum anderen die Frage, ob ein anerkanntes Verfahren zur Aggregation der im einzelnen gemessenen Unschärfe existiert. Für das vorliegende Prüfungsmodell ist festzustellen, daß zur Messung der Unschärfe für die einzelnen Prüfungsergebnisse kein objektives Unschärfemaß zur Verfügung steht. Zwar liefern die mathematisch-statistischen Verfahren objektive Wahrscheinlichkeiten für das Eintreten bestimmter Stichprobenergebnisse. Die Entscheidung, ob die Voraussetzungen, die an die Anwendung der Verfahren gestellt werden, zutreffen und eine Verwendung des Konfidenzniveaus als Glaubwürdigkeitsmaß zulässig ist, muß vom Prüfer getroffen werden, so daß insoweit die resultierende Glaubwürdigkeitseinschätzung subjektiv ist. Eine Objektivierung im Sinn einer Konventionalisierung ist ausschließlich dann möglich, wenn feste Regeln für die Anwendbarkeit der Verfahren etabliert werden.

Die Objektivierung der Ergebnisse von Verfahren mit bewußter Auswahl der Stichprobenelemente erfordert intersubjektiv überprüfbare Aussagen über die nicht geprüften Bereiche der Prüfungsgesamtheit. Diese können zum einen durch mathematisch-statistische Verfahren erlangt werden. Zum anderen wurden Konzeptionen gezeigt, die, falls Verfahren zur Messung der Zuverlässigkeit des datenerzeugenden Systems gefunden werden, eine unscharfe Formulierung des zweifelsohne bestehenden Zusammenhang zwischen dem Grad der Zuverlässigkeit des datenerzeugenden Systems und dem Fehleranteil in der Prüfungsgesamtheit erlauben. Soweit dabei allerdings Ansätze aus dem Bereich des plausiblen Schließens Anwendung finden sollen, fehlen insbesondere die theoretischen Grundlagen für die Auswahl von geeigneten Implikationsoperatoren. Eine Auswahl von Operatoren nach den damit erzielten Ergebnissen, wie sie im Rahmen der unscharfen Regelung praktiziert wird, scheidet für die realen Prüfungen aus, da zumindest im Hinblick auf die Ordnungsmäßigkeit der Prüfungsgesamtheit regelmäßig auch nach der Prüfungsdurchführung keine sichere Aussage gemacht werden kann, wenn nicht die Nichtordnungsmäßigkeit mit Sicherheit festgestellt wurde.

Zur Beantwortung der zweiten, oben angeführten Frage wurde im Rahmen der vorliegenden Arbeit zumindest ein Ansatz gezeigt, der eine Aggregation der einzelnen bei den Prüfungsergebnissen auftretenden Unschärfen zu einem Gesamtrisikomaß erlaubt. Insoweit wird die Ableitung des Prüfungsrisikos aus den einzelnen subjektiven Einschätzungen des Prüfers intersubjektiv überprüfbar. Darüber hinaus ist eine Überprüfung der Plausibilität einzelner Glaubwürdigkeitsbeurteilungen eher möglich, als die Überprüfung eines dem Prüfungsurteil insgesamt zugewiesenen Risikomaßes. Soweit die dem Glaubwürdigkeitsmaß zugrundeliegenden Axiome als Anforderungen anerkannt werden, die an eine Risikobeurteilung durch den Prüfer zu stellen sind, besteht auch die Möglichkeit, die Einhaltung der Axiome im Einzelfall zu prüfen.

Im Hinblick auf die Anwendung der hier vorgeschlagenen Vorgehensweise zur Risikomessung in realen Prüfungen darf aus dem Umstand, daß die Auswirkungen der Verknüpfung von Glaubwürdigkeiten ausschließlich anhand von sehr einfachen Beispielen untersucht wurden, nicht geschlossen werden, daß in komplexen Prüfungssituationen Glaubwürdigkeitsbeurteilungen nicht aggregiert werden können[6]. Ansätze zur Abbildung komplexer Glaubwürdigkeitsstrukturen mit Hilfe von Hypergraphen und Verfahren zur Berechnung der Glaubwürdigkeiten in solchen Hypergraphen auch bei größeren Datenmengen stehen zur Verfügung[7]. Voraussetzung ihrer Anwendung ist allerdings, daß der Prüfer die komplexen Zusammenhänge und Abhängigkeiten, die er zwischen einzelnen Erkenntnissen unterstellt, auch beschreiben kann. Insoweit wären zunächst empirische Untersuchungen notwendig, um diese komplexen Prüfungsstrukuren zu identifizieren.

B. Risikoorientierung der Prüfungsplanung

Neben der Frage der Risikomessung sollte die vorliegende Arbeit auch die Frage beantworten, ob mit Hilfe des ermittelten Risikomaßes eine am Risiko orientierte Planung der Prüfungshandlungen möglich ist. Wie die Ausführungen zur Risikomessung gezeigt haben, ist das Risiko außer von der jeweiligen Handlungsalternative insbesondere von den vorliegenden Prüfungsergebnissen abhängig. Im anfangs vorgestellten Entscheidungsmodell kommt dies durch die Abhängigkeit der Einschätzung über das Eintreten bestimmter Umweltzustände von den Prüfungsergebnissen zum Ausdruck. Es müßte demnach zunächst für alle möglichen Prüfungsergebnisse, diese entsprechen im Entscheidungsmodell

---

6 Vgl. *Sullivan*, Response, 1988, S. 80.

7 Vgl. *Shenoy/Shafer*, Axioms, 1990.

der Menge *I*, die Auswirkung auf die Glaubwürdigkeit der Handlungsalternativen festgestellt werden. Im Anschluß daran wären die Eintrittswahrscheinlichkeiten für die einzelnen Prüfungsergebnisse zu bestimmen. Eine begründete Prognose der Ergebnisse von Prüfungshandlungen für den Einzelfall ist ausschließlich dann möglich, wenn Informationen über den tatsächlichen Zustand der Prüfungsgesamtheit vorliegen. Da diese erst mit Hilfe der Prüfungshandlungen ermittelt werden sollen, erscheint die Prognose von Prüfungsergebnissen und damit die Abschätzung des Einflusses bestimmter Prüfungshandlungen auf das Risiko vor der Durchführung der Prüfungshandlungen nicht möglich.

Das angeführte Problem besteht dabei unabhängig davon, welches Prüfungsverfahren eingesetzt wird, da selbst dann, wenn der Prüfer eine bestimmte Ausprägung der Prüfungsgesamtheit vermutet, für keines der zur Verfügung stehenden Prüfungsverfahren angegeben werden kann, in welcher Höhe sich das Risiko verändern wird. Dies wird insbesondere bei den mathematisch-statistischen Schätzverfahren deutlich. Bei diesen ist, falls der Stichprobenumfang und das Konfidenzintervall vorgegeben sind, das Konfidenzniveau von der realisierten Ausprägung der Stichprobenfunktion abhängig und insoweit vor Ziehung der Stichprobe eine Zufallsgröße.

Insgesamt bleibt festzustellen, daß eine am Risiko orientierte Planung von Prüfungshandlungen i.S. einer Optimierung des Prüfungsprogramms nicht möglich ist. Allerdings besteht analog der Vorgehensweise bei sequentiellen Hypothesentests die Möglichkeit, ein ausreichend geringes Risiko für die Prüfung vorzugeben, bei dessen Erreichen die Prüfung beendet wird. Die hier vorgestellten Konzeptionen zur Risikomessung können insoweit zumindest ein Abbruchkriterium für einzelne Prüfungen liefern. Darüber hinaus können für den Fall, daß es gelingt, die einer Prüfung zugrundeliegenden Glaubwürdigkeitsstrukturen mit Hilfe von Hypergraphen EDV-gestützt abzubilden, What-If-Analysen durchgeführt werden, die zeigen, in welcher Größenordnung sich das gesamte Prüfungsrisiko verändert, wenn zu einzelnen Bereichen zusätzliche Erkenntnisse gewonnen werden. Insoweit könnte zumindest der Einsatz von Ressourcen in Bereichen, für die keine neuen Erkenntnisse mehr erforderlich sind, vermieden werden.

Literaturverzeichnis

Adenauer, P., [Kontrollsystem], Berücksichtigung des internen Kontrollsystems bei Jahresabschlußprüfung, Bergisch Gladbach u.a. 1989.

Adler/Düring/Schmaltz, Rechnungslegung und Prüfung der Unternehmen, Kommentar zum HGB, AktG, GmbHG, PublG nach den Vorschriften des Bilanzrichtlinien-Gesetzes, hrsg. von Karl-Heinz Forster u.a., 5. Aufl., Stuttgart 1987.

Akresh, A. D., [Response], Discussant's Response to "Analytical Procedure Results as Substantive Evidence", in: Proceedings of the 1990 Touche Ross University of Kansas Symposium on Auditing Problems, edited by Rajendra P. Srivastava, Lawrence 1990, S. 104-108.

Alderman, C. W./Tabor, R. H., [risk-driven audits], The case for risk-driven audits, in: Journal of Accountancy Vol. 167, No. 3 (1989), S. 55-61.

American Institute of Certified Public Accountants (AICPA), [SAS No. 39], Statement on Auditing Standards No. 39, Audit Sampling, 1981, in: AICPA Codification of Statements on Auditing Standards, edited by American Institute of Certified Public Accountants, Chicago, Illinois 1994, AU Sec. 350.

– [SAS No. 47], Statement on Auditing Standards No. 47, Audit Risk and Materiality in Conducting an Audit, 1983, in: AICPA Codification of Statements on Auditing Standards, edited by American Institute of Certified Public Accountants, Chicago, Illinois 1994, AU Sec. 312.

– [SAS No. 56], Statement on Auditing Standards No. 56, Analytical Procedures, 1988, in: AICPA Codification of Statements on Auditing Standards, edited by American Institute of Certified Public Accountants, Chicago, Illinois 1994, AU Sec. 329.

Ballwieser, W., [Informationsökonomie], Informationsökonomie, Rechnungslegungstheorie und Bilanzrichtlinie-Gesetz, in: ZfbF 37. Jg. (1985), S. 47-66.

Bamberg, G./Bauer, F., [Statistik], Statistik, 9. Aufl., München, Wien 1996.

Bandemer, H./Gottwald, S., [Einführung], Einführung in Fuzzy-Methoden, 4. Aufl., Berlin 1993.

Blocher, E./Willingham, J. J., [Review], Analytical Review, New York u. a. 1985.

Bochenski, I.M./Menne, A., [Grundriß], Grundriß der formalen Logik, 5. Aufl., Paderborn 1983.

Buchner, R., [Diskussion], Zur Diskussion um die Frage "Zufalls- oder Urteilsstichprobe" bei Buchprüfungen, in: ZfbF 35. Jg. (1983), S. 478-502.

– [Prüfungswesen], Wirtschaftliches Prüfungswesen, 2. Aufl., München 1997.

Bücker, R., [Lösungsmöglichkeiten], Lösungsmöglichkeiten von Entscheidungsproblemen unter Anwendung der Bayesschen Entscheidungslehre, Diss. Münster 1973.

Bühler, W., [Stichprobenumfang], Zum optimalen Stichprobenumfang bei der Stichprobeninventur, in: ZfbF 36. Jg. (1984), S. 699-722.

Clemm/Nonnenmacher, [Beck Bil.-Komm.], Kommentierung zu § 249 HGB, in: Beck Bil.-Komm., 3. Aufl., München 1995.

Cochran, W. G., [Sampling], Sampling Techniques, 3rd ed., New York u. a. 1977.

– [Stichprobenverfahren], Stichprobenverfahren, Berlin, New York 1972.

Coenenberg, A. G./Hanisch, H., [Stichprobenprüfung], Stichprobenprüfung, Entdekkungsstichprobe, in: HWRev, 2. Aufl. Stuttgart 1992, Sp. 1862-1874.

Cooley, J. W./Hicks, J. O., [Fuzzy Set Approach], A Fuzzy Set Approach to Aggregating Internal Control Judgments, in: Management Science Vol. 29, No. 3 (1983), S. 317-334.

Cushing, B. E., [Mathematical Approach], A Mathematical Approach to the Analysis and Design of Internal Control Systems, in: The Accounting Review Vol. 49 (1974), S. 24-41.

Cushing, B. E./Loebbecke, J. K., [Audit Risk], Analytical Approaches to Audit Risk: A Survey and Analysis, in: Auditing: A Journal of Practice & Theory Vol. 3, No. 1 Fall (1983), S. 23-41.

De Finetti, B., [Theory], Theory of Probability, Vol. 1, London u. a. 1974.

Demant, B., [Fuzzy-Theorie], Fuzzy-Theorie oder die Faszination des Vagen, Braunschweig u.a. 1993.

Dempster, A. P., [Probabilities], Upper and Lower Probabilities induced by a Multivalued Mapping, in: Annals of Mathematical Statistics Vol. 38 (1967), S. 325-339.

Diehl, C.-U., [Prüfungsvorgehen], Strukturiertes Prüfungsvorgehen durch risikoorientierte Abschlußprüfung, in: Aktuelle Fachbeiträge aus Wirtschaftsprüfung und Beratung, Festschrift zum 65. Geburtstag von Prof. Dr. Hans Luik, hrsg. von Schitag Ernst & Young-Gruppe, Stuttgart 1991, S. 187-215.

Dörner, D., [Audit Risk], Audit Risk, in: HWRev, 2. Aufl. Stuttgart 1992, Sp. 81-95.

Dubois, D./Prade, H., [Fuzzy Sets], Fuzzy Sets and Systems: Theory and Applications, New York 1980.

– [Statistical Data], Fuzzy Sets and Statistical Data, in: European Journal of Operational Research Vol. 25 (1986), S. 345-356.

– [Properties], Properties of Measures of Information in Evidence and Possibility Theories, in: Fuzzy Sets and Systems Vol. 24 (1987), S. 161-182.

– [Possibility], Possibility Theory, New York 1988.

– [Fuzzy Sets], Fuzzy sets, probability and measurement, in: European Journal of Operational Research Vol. 40 (1989), S. 135-154.

– [Approximate Reasoning], Fuzzy sets in approximate reasoning, in: Fuzzy Sets and Systems Vol. 40 (1991), S. 143-202.

Egner, H., [Prüfungslehre], Betriebswirtschaftliche Prüfungslehre, Berlin, New York 1980.

– [Prüfungstheorie], Prüfungstheorie, verhaltensorientierter Ansatz (Syllogistischer Ansatz), in: HWRev, 2. Aufl., Stuttgart 1992, Sp. 1566-1578.

Elliot, R. K./Rogers, J. R., [Sampling], Relating Statistical Sampling to Audit Objectives, in: Journal of Accountancy, July 1972, S. 46-55.

Fishburn, P. C., [Utility Theory], Utility Theory for Decision Making, New York u. a. 1970.

Fisz, M., [Wahrscheinlichkeitsrechnung], Wahrscheinlichkeitsrechnung und mathematische Statistik, 9. Aufl., Berlin 1978.

Förschle/Kofahl, [Beck Bil.-Komm.], Kommentierung zu § 247 HGB, in: Beck Bil.-Komm., 3. Aufl., München 1995.

Gans, C., [Prüfungen], Betriebswirtschaftliche Prüfungen als heuristische Suchprozesse, Bergisch Gladbach, Köln 1986.

Gärtner, M., [Prüfungshandlungen], Analytische Prüfungshandlungen im Rahmen der Jahresabschlußprüfung - Ein Grundsatz ordnungsmäßiger Abschlußprüfung, Marburg 1994.

Göbel, S., [Prüfung], Prüfung von EDV-Programmsystemen im Rahmen der Jahresabschlußprüfung, zugl. Diss. Regensburg 1989, Düsseldorf 1990.

Hagest, J., [Logik], Zur Logik der prüferischen Urteilsbildung bei Jahresabschlussprüfungen, Diss. München 1975.

– [Urteilsstichprobe], Die Urteilsstichprobe des Abschlußprüfers, eine Stichprobe zweiter Klasse?, in: Praxis des Prüfungswesens, hrsg. von Klaus v. Wysocki und Joachim Hagest, München 1976, S. 113-127.

Halpern, J. J./Fagin, R., [View], Two Views of Belief: Belief as Generalized Probability and Belief as Evidence, in: Artificial Intelligence Vol. 54 (1992), S. 275-317.

Hömberg, R., [Prüfungen], Betriebswirtschaftliche Prüfungen als heuristische Suchprozesse, Habilitationsschrift Wien 1981.

– [Prüfung], Prüfung, externe, in: HWB, 5. Aufl., Stuttgart 1993, Sp. 3570-3583.

Institut der Wirtschaftsprüfer in Deutschland e.V. (IDW), [HFA 1/1988], HFA Stellungnahme 1/1988: Zur Anwendung stichprobengestützter Prüfungsmethoden bei der Jahresabschlußprüfung, in: WPg, 41. Jg. (1988), S. 240-247.

– (IDW), [FG 1/1988], Fachgutachten 1/1988: Grundsätze ordnungsmäßiger Durchführung von Abschlußprüfungen, in: WPg 42. Jg. (1989), S. 9-19.

– (IDW), [FG 3/1988], Fachgutachten 3/1988: Grundsätze ordnungsmäßiger Durchführung von Abschlußprüfungen, in: WPg 42. Jg. (1989), S. 27-36.

– (IDW), [WP-Handbuch], Wirtschaftsprüfer-Handbuch 1996, Handbuch für Rechnungslegung, Prüfung und Beratung, Bd. I, bearb. von Wolfgang Dieter Budde u.a., 11. Aufl., Düsseldorf 1996.

Kaplan, S./Garrick, B. J., [Bestimmung], Die quantitative Bestimmung von Risiko, in: Die quantitative Bestimmung von Risiko - Grundlagen und Ergebnisse interdisziplinärer Risikoforschung, hrsg. von Gotthard Bechmann, Opladen 1993, S. 91-124.

Keynes, J. M., [Treatise], A Treatise on Probability, London 1963.

Kinney, W. R. Jr./Haynes, C. M., [Analytical Procedure], Analytical Procedure Results as Substantive Evidence, in: Proceedings of the 1990 Touche Ross University of Kansas Symposium on Auditing Problems, edited by Rajendra P. Srivastava, Lawrence 1990.

Klir, G. J./Folger, T. A., [Fuzzy Sets], Fuzzy Sets, Uncertainty and Information, Englewood Cliffs 1988.

Knop, W., [Planung], Eine Möglichkeit zur optimalen Planung einer einzelnen Jahresabschlußprüfung unter besonderer Berücksichtigung der Beurteilung des internen Kontrollsystems, Thun u.a. 1983.

Koch, H., [Unsicherheit], Unsicherheit, Techniken zur Handhabung von, in: Handwörterbuch der Planung, hrsg. von Norbert Szyperski, Stuttgart 1989, Sp. 2060-2073.

Ködel, W., [Abschlußprüfung], Risikoorientierte Abschlußprüfung, Integration in das Risiko-Management von Prüfungsunternehmen, zugl. Diss. München 1996, Wiesbaden 1997.

Kolarik, F. G., [Buchprüfung], Die Buchprüfung als Allokationsproblem von heterograden Wahrscheinlichkeitsstichproben, Diss. Wien 1964.

Kolmogoroff, A., [Grundbegriffe], Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin 1933, Reprint Berlin u.a. 1973.

Konrath, L. F., [Auditing], Auditing Concepts and Applications - A Risk-Analysis Approach, 2nd ed., St. Paul u. a. 1993.

Krantz, D. H./Luce, R. D./Suppes, P./Tversky. A., [Measurement], Foundations of Measurement, Vol. I: Additive and Polynomial Representations, New York, London 1971.

Kreutzfeldt, R. W., [Response], Discussant's Response to "Assessing Control Risk: Effects of Procedural Differences on Auditor Consensus", in: Proceedings of the 1990 Touche Ross University of Kansas Symposium on Auditing Problems, edited by Rajendra P. Srivastava, Lawrence 1990 , S. 132-146.

Kruse, R./Gebhardt, J./Klawonn, F., [Fuzzy-Systeme], Fuzzy-Systeme, Stuttgart 1993.

Kruse, R./Gebhardt, J./Klawonn, F., [Fuzzy-Systeme], Fuzzy-Systeme, 2. Aufl., Stuttgart 1995.

Kupsch, P., [BHR], Kommentierung zu § 249 HGB, in: BHR, Bonn 1986.

– [BHR], Kommentierung zu § 246 HGB, in: BHR, Bonn 1986.

v. Kutschera, F., [Wissenschaftstheorie], Wissenschaftstheorie, München 1972.

Lachnit, L., [Globalabstimmung], Globalabstimmung und Verprobung, in: HWRev, 2. Aufl., Stuttgart 1992, Sp. 719-742.

Lanfermann, J., [Stichprobenprüfung], Stichprobenprüfung, bewußte Auswahl, in: HWRev, 2. Aufl., Stuttgart 1992, Sp. 1855-1862.

Lang, J., [Grundsätze], Grundsätze ordnungsmäßiger Buchführung II - Überblick über die Hauptgrundsätze ordnungsmäßiger Buchführung -, in: Handwörterbuch unbestimmter Rechtsbegriffe im Bilanzrecht des HGB, hrsg. von Ulrich Leffson u. a., Köln 1986, S. 240-246.

Laux, H., [Entscheidungstheorie I], Entscheidungstheorie I - Grundlagen, 2. Aufl., Berlin u.a. 1991.

Leffson, U., [Wesentlich], Wesentlich, in: Handwörterbuch unbestimmter Rechtsbegriffe im Bilanzrecht des HGB, hrsg. von Ulrich Leffson u. a., Köln 1986, S. 434-447.

– [Grundsätze], Die Grundsätze ordnungsmäßiger Buchführung, 7. Aufl., Düsseldorf 1987.

– [Wirtschaftsprüfung], Wirtschaftsprüfung, 4. Aufl., Wiesbaden 1988.

Leffson, U./Bönkhoff, F. J., [Materiality-Entscheidungen], Zu Materiality-Entscheidungen bei Jahresabschlußprüfungen, in: WPg, 35. Jg. (1982), S. 389-397.

Leffson, U./Lippmann, K./Baetge, J., [Sicherheit], Zur Sicherheit und Wirtschaftlichkeit der Urteilsbildung, Düsseldorf 1969.

Lenz, H., [Urteilsbegründung], Urteilsbegründung bei betriebswirtschaftlichen Prüfungen - Indirekte Prüfung als statistische Begründung rationaler Erwartungen, in: ZfB 59. Jg. (1989), S. 1353-1366.

Leslie, D. A., [Analysis], An Analysis of the Audit Framework Focusing on Inherent Risk and the Role of Statistical Sampling in Compliance testing, in: Proceedings of the 1984 Touche Ross University of Kansas Symposium on Auditing Problems, edited by Howard F. Stettler and N. Allen Ford, Lawrence 1984, S. 89-125.

Loitlsberger, E., [Treuhand], Treuhand- und Revisionswesen, 2. Aufl., Stuttgart 1966.

– [Fehlergewichtung], Fehlergewichtung als Problem der Prüfungstheorie, in: Der Wirtschaftsprüfer im Schnittpunkt nationaler und internationaler Entwicklungen, Festschrift für Klaus v. Wysocki, hrsg. von Gerhard Gross, Düsseldorf 1985, S. 187-200.

– [Prüfungstheorie], Prüfungstheorie, spieltheoretischer Ansatz, in: HWRev, 2. Aufl., Stuttgart 1992, Sp. 1558-1565.

Luce, R. D./Raiffa, H., [Games], Games and Decisions - introduction and critical survey, New York, London, Sydney 1957.

Lück, W., [Jahresabschlußprüfung], Jahresabschlußprüfung, Stuttgart 1993.

Mackensen, L., [Deutsches Wörterbuch], Deutsches Wörterbuch, 10. Aufl., Köln u.a. 1982.

Mandl, G., [Auswahl], Zur Auswahl statistischer Stichprobenverfahren im heterograden Fall der Buchprüfung, in: Management und Kontrolle - Festgabe für Erich Loitlsberger zum 60. Geburtstag, hrsg. von Gerhard Seicht, Berlin 1981, S. 173-196.

– [Anwendungsvoraussetzungen], Untersuchungen über Anwendungsvoraussetzungen und Effizienz statistischer Stichprobenverfahren in der Buchprüfung, Wien 1984.

Minz, W., [Prüfungsmethoden], Prüfungsmethoden, in: WPg 13. Jg. (1960), S. 89-97.

v. Mises, R., [Wahrscheinlichkeit], Wahrscheinlichkeit, Statistik und Wahrheit, Wien, New York 1972.

Müller, W., [Risiko], Risiko und Ungewißheit, in: HWB, Stuttgart 1993, Sp. 3813-3825.

Niemeyer, G., [Systemtheorie], Kybernetische System- und Modelltheorie - system dynamics, München 1977.

Norwich, A. M./Turksen, I. B., [Model], A Model for the Measurement of Membership and the Consequences of its Empirical Implementation, in: Fuzzy Sets and Systems Vol. 12 (1984), S. 1-25.

Obermeier, I., [Abschlussprüfung], Statistische Abschlussprüfung, Bern, Stuttgart 1983.

Ossadnik, W., [Materiality], Materiality als Grundsatz externer Rechnungslegung, in: WPg, 48. Jg. (1995), S. 33-42.

Pearl, J., [Reasoning], Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, San Mateo 1988.

Popper, K., [Propensity], The Propensity Interpretation of Probability, in: British Journal for the Philosophy of Science Vol. 10 (1959), S. 25-42.

– [Logik], Logik der Forschung, 10. Aufl., Tübingen 1994.

Pougin, E., [Kontrollsystem], Die Berücksichtigung des internen Kontrollsystems als Grundlage ordnungsmäßiger Abschlussprüfung, Düsseldorf 1959.

Quick, R., [Risiken], Die Risiken der Jahresabschlußprüfung, Düsseldorf 1996.

Reuter, H. H., [Prüfungsumfang], Prüfungsumfang und Urteilsbildung im Rahmen einer Buchprüfung auf Stichprobenbasis, Frankfurt/M. 1975.

Roberts, D. A., [Statistical Auditing], Statistical Auditing, New York 1978.

Rommelfanger, H., [Decision], Fuzzy Decision Support-Systeme, Entscheiden bei Unschärfe, 2. Aufl., Berlin u. a. 1994.

Savage, L. J., [Foundations], The Foundations of Statistics, 2nd ed., New York 1972.

Schettler, K., [Planung], Planung der Jahresabschlußprüfung, Wiesbaden 1971.

Schildbach, T., [Stichprobenprüfung], Stichprobenprüfung, Annahmestichprobe, in: HWRev, 2. Aufl., Stuttgart 1992, Sp. 1846-1854.

Schneeweiß, H., [Entscheidungskriterien], Entscheidungskriterien bei Risiko, Berlin u.a. 1967.

– [Bemerkungen], Kritische Bemerkungen zur Kritik am Wahrscheinlichkeitssubjektivismus, in: Statistische Hefte 18. Jg. (1977), S. 218-232.

Schneider, D., [Meßbarkeitsstufen], Meßbarkeitsstufen subjektiver Wahrscheinlichkeiten als Erscheinungsformen der Ungewißheit, in: ZfbF 31. Jg. (1979), S. 89-122.

– [Informationstheorie], Informations- und Entscheidungstheorie, München, Wien 1995.

Schröder, M., [Zeitreihenprognose], Einführung in die kurzfristige Zeitreihenprognose und Vergleich der einzelnen Verfahren, in: Prognoserechnung, hrsg. von Peter Mertens, Würzburg, Wien 1994, S. 7-39.

Schulte, E. B., [Methoden], Quantitative Methoden der Urteilsgewinnung bei Unternehmensprüfungen, Düsseldorf 1970.

Srivastava, R./Shafer, G., [Belief-Function], Belief-Function Formulas for Audit Risk, in: The Accounting Review Vol. 67, No. April (1992), S. 249-283.

Selchert, F. W., [Begriff], Begriff und Prozeß betriebswirtschaftlicher Prüfungen, in: ZfbF 30. Jg. (1978), S. 125-145.

Shafer, G., [Theory], A Mathematical Theory of Evidence, Princeton and London 1976.

– [Nonadditive Probability], Nonadditive Probability, in: Encyclopedia of statistical sciences, Bd. 6 Multivariate analysis to Plackett and Burman designs, New York u. a. 1985.

– [Probability], Probability Judgement in Artificial Intelligence and Expert Systems, in: Statistical science Vol. 2. (1987), S. 3-44.

Shafer, G./Srivastava, R., [Bayesian], The Bayesian and Belief-Function Formalisms - A General Perspective for Auditing, in: Readings in Uncertain Reasoning, edited by Glenn Shafer and Judea Pearl, San Mateo 1990, S. 482-521.

Shafer, G./Shenoy, P. P./Srivastava, R. P., [Auditor's Assistant], AUDITOR´S ASSISTANT: A Knowledge Engineering Tool for Audit Decisions, in: Proceedings of the 1988 Touche Ross University of Kansas Symposium on Auditing Problems, edited by Rajendra P. Srivastava and James E. Rebele, Lawrence 1988, S. 61-79.

Shenoy, P. P./Shafer, G., [Axioms], Axioms for Probability and Belief-Function Propagation, in: Readings in Uncertain Reasoning, edited by Glenn Shafer and Judea Pearl, San Mateo 1990, S. 575-610.

Sieben, G./Bretzke, W.-R., [Typologie], Zur Typologie betriebswirtschaftlicher Prüfungssysteme, in: BFuP 25. Jg. (1973), S. 625-630.

Smets, P., [Belief Functions], Belief Functions, in: Non-Standard Logics for Automated Reasoning, edited by Phillipe Smets u. a., London u. a. 1988, S. 253-277.

Sperl, A., [Prüfungsplanung], Prüfungsplanung, Düsseldorf 1978.

Spies, M., [Unsicheres Wissen], Unsicheres Wissen - Wahrscheinlichkeit, Fuzzy-Logik, neuronale Netze und menschliches Denken, Heidelberg u. a. 1993.

Stegmüller, W., [Wahrscheinlichkeit, 1. Halbband], Personelle und Statistische Wahrscheinlichkeit, 1. Halbband, Berlin u. a. 1973.

– [Wahrscheinlichkeit, 2. Halbband], Personelle und Statistische Wahrscheinlichkeit, 2. Halbband, Berlin u. a. 1973.

Stibi, E., [Prüfungsrisikomodell], Prüfungsrisikomodell und risikoorientierte Abschlußprüfung, zugl. Diss. Göttingen 1994, Düsseldorf 1995.

Strawser, J. R., [Information Processing], Human Information Processing and the Consistency of Audit Risk Judgments, in: Accounting and Business Research Vol. 21, No. 81 (1990), S. 67-75.

Sullivan, J. B., [Response], Discussant's Response to "AUDITOR´S ASSISTANT: A Knowledge Engineering Tool for Audit Decisions", in: Proceedings of the 1988 Touche Ross University of Kansas Symposium on Auditing Problems, edited by Rajendra P. Srivastava and James E. Rebele, Lawrence 1988, S. 80-83.

Suppes, P./Krantz, D. H./Luce, R. D./Tversky. A., [Measurement], Foundations of Measurement, Vol. II: Geometrical, Threshold, and Probabilistic Representations, San Diego u. a. 1989.

Tarski, A., [Einführung], Einführung in die mathematische Logik, 5. Aufl., Göttingen 1971.

Thoennes, H. O., [Prüfungsansatz], Der risikoorientierte Prüfungsansatz, in: Rechnungslegung und Prüfung 1994, hrsg. von Jörg Baetge, Düsseldorf 1994, S. 31-51.

Tilli, T., [Fuzzy-Logik], Fuzzy-Logik, 2. Aufl., München 1992.

Tversky, A./Kahneman, D., [Judgement], Judgement under Uncertainty: Heuristics and Biases, in: Science Vol. 185 (1974), S. 1124-1131.

Werners, B., [Inferenz], Approximative Inferenz mit linguistischen Variablen, in: Operations Research - Reflexionen aus Theorie und Praxis - Festschrift zum 60. Geburtstag von Hans-Jürgen Zimmermann, hrsg. von Brigitte Werners, und Roland Gabriel, Berlin u. a. 1994.

Wiedmann, H., [Prüfungsansatz], Der risikoorientierte Prüfungsansatz, in: WPg 46. Jg. (1993), S. 13-25.

Wirtschaftsprüferkammer (WPK), [Bericht], Bericht über die Durchsicht der im Bundesanzeiger 1994 veröffentlichten und hinterlegten Abschlüsse sowie Zusammenstellung der Einschränkungen und Zusätze in Bestätigungsvermerken durch die Wirtschaftsprüferkammer, in: WPK-Mitteilungen Beilage zu Heft Nr. 3 (1995).

DOI https://doi.org/10.3790/978-3-428-48909-1

Wittmann, A., [Systemprüfung], Systemprüfung und ergebnisorientierte Prüfung - Die Systemprüfung als Grundlage der ergebnisorientierten Prüfung im Rahmen der aktienrechtlichen Jahresabschlußprüfung, Diss. Regensburg 1980.

Würtele, G., [Operationalisierung], Die Operationalisierung des Grundsatzes der Materiality bei Abschlußprüfungen, Pfaffenweiler 1989.

v. Wysocki, K., [meßtheoretische Ansatz], Der meßtheoretische Ansatz einer Prüfungslehre, in: Treuhandwesen, Prüfung, Begutachtung, Beratung, hrsg. von K. Lechner, Wien 1978, S. 105-168.

– [Auswahl], Auswahl von Prüfelementen bei der Jahresabschlußprüfung, in: Der Schweizer Treuhänder 60. Jg. (1986), S. 391-396.

– [Grundlagen], Grundlagen des betriebswirtschaftlichen Prüfungswesens, 3. Aufl., München 1988.

– [Prüfungstheorie], Prüfungstheorie, meßtheoretischer Ansatz, in: HWRev, 2. Aufl., Stuttgart 1992, Sp. 1545-1557.

– [Wirtschaftlichkeit], Wirtschaftlichkeit von Prüfungen, in: HWRev, 2. Aufl., Stuttgart 1992, Sp. 2171-2180.

Yu, S., [Model], A Markovian Model for the Review of the Internal Control System, Diss. University of Minnesota 1972.

Zadeh, L. A., [Fuzzy Sets], Fuzzy Sets, in: Information and Control Vol. 8 (1965), S. 338-353.

– [Concept, Part 2], The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Part 2, in: Information Sciences Vol. 8 (1975), S. 301-357, abgedruckt in: Fuzzy Sets and Applications, selected Papers by L. A. Zadeh, edited by R. R. Yager u. a., New York u. a. 1987, S. 271-327.

– [Concept, Part 3], The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Part 3, in: Information Sciences Vol. 9 (1976), S. 43-80.

– [Possibility], Fuzzy Sets as a Basis for a Theory of Possibility, in: Fuzzy Sets and Systems Vol. 1 (1978), S. 3-28.

– [Theory], A Theory of Approximate Reasoning, in: Machine Intelligence 9, edited by J. E. Hayes, Donald Michie and L. I. Mikulich, New York u. a. 1979, S. 149-194.

Zimmer, A., [Risiko], Warum müssen Wissenschaftler über Risiko reden? Und welches Risiko gehen sie damit ein?, in: Blick in die Wissenschaft - Forschungsmagazin der Universität Regensburg 3. Jg. Nr. 5 (1994), S. 4-13.

Zimmermann, H.-J., [Fuzzy Set Theory], Fuzzy Set Theory and its Applications, 2nd ed., Boston u. a. 1991.

– [Stichprobenprüfung], Stichprobenprüfung, Schätzstichprobe, in: HWRev, 2. Aufl., Stuttgart 1992, Sp. 1874-1881.

Zimmermann, H.-J./Zysno, P., [Connectives], Latent Connectives in Human Decision Making, in: Fuzzy Sets and Systems Vol. 4 (1980), S. 37-51.

– [Decisions], Decisions and Evaluations by Hierarchical Aggregation of Information, in: Fuzzy Sets and Systems Vol. 10 (1983), S. 243-260.

Sachregister

DOI https://doi.org/10.3790/978-3-428-48909-1