

Delphi und Kognitionspsychologie: ein Zugang zur theoretischen Fundierung der Delphi-Methode

Häder, Michael; Häder, Sabine

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Häder, M., & Häder, S. (1995). Delphi und Kognitionspsychologie: ein Zugang zur theoretischen Fundierung der Delphi-Methode. *ZUMA Nachrichten*, 19(37), 8-34. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-208882>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

DELPHI UND KOGNITIONSPSYCHOLOGIE: EIN ZUGANG ZUR THEORETISCHEN FUNDIERUNG DER DELPHI-METHODE*)

MICHAEL HÄDER UND SABINE HÄDER

Die Delphi-Methode ist ein vergleichsweise stark strukturierter Gruppenkommunikationsprozeß, in dessen Verlauf Sachverhalte, über die unsicheres und unvollständiges Wissen existiert, von Experten beurteilt werden. Die Delphi-Methode wird in den deutschen Sozialwissenschaften bislang kaum angewendet. Der Grund für diese Zurückhaltung dürfte u.a. in der vorhandenen Ungewißheit über die theoretischen Grundlagen und methodischen Prinzipien bei der Anwendung der Methode zu finden sein. Solche Vorbehalte können nur durch eine Evaluation der Delphi-Methode und durch empirische Tests abgebaut werden. Dies gilt sowohl für die Behebung des Theoriedefizits als auch für die methodologische Aufarbeitung. Daraufhin können sich u. E. spezifische Gebiete im Rahmen der sozialwissenschaftlichen Forschung erschließen, für die der Einsatz von Delphi geeignet ist. Der Artikel wird zeigen, daß die Nutzung kognitionspsychologischer Erkenntnisse, insbesondere die Theorie mentaler Modelle, neue Möglichkeiten eröffnet, um einige wesentliche Aspekte der Funktionsweise von Delphi zu erklären. Anschließend werden die Ergebnisse eines Experiments zur empirischen Validierung unserer theoretischen Annahmen vorgestellt.

The Delphi method is a highly structured communication process performed by groups of experts. The method focuses on facts about which we have only uncertain or incomplete knowledge. These facts are evaluated by the involved experts in the course of the Delphi process. The Delphi method has been rarely used in the German social sciences until now. One of the reasons for this might be the uncertainty about the theoretical basis and about the methodological rules of this method. This uncertainty can be reduced only by an evaluation of Delphi and through empirical tests. The theoretical deficit must be removed and the methodological basis must be clarified. After this, it should be possible to solve specific problems in the social sciences with the Delphi method. The article will explain important aspects of Delphi through the application of

results from social cognition research. In this context, the theory of mental models is a particularly useful instrument. Following the theoretical introduction, we will present the results of an experiment for the validation of our assumptions.

1. Die Delphi-Methode in den Sozialwissenschaften - eine Situationsbeschreibung

Der erste Verweis auf die Benutzung eines Delphi-Ansatzes stammt aus dem Jahr 1948. Damals soll diese Technik dazu benutzt worden sein, um die Ergebnisse eines Hund- oder Pferderennens¹⁾ vorauszusagen. In den folgenden Jahren wurde die Delphi-Methode von der RAND Corporation in 14 Experimenten für militärische Zwecke eingesetzt, beispielsweise um mögliche Ziele sowjetischer Angriffe auf die USA vorherzusagen. Wissenschaftliche Ergebnisse dieser Studien sind jedoch nicht publiziert worden (vgl. Linstone/Turoff 1975: 10). Bekannt wurde die Delphi-Methode schließlich durch einen 1964 ebenfalls von der RAND Corporation erarbeiteten "Report on a Long-Range Forecasting Study" (Gordon/Helmer 1964). Das Ziel dieser Studie bestand in der langfristigen Vorhersage wissenschaftlicher und technischer Entwicklungen, wobei eine Zeitspanne von zehn bis fünfzig Jahren als langfristig definiert wurde. Diese Untersuchung wird in der deutschsprachigen Literatur oft als die erste bezeichnet (vgl. Albach 1970; Geschka 1977; Saliger/Kunz 1981). Seit den 70er Jahren fand dann eine Ausbreitung der Delphi-Methode auch in Westeuropa - einschließlich Deutschland - statt. Die Anlagen dieser Folgestudien waren zunächst stark vom erwähnten RAND Corporation-Projekt beeinflusst (vgl. Helmer/Rescher 1960). Insbesondere in der Betriebswirtschaft kam es zur Rezeption der Delphi-Methode für Prognosezwecke. Die Einsatzgebiete wurden aber schnell vielfältiger, so daß Seeger bereits 1979 schätzte, es habe innerhalb der 15jährigen Anwendungsdauer circa 1500 Delphi-Untersuchungen verschiedenster Art gegeben (vgl. Seeger 1979: 32).

In einer Literaturstudie (vgl. Häder/Häder 1994a: 3) konnte gezeigt werden, daß die Sozialwissenschaften in Deutschland - anders als in den USA und auch im Unterschied zu Japan - bisher vergleichsweise wenig von der Delphi-Methode Gebrauch gemacht haben. Ein Indiz für die Zurückhaltung gegenüber der Delphi-Methode ist beispielsweise der Umfang, mit dem diese Technik in der aktuellen Studienliteratur im Fach „Methoden der empirischen Sozialforschung“ behandelt wird. Nur in zwei von zehn daraufhin untersuchten Fachlehrbüchern findet sie überhaupt Erwähnung - und dies lediglich auf wenigen Zeilen.

2. Der Stand der Erforschung der Delphi-Methode

2.1 Empirische Befunde zum Funktionieren von Delphi

Sowohl die internationale sozialwissenschaftliche als auch die deutsche betriebswirtschaftliche Literatur liefert eine Vielzahl von Hinweisen auf erfolgreiche Anwendungen der Delphi-Methode. Die beiden folgenden Zitate sollen zunächst stellvertretend für zahlreiche andere stehen:

"Die Delphi-Methode kann also insgesamt als sehr leistungsfähiges Vorhersageverfahren angesehen werden. Das ist ihr größter Vorteil" (Kaufmann 1972: 98).

"Das heute verfügbare bestmögliche Instrumentarium zur langfristigen Technikvorschau sind Delphi-Untersuchungen" (BMFT 1993: XI).

Diese positiven Urteile über die Leistungsfähigkeit der Methode stützen sich auf Ergebnisse, die bei Studien zur empirischen Validierung der Ergebnisse von Delphi-Befragungen gefunden worden sind²⁾. Dabei wurden bisher vor allem zwei Wege genutzt: Erstens wurde mit Almanachfragen³⁾ überprüft, ob mit der Delphi-Methode eine Annäherung an die richtigen Antworten erzielt werden kann. Die RAND Corporation erarbeitete dazu zwei experimentelle Panels (vgl. Dalkey 1969, Albach 1970, Geschka 1977). Im ersten Panel wurden 14 Gruppen insgesamt 350 Almanachfragen vorgelegt, im zweiten wurden 16 Gruppen insgesamt 160 Fragen gestellt. Die Versuchspersonen, Studenten der Universität von Kalifornien in Los Angeles, sollten Schätzungen zu Sachverhalten abgeben, deren tatsächliche Ausprägungen ihnen nicht bekannt waren. Die erfragten Sachverhalte waren so allgemein, daß die Studenten sie auf der Basis ihres Allgemeinwissens schätzen konnten. Zusammengefaßt erbrachten diese Tests u.a. eine Konvergenz der Meinungen in Richtung auf den wahren Wert, einen Rückgang der Schätzfehler von der ersten zur letzten Befragungsrunde und eine weitere Verbesserung des Gruppenurteils, wenn in den Rückmeldungen die Streuung der Antworten mitgeteilt wurde (vgl. auch Geschka 1977: 34; Albach 1970; Becker 1974).

Zweitens können die Ergebnisse zurückliegender Prognose-Delphis daraufhin untersucht werden, ob die vorausgesagten Ereignisse inzwischen eingetreten sind. Ein Beispiel für diese Art der Validierung ist eine japanische Studie von 1971. Gegenstand dieser Untersuchung war eine Beurteilung der künftigen Entwicklung von Wissenschaft und Technik. Hier ließen sich im Jahr 1991 bereits 530 Fragen überprüfen. „Unter den beurteilungsfähigen damaligen Voraussagen sind fast 30% bis zum Jahr 1991 in vollem Umfang eingetroffen. ... Andere Voraussagen waren nur teilweise (ein weiteres Drittel) richtig. ... Zusammenfassend erscheint der Verlässlichkeitsgrad der Delphi-Untersuchung erstaunlich hoch zu sein.“ (BMFT 1993:XXff.) Auch mittelfristige Vorhersagen, z.B. zur Zinsent-

wicklung in einem Zeitraum von drei bzw. sechs Monaten konnten von anderen Autoren inzwischen erfolgreich validiert werden (vgl. Brockhoff 1979).

In zwei weiteren Tests bei ZUMA und mit Studenten der Universität Marburg zur empirischen Validierung des Delphi-Verfahrens wurden Experten um die Schätzung von ihnen nicht bekannten Ergebnissen einer Bevölkerungsbefragung gebeten (vgl. Häder/Häder 1994b; Häder/Häder/Ziegler 1995). Es zeigte sich, daß es mit Hilfe einer Delphi-Befragung möglich ist, Aussagen über die Antworten bei einer repräsentativen Befragung zu treffen. Ein Vergleich der individuellen Schätzfehler über die Befragungsrunden ergab, daß sich bei allen eingesetzten Typen von Schätzaufgaben die Fehler in der dritten Welle gegenüber der ersten Welle verringert hatten.

2.2 Zur Kritik an Delphi

Trotz des breiten Spektrums von erfolgreichen Anwendungen wird die Diskussion um die Leistungsfähigkeit der Delphi-Methode auch Jahrzehnte nach ihrer Implementation in das Methodenarsenal noch kontrovers geführt. Unseres Erachtens betreffen die verschiedenen kritischen Positionen drei Grundprobleme.

1. Es mangelt nach wie vor generell an einer systematischen Beschreibung der Voraussetzungen und der Grenzen für die erfolgreiche Nutzung der Delphi-Methode. Der Einsatz von so ungewöhnlichen Begriffen wie „ideales Verfahren“ (BMFT 1993: 15) und „Allheilmittel“ (Brockhoff 1979: 166) zur Charakterisierung dieser empirischen Methode deutet schon formal auf die noch bestehende Ratlosigkeit bzw. auch auf Verwirrung beim Umgang mit der Delphi-Methode hin.⁴⁾

In der Diskussion konkreter Aspekte der Delphi-Methode spielt häufig eine Rolle, inwiefern der Delphi-Ansatz zu besseren Ergebnissen führt als der Einsatz alternativer empirischer Methoden. Daran wird mitunter die Frage nach der grundsätzlichen Legitimierung der Anwendung von Delphi für bestimmte Aufgabenstellungen geknüpft. Insbesondere wird eine Diskussion um das Verhältnis von Delphi-Erhebungen zu anderen Gruppeninteraktions- und Prognoseverfahren wie etwa Face-to-Face-Gruppendiskussion, Brainstorming und Expertenbefragungen geführt.

Ein erster Schritt zur systematischen Aufarbeitung der Stellung von Delphi im empirischen Methodenarsenal wäre u. E. eine Abgrenzung des inhaltlichen und methodischen Anwendungsraumes von Delphi im Sinne einer Definition, die die potentiellen Einsatzgebiete sachlich umreißt. In unserer Literaturstudie (vgl. Häder/Häder 1994a) konnten wir zeigen, daß bislang die Definitionen der Delphi-Methode bei unterschiedlichen Autoren ebenso variieren wie ihre Anwendungsgebiete. Sie reichen von eng begrenzten

Sichtweisen (A: „Die Delphi-Methode ist ein Verfahren, um aus Expertenmeinungen Prognosen zu gewinnen und Konsens und Dissens zwischen den Expertenmeinungen deutlich zu machen.“ Köhler 1992: 325) bis zu sehr allgemeinen Bestimmungen (B: „Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole to deal with a complex problem.“ Linstone/Turoff 1975: 3).

Ohne unseren Standpunkt an dieser Stelle ausführlich darlegen zu können, wollen wir doch auf die vorgenannten Begriffsbestimmungen kurz eingehen.

Zu A: Auffassungen, die Delphi lediglich als Prognoseinstrument beschreiben, folgen wir nicht. Wie wir noch zeigen werden, gibt es insbesondere in den Sozialwissenschaften eine Reihe weiterer Anwendungsmöglichkeiten. Darüber hinaus fehlen in dieser Art von Bestimmung Hinweise auf die Art und Weise der Meinungsbildung.

Zu B: Diese Charakterisierung gilt es u.E. wie folgt zu spezifizieren:

- Inhalte von Delphi-Studien sind stets Sachverhalte, über die *unsicheres* bzw. *unvollständiges Wissen* existiert. Andernfalls gäbe es effizientere Methoden zur Entscheidungsfindung.
- Bei Delphi handelt es sich um *Urteilsprozesse* unter Unsicherheit. Die an Delphi-Studien beteiligten Personen geben also jeweils Schätzungen ab.
- Für die Teilnahme an Delphi-Studien sind *Experten* zu rekrutieren, die aufgrund ihres Wissens und ihrer Erfahrungen in der Lage sind, kompetent zu urteilen⁵⁾.

Daraus leitet sich unsere folgende Arbeitsdefinition ab:

Die Delphi-Methode ist ein vergleichsweise stark strukturierter Gruppenkommunikationsprozeß, in dessen Verlauf Sachverhalte, über die naturgemäß unsicheres und unvollständiges Wissen existiert, von Experten beurteilt werden.

2. Ein weiterer wesentlicher Aspekt der Diskussion betrifft die Offenlegung der kognitiven Prozesse, die bei den in eine Delphi-Befragung einbezogenen Teilnehmern ablaufen. Es gilt theoretisch zu begründen, warum es den Experten gelingt, Urteile über Sachverhalte zu fällen, über die kein sicheres bzw. vollständiges Wissen existiert.

Dalkey, einer der Begründer der Delphi-Methode, ging bei seiner Rezeption des Verfahrens davon aus, daß in „n“ Köpfen mindestens so viel Information enthalten ist, wie in einem (vgl. 1969: 411), wahrscheinlich aber mehr, was letztlich den Einsatz von Delphi zur Wissensakkumulation legitimiere. Die Defizite einer solchen Bestimmung werden inzwischen kritisiert und es wird angemahnt, den Hintergrund offen zu legen, „auf dem die Befragten ihre Meinung aufbauen und ableiten“ (Seeger 1979: 151). Mit Hilfe der Nutzung

von neueren Forschungsergebnissen aus der Kognitionspsychologie scheint es uns möglich, zur Klärung dieses Problems einen Beitrag zu leisten (vgl. Abschnitt 3).

In diesem Zusammenhang könnte die folgende Textstelle als Motto eines in diese Richtung orientierten Forschungsprogramms betrachtet werden: „It is time for the oracles to move out and for science to move in“ (Wechsler 1978: 596).

3. Gegenstand von Diskussionen sind weiterhin die Konsequenzen verschiedener methodischer Varianten des Vorgehens bei Delphi-Befragungen. Es wurde bisher - mitunter relativ wahllos und ohne ausreichende methodische Begründung - nach immer neuen Variationen der Methode gesucht, von denen dann eine Verbesserung der Ergebnisse von Delphi-Befragungen erwartet wurde. Gerade diese praktizierte relative Beliebigkeit bei der Gestaltung der Delphi-Methode und der damit in Verbindung stehende Mangel an professionellen Standards machen dieses Verfahren für Kritiker jedoch besonders anfechtbar (vgl. auch Seeger 1979: 148ff.).

Zusammenfassend ist der Stand der Diskussion damit durch ein gewisses Dilemma - mitunter wird auch von einem „These-Antithese-Verhältnis“ gesprochen (vgl. Wechsler ebenda) - gekennzeichnet: Auf der einen Seite handelt es sich bei Delphi um ein offensichtlich (unter bestimmten Voraussetzungen) gut funktionierendes Instrument, welches bereits in vielen Fällen zu wertvollen Schätzungen geführt hat (vgl. Abschnitt 2.1.). Auf der anderen Seite wird prinzipiell kritisiert, daß es an Einsicht über die Prinzipien des Funktionierens von Delphi fehle. Neben der Frage, woher eigentlich das Wissen stammt, aufgrund dessen die Experten ihre Urteile fällen und wieso von Befragungswelle zu Befragungswelle mit einer Verbesserung der abgegebenen Schätzungen gerechnet werden kann, steht das Problem der Standardisierung des methodischen Vorgehens bei Delphi im Vordergrund. Darüber hinaus wird eine „Standortbestimmung“ für die Einsatzmöglichkeiten von Delphi-Studien gefordert.

3. Kognitionspsychologische Grundlagen der Urteilsbildung bei Delphi-Studien

Dieser Artikel wendet sich nun einem Hauptproblem der Diskussion um die Grundlagen der Delphi-Methode zu: der kognitionspsychologischen Begründung, auf welche Weise die Urteile bei einer Delphi-Befragung generiert werden. Im Mittelpunkt der folgenden Argumentation steht vor allem die Frage, warum angenommen werden kann, daß bei der Delphi-Methode, die mehrere Wiederholungen der Urteilsabgabe vorsieht, die Schätzungen in der Regel von Welle zu Welle qualitativ hochwertiger ausfallen.

Ausgangspunkt unserer Überlegungen sind neuere Forschungen zu Urteilsprozessen unter suboptimalen Bedingungen, insbesondere bei unvollständigem bzw. unsicherem Wissen. Zunächst wird gezeigt, daß es sich bei den Expertenschätzungen innerhalb einer Delphi-Studie aus kognitionspsychologischer Sicht um Urteilen unter Unsicherheit handelt. Im folgenden sollen dann neuere Ergebnisse aus der Kognitionspsychologie referiert werden, die zu diesem Problembereich vorliegen. Schließlich geht es um die Frage, wie diese Erkenntnisse für die wissenschaftliche Legitimierung von Delphi-Befragungen genutzt werden können.

3.1 Urteilen unter Unsicherheit - die allgemeine Anforderung an die Experten bei der Delphi-Methode

Zunächst sollen die kognitionstheoretischen Grundlagen, die die Experten befähigen, schon in der ersten Befragungsrunde ein kompetentes Urteil abzugeben, dargestellt werden. Gleichzeitig werden dabei Anforderungen an die Rekrutierung der Experten spezifiziert.

Die bei Delphi-Studien durch die Experten grundsätzlich zu lösende Aufgabe besteht im weitesten Sinne im Beurteilen von Sachverhalten, für die - selbst beim kompetentesten Experten - nicht alle Informationen vorliegen (können), um ein sicheres Urteil zu fällen. Dies stellt einen Unterschied zu Einstellungsmessungen dar, bei denen z.B. über den Grad der eigenen Zufriedenheit Auskunft gegeben werden soll.

„Problemlösen (und auch Urteilen, M.H./S.H.) stellt erhebliche Anforderungen an die menschliche Informationsverarbeitung. Dies gilt besonders dann, wenn komplexe Ziele vorliegen, viele Handlungsalternativen denkbar und Konsequenzen in zahlreichen Dimensionen abzuwägen sind. Entsprechend hängt der Erfolg entscheidend davon ab, wie gut Informationssuche, Informationsbewertung und Informationsverknüpfung gelingen“ (Zimolong/Rohrman 1988: 625). Um nun die Mechanismen zu verstehen, die bei Urteilsprozessen unter Unsicherheit ablaufen, wird die Kognitionspsychologie herangezogen, denn sie schreibt geistige Leistungen wie Verstehen, Denken, Wissen und schließlich Urteilen einem komplexen Informationsverarbeitungssystem zu (vgl. Wimmer/Perner 1979) und versucht, dieses transparent zu machen.

3.2 Aspekte der Bildung des Expertenurteils in der ersten Befragungswelle

In der Kognitionspsychologie wird der Mensch als ein System aufgefaßt, das aktiv Informationen aus der Umwelt aufnimmt, speichert, manipuliert und z.T. zielgerichtet weiterverwendet (Dutke 1994: 10). Dieses Informationsverarbeitungsparadigma ist zwar sehr

facettenreich (vgl. Scane 1987: 79ff.), beruht aber letztlich auf einigen wenigen Kernannahmen (Weidenmann 1988: 20ff.), von denen die folgenden drei im Zusammenhang mit der Bildung des Expertenurteils in einer Delphi-Studie interessieren:

- Informationen, die aktuell in der Umwelt nicht vollständig gegeben sind, können aus dem Wissensbestand des Individuums ergänzt werden.
- Die innere Repräsentation von Umweltgegebenheiten ist kein Abbild im passiv-photographischen Sinne, sondern eine aktive Konstruktion und Rekonstruktion. Diese Aufbauprozesse sind in der informationsverarbeitenden Aktivität und damit z.T. auch in der Intentionalität des Individuums begründet.
- Es erfolgt eine ständige zyklische Rückkopplung zwischen Wahrnehmung, Gedächtnis und Informationssuche: Der Gedächtnisbestand leitet durch Erwartungsbildung die Informationssuche und Wahrnehmung, die ihrerseits den Gedächtnisbestand verändert.

Diese Kernannahmen erlauben bereits die Formulierung von konkreten Erwartungen an die Kompetenz eines Experten:

- Wissensbestand und Kompetenz des Experten für eine gegebene Urteilsaufgabe bei Delphi verhalten sich direkt proportional, da bei der Lösung von Problemen bei Delphi stets (nur) unvollständige Informationen aus der Umwelt bereitstehen.
- Die Kompetenz eines Experten wird für die Lösung von Delphi-Aufgaben weiterhin durch seine Fähigkeiten zur internen Informationsspezifikation, -strukturierung und -bewertung sowie seine Motivation zur Problembewältigung bestimmt.
- Experten sollten möglichst stark mit dem Untersuchungsgegenstand vertraut sein und sich häufig mit dem angezielten Inhalt beschäftigen.

Wie funktioniert jedoch der Informationsverarbeitungsprozeß, insbesondere der des Urteilens unter Unsicherheit, konkret? Diese Frage zielt letztlich darauf, wie es den Experten bei Delphi-Studien gelingt, ein Urteil zu generieren, obwohl über den erfragten Sachverhalt kein vollständiges Wissen vorliegen kann.

Innerhalb der Kognitionspsychologie bietet vor allem die Theorienklasse mentaler Modelle, deren Forschungsgegenstand die menschliche Fähigkeit ist, „Strukturen und Prozesse realer Systeme in analoger Weise intern zu repräsentieren und sie damit verstehen, vorhersagen und erklären zu können“ (Conrad 1993: 129), einen Zugang zum näheren Verständnis solcher Urteilsprozesse. Zunächst soll dieser Ansatz kurz umrissen werden: Nach Dutke werden mentale Modelle dazu benutzt, „unabhängig von äußeren Vorgaben, Alltagswissen in Form gedanklicher Modelle (zu) organisieren, um sich das Verstehen oder Behalten bestimmter Sachverhalte zu erleichtern“ (Dutke 1994: 2). In der Kogni-

tionspsychologie dienen sie „der Erklärung menschlicher Informationsverarbeitungsleistungen. Mentale Modelle werden als kognitive Konstruktionen aufgefaßt, die auf einer Interaktion von Wahrnehmung und Gedächtnis beruhen. Sie sind von den Intentionen des Informationsverarbeiters abhängig und damit auch indirekt abhängig von der zu bewältigenden Aufgabe“ (ebenda: 12). Mentale Modelle beziehen schließlich „Alltagswissen in die Lösung von Problemen des logischen Schließens und des Urteilens mit ein“ (ebenda: 29).

Wichtig für die Beschreibung des Zustandekommens der Expertenurteile in einer Delphi-Befragung sind insbesondere die Aussagen, die die Theorie mentaler Modelle zum Urteilen unter Unsicherheit trifft. Dabei handelt es sich zunächst um eine alltägliche Anforderung an die menschliche Informationsverarbeitung, die deutliche Parallelen zu den bei Delphi-Studien auftretenden Anforderungen besitzt: „Es geht um die Frage, wie Individuen vorgehen, wenn sie unter suboptimalen Bedingungen (z.B. zu wenig Information, hohe Komplexität der Aufgabe, Zeitdruck) Einschätzungen, Ursachenerklärungen, Schlußfolgerungen, Vorhersagen usw. abzugeben haben“ (Strack 1985: 241).

Der Theorie mentaler Modelle nach Gigerenzer et al. folgend ist für den Erfolg einer Schätzung, d.h. eines Urteils bei unsicherem Wissen, entscheidend, ob für das Urteilen ein lokales mentales Modell (lokales MM) oder ein probabilistisches mentales Modell (PMM) benutzt wird. Während bei lokalen MM nur direkt auf die Aufgabe bezogenes Wissen aktiviert wird, wird bei der Konstruktion von PMM in größerem Umfang Erfahrungswissen aus dem Alltag (oder Expertenwissen) herangezogen, das sich u.a. in Häufigkeits- und Wahrscheinlichkeitsschätzungen manifestiert.

Für das Verständnis der bei Delphi ablaufenden Urteilsprozesse soll besonders auf die Ergebnisse von Studien verwiesen werden, die ermittelten, daß Versuchspersonen bei Experimenten dazu in der Lage waren, Häufigkeitsschätzungen relativ korrekt abzugeben, während ihnen jedoch die Beurteilung von Einzelfällen nicht gelang. Gigerenzer et al. (1991: 513ff.) schildern dazu das folgende Experiment.

Versuchspersonen bekamen die Aufgabe gestellt, zu beurteilen, welche von zwei Städten die größere Einwohnerzahl besitze. Dazu waren jeweils zwei Städte (z.B. Solingen und Heidelberg) zu vergleichen („Welche Stadt hat mehr Einwohner?“). Nach der Abgabe ihres Urteils wurden die Zielpersonen danach gefragt, wie sicher sie wären, daß ihre Antwort richtig sei. Diese Einzelfallbeurteilung gelang den Versuchspersonen nicht: Sie überschätzen im allgemeinen ihre eigene Urteilsfähigkeit. Dieses Phänomen wurde bereits auch in anderen Experimenten ermittelt und mit dem Terminus „overconfidence bias“ belegt (vgl. Nisbett/Ross 1980: 119f.). Wurden jedoch die gleichen Versuchsperso-

nen, nachdem sie 50 solcher Urteile abgegeben hatten, danach gefragt, wieviele dieser Urteile richtig gewesen seien, so gelang ihnen diese Häufigkeitsschätzung „erstaunlich akkurat“ (Dutke 1994: 35). Ein solches Phänomen wird von Gigerenzer et al. (1991: 511) als „Confidence-Frequency-Effekt“ bezeichnet. Dieser zunächst überraschende Effekt wird mit Hilfe der Theorie mentaler Modelle erklärbar:

Bei Entscheidungssituationen wird zunächst versucht, ein lokales mentales Modell zu konstruieren. Dieses ermöglicht den direkten Abruf des in der Situation erforderlichen Wissens. Lokale MM sind demzufolge auf die Lösung der jeweiligen Aufgabenstellung begrenzt und berücksichtigen kein weiteres Wissen. Außerdem benutzen sie lediglich elementare logische Operationen.

Kann ein entsprechendes lokales MM aus dem Gedächtnis abgerufen werden, wird das Urteil subjektiv als sicher eingestuft (Urteilen bei sicherem Wissen). Diese begrenzte Sichtweise auf ein Problem führt jedoch häufig zu dem bereits beschriebenen und in Experimenten angetroffenem „overconfidence bias“.

„If no local MM can be activated, it is assumed that a PMM is constructed next. A PMM solves the task by inductive inference, and it does so by putting the specific task into a larger context. A PMM connects the specific structure of the task with a probability structure of a corresponding natural environment (stored in long-term memory). In our example, a natural environment could be the class of all cities in Germany with a set of variables defined on this class, such as the number of inhabitants. This task selects the number of inhabitants as the target variable and the variables that covary with this target as the cues.

A PMM is different from a local MM in several respects. First, it contains a *reference class* of objects that includes the objects a and b. Second, it uses a network of variables in addition to the target variable for indirect inference. Thus, it is neither local nor direct...Probabilistic inference is part of the cognitive process, and uncertainty is part of the outcome“ (Gigerenzer et al. 1991: 507).

Damit geht die Theorie mentaler Modelle bei der Erklärung von Urteilsfindungen bei unsicherem Wissen über den Erkenntnisstand hinaus, den z.B. die Theorie der Urteilsheuristiken bietet. Hierbei werden replizierbare und stabile Fehlleistungen erklärt, indem argumentiert wird, menschliches Urteilen sei, besonders im sozialen Bereich, eher durch einen überhöhten Gebrauch von „primitiven intuitiven Strategien“ (auch: Faustregeln) denn durch die Anwendung „normativ angemessener Strategien“ gekennzeichnet (vgl. Nisbett/Ross 1980: 3). Diese würden die Urteilsfindung in natürlichen Situationen erleichtern, jedoch unter bestimmten Randbedingungen zu systematischen Verzerrungen

führen, was letztlich kognitive Täuschungen, wie z.B. den „overconfidence bias“ zur Folge hätte.

Ihre theoretischen Ausführungen und Experimente zusammenfassend schreiben Gigerenzer et al. dazu: „One cannot speak of a general overconfidence bias anymore, in the sense that it relates to deficient processes of cognition or motivation. In contrast, subjects seem to be able to make fine conceptual distinctions - confidence versus frequency - of the same kind as probabilists and statisticians do. Earlier attempts postulating general deficiencies in information processing or motivation cannot account for the experimental results predicted by PMM theory and confirmed in two experiments. PMM theory seems to be the first theory in this field that gives a coherent account of these various effects by focusing on the relation between the structure of the task, the structure of a corresponding environment, and a PMM“ (Gigerenzer et al. 1991: 526).

Mit den hier aufgegriffenen Gedanken zum Urteilen unter suboptimalen Bedingungen wurden zunächst einige u. E. wesentliche kognitionspsychologische Aspekte vorgestellt, die auch für die Urteilsbildung bei Delphi-Studien von Bedeutung sein dürften. Nun soll versucht werden, die bei den Experten in einer Delphi-Befragung ablaufenden kognitiven Prozesse mit Hilfe der Theorie der PMM konkreter zu erläutern.

Da kein sicheres Wissen für die Urteilsfindung zur Verfügung stehen kann, sind zunächst auch lokale mentale Modelle nicht anwendbar. Vielmehr müssen die Experten zur Schätzung umfangreiches Wissen (auch aus dem Langzeitgedächtnis) heranziehen, Referenzklassen bilden, komplexe Schlußfolgerungen ziehen usw. und letztlich aufgrund von Wahrscheinlichkeitshinweisen urteilen. Die Experten bilden aufgrund von - mitunter sicherlich unbewußten - Wahrnehmungen, aufgrund ihres Fachwissens und schließlich aufgrund ihrer jeweils fachspezifischen Intentionen hypothetische Modelle zur Lösung des erfragten Sachverhalts. Auf diese drei zentralen Elemente soll kurz konkreter eingegangen werden:

Die Experten beziehen ihre *Wahrnehmungen* hinsichtlich der von ihnen zu lösenden Aufgabe zunächst aus unterschiedlichen Quellen. Vor allem ihre eigene (soziale) Umgebung sowie Massenkommunikationsmittel spielen dabei wahrscheinlich eine entscheidende Rolle. Die jeweilige soziale Situation des Experten, aber auch die Selektivität der Wahrnehmungen sorgen für differenzierte Ergebnisse dieses Prozesses. Außerdem wird ein im Rahmen einer Delphi-Befragung von den Experten erwartetes Urteil stark von in der eigenen Umgebung wahrgenommenen Mehrheitsmeinungen (vgl. Noelle-Neumann 1989) geprägt werden.

Das *Fachwissen* eines Experten enthält sowohl empirische Erkenntnisse als auch spezifische Theorien, mit deren Hilfe Aspekte der Wirklichkeit erklärt werden. So wird beispielsweise die Sicht auf Jugendliche dadurch geprägt, ob es sich bei den Experten um Psychologen, Soziologen, Theologen, Mediziner, Juristen, Pädagogen oder Politologen handelt, die jeweils aufgrund ihres Fachwissens mit bestimmten Eigenschaften dieses Personenkreises vertraut sind. Damit werden auch die von ihnen bei der Bildung probabilistischer mentaler Modelle benutzten „cues“ jeweils fachgebietsspezifischen Charakter tragen.

In die Bildung von kognitiven Konstrukten zur Lösung einer Delphi-Aufgabe fließen weiterhin die *Intentionen* der Experten ein. Hier spielt offenbar der eigene Standpunkt, von dem aus ein Urteil gebildet wird, eine wichtige Rolle. Selbst ein ähnlich strukturiertes Fachwissen wird vor dem Hintergrund unterschiedlicher politischer Standpunkte und unterschiedlicher wirtschaftlicher Interessen zu divergierenden Urteilen führen.

Allen Expertenurteilen bei Delphi-Erhebungen ist zunächst gemeinsam, daß von ihnen lediglich unter Unsicherheit ein Urteil über den erfragten Sachverhalt abgegeben werden kann. Die Vielfalt an Wahrnehmungen, Fachwissen und Intentionen, über die die Experten verfügen, erlaubt es ihnen allerdings mentale Modelle zu erstellen, auf deren Grundlage sie zu qualitativ hochwertigen Urteilen gelangen.

3.3 Die Bildung der Expertenurteile in den folgenden Befragungswellen

Für die psychologische Begründung der in Experimenten nachgewiesenen Verbesserungen der Qualität der Schätzungen bei Delphi-Studien in den folgenden Wellen können zunächst grundsätzliche Erkenntnisse über Lernprozesse herangezogen werden. So stellen Zimolong/Rohrman fest: „Jedes gelernte Verhalten verändert sich mit der Anzahl seiner Ausführungen und der Qualität der erhaltenen Rückmeldungen. Der Zeitbedarf verringert sich exponentiell zur Anzahl der Wiederholungen, die Ausführungsqualität verbessert und stabilisiert sich gegenüber Störungen aus der Umwelt, die willentliche Steuerung und Kontrolle des Ablaufs wird durch eine unbewußte, automatische Regelung ersetzt. Als Ergebnis vermindert sich die erlebte Beanspruchung. Die als 'Potenzgesetz des Lernens' bekannte Beziehung gilt nicht nur für das sensumotorische Lernen, sondern darüber hinaus für jede Art von kognitiver Aktivität“ (Zimolong/Rohrman 1988: 628). Viele der hier beschriebenen Effekte, die sich als Lernerfolg interpretieren lassen, sind bei Delphi-Studien beobachtet worden (Häder/Häder 1994b; Häder/Häder/Ziegler 1995).

Besondere Bedeutung für die beobachtete Qualitätsverbesserung der Urteile kommt bei Delphi-Erhebungen den Rückmeldungen zu, die die Experten nach jeder Befragungs-

runde erhalten. Sie können ebenfalls direkt und indirekt einen Informationsgewinn bewirken, der bei der erneuten Urteilsbildung berücksichtigt wird.

So stellt Schwarz (1991; vgl. auch Hippler/Schwarz/Noelle-Neumann 1989; Schwarz/Sudman 1992; Schwarz/Strack/Hippler 1990) auf der Grundlage von Arbeiten zur flexiblen Konstruktion mentaler Repräsentationen (vgl. Barsalou 1987; Barsalou 1989) zur Urteilsbildung innerhalb einer Befragung fest, daß die Untersuchungspersonen hierbei auf verschiedene Arten von Wissen zurückgreifen. Bestimmtes Wissen ist den Befragten dabei permanent verfügbar, andere Informationen dagegen nur temporär. Wichtig für die Diskussion um die kognitionspsychologischen Grundlagen von Delphi-Studien ist, daß (zunächst) nicht alle potentiell relevanten Informationen aus dem Gedächtnis abgerufen werden, sondern daß der Suchprozeß abgebrochen wird, sobald die Befragten genügend Informationen erinnert haben, um sich mit hinreichender Sicherheit ein Urteil bilden zu können (vgl. Bodenhausen/Wyer 1987). „Das Urteil beruht daher nur auf der Teilmenge potentiell relevanter Information, die zum Urteilszeitpunkt leicht verfügbar ist - sei dies chronisch oder temporär.“ (Schwarz 1991: 72).

Die den Teilnehmern einer Delphi-Befragung dann nach jeder Welle zurückgemeldeten Informationen bewirken zusätzlich, daß diese den „Suchprozeß“ in ihrem Gedächtnis nochmals aufnehmen und nach weiteren, für den jeweiligen Sachverhalt relevanten Informationen fahnden. Kognitionspsychologisch soll die rückgemeldete Gruppenantwort damit einen gewissen Kontexteffekt bewirken und letztlich in indirekter Weise zur Verbesserung des abgegebenen Urteils beitragen.

Zugleich ist die Rückinformation aber auch als direkte, neue Information zu interpretieren. Sie muß in das bei den Experten bereits bestehende mentale Modell eingepaßt oder abgewiesen werden. Dazu muß eine Konsistenzprüfung darüber erfolgen, ob die neue Information in bezug auf das bestehende Modell als widerspruchsfrei integrierbar oder abzuweisen eingeschätzt wird. Folgende Ausgänge sind denkbar:

- Die Rückinformation ist widerspruchsfrei zum bestehenden mentalen Modell (die rückgemeldete Werte über das Gruppenmittel sind beispielsweise denen sehr ähnlich, die der Experte in der vorangegangenen Runde geschätzt hatte). In diesem Fall ist eine Änderung des mentalen Modells und damit des Urteils unwahrscheinlich. Dies gilt freilich nur unter der Voraussetzung, daß der Experte inzwischen kein eigenes weiteres Wissen aktiviert hat, welches sein ursprüngliches Urteil in Frage stellt.
- Die Rückinformation ist mit dem bisher abgegebenen Urteil unverträglich. In diesem Fall ist vor allem entscheidend, für wie kompetent das Expertengremium eingeschätzt wird, welche „Kosten“ dem einzelnen Experten bei einer Änderung seines Urteils

entstehen (vgl. Häder/Häder 1994) und wie sicher der Experte sich bei der Abgabe seines ersten Urteils war. Es muß allerdings auch berücksichtigt werden, daß bei dem betreffenden Experten möglicherweise Effekte des Selbstwertschutzes (vgl. Stahlberg/Osnabrügge/Frey 1985: 79ff.) eine Rolle spielen. Insofern ist es möglich, (a) daß die Information aus der Rückinformation zurückgewiesen und das eigene Urteil beibehalten wird, (b) das eigene Urteil korrigiert bzw. sogar verworfen und ein neues mentales Modell konstruiert wird, das widerspruchsfrei zur Rückinformation ist.

Für die Qualität des finalen Urteils ist auch der „tatsächliche“ Wahrheitsgehalt der einzelnen Rückmeldungen zu berücksichtigen. Würde man in Betracht ziehen, ob eine Veränderung in die richtige (oder falsche) Richtung erfolgt, dann erhöht sich die Zahl der oben aufgeführten möglichen Kombinationen. Insgesamt wird deutlich, daß sich sehr diffizile Prozesse bei der Integration der Rückmeldungen in ein bereits bestehendes mentales Modell abspielen. Generell lediglich eine Urteilsänderung in Richtung des Gruppenmittels aufgrund von Konformitätsdruck anzunehmen, wie es von Kritikern der Delphi-Methode gelegentlich vorgeworfen wird, ist damit nur *ein* möglicher Ausgang der Konsistenzprüfung und darüber hinaus bei hochkompetenten Experten kein sehr wahrscheinlicher.

Ein weiterer umfangreicher Problemkreis würde aufgespannt werden, wenn die Frage nach spezifischen Persönlichkeitsmerkmalen der Experten gestellt würde. Hier geht es um solche Verhaltensdispositionen, die die Experten dazu veranlassen, ihre Meinungen zu korrigieren bzw. beizubehalten.

4. Punkt- vs. Verteilungsschätzungen: ein Test

4.1 Begründung der Arbeitshypothese

In dem im folgenden vorzustellenden Test soll ein Aspekt der Urteilsfindung mit Hilfe probabilistischer mentaler Modelle aufgenommen und in einem Delphi-Experiment näher untersucht werden. Die Grundidee unseres Testaufbaus ist an das bereits im 3. Abschnitt erwähnte Experiment von Gigerenzer et al. angelehnt, in dem Untersuchungspersonen die gleiche Aufgabenstellung auf zwei unterschiedliche Arten zu bearbeiten hatten.

Für unseren Test rekrutierten wir zwei Gruppen von Untersuchungspersonen, für die hinsichtlich ihrer Fähigkeit zur Aufgabenbewältigung (inhaltliche Kompetenz, Kenntnisse empirischer Verteilungen, Motivation, soziodemographische Merkmale) Homogenität nachgewiesen werden konnte (vgl. Häder/Häder/Ziegler 1995)⁶. Beide Gruppen sollten die Ergebnisse zweier fünfstufig skalierten Indikatoren aus einer Befragung von Jugendlichen (eine spezifische Subpopulation der Shell-Jugendstudie) schätzen, die eine Gruppe

jedoch lediglich die in der Befragung ermittelten Mittelwerte, die andere Gruppe dagegen jede einzelne Antwortstufe für beide Indikatoren.

Im Rahmen der Theorie der PMM wird dargelegt, das erstens die Bildung von Referenzklassen (vgl. Brunswik 1943: 257), deren Bestandteile die sogenannten „probability cues“ sind, sowie zweitens die eigentliche Einschätzungsart, die „target variable“ (Gigerenzer et al. 1991: 507ff.), für die Vorhersage der Qualität eines Urteils unter Unsicherheit die entscheidende Rolle spielen.⁷⁾ Um die Vermutung zu unterstützen, daß zur Lösung der Aufgaben innerhalb dieses Tests PMM benutzt worden sind, müßten sich zunächst solche Referenzklassen benennen lassen. Derartige probability cues könnten z.B. sein:

- die Unterstellung, daß der zu schätzende Wert bekannten Stereotypen, wie z.B. über die Jugend entspricht (nach Luhmann sind Stereotype die wichtigsten Bausteine der öffentlichen Meinung, die eine Reduktion von Komplexität bewirken, vgl. dazu auch Noelle-Neumann 1989: 208),
- der Verdacht, daß die zu schätzenden Antworten denen auf ähnliche, bekanntere Fragen entsprechen und so von den Experten ein gewisser Analogieschluß vorgenommen werden kann,
- das Wissen über (z.B. in den Medien) spektakulär dargestellte Ereignisse, beispielsweise im Zusammenhang mit der Freude über den Fall der Mauer, die im Zusammenhang mit dem erfragten Sachverhalt stehen (die wiederum auf die Haltung der Menschen schließen lassen),
- die Kenntnis von sozialwissenschaftlichen Theorien, die eine Erklärung der zu lösenden Aufgabe enthalten,
- die Annahme, daß Antworten bei Befragungen häufig dazu tendieren, normalverteilt zu sein, d.h., daß der jeweilige zu schätzende Mittelwert in etwa bei der mittleren Antwortkategorie liegt,
- und schließlich (in der zweiten und dritten Befragungsrunde) die in den Rückmeldungen der einzelnen Wellen über Mittelwerte und Streuung enthaltenen Informationen über die Antworten der anderen Experten.

Bei Delphi-Befragungen vermögen die Experten nun die jeweiligen Wahrscheinlichkeiten, mit denen die einzelnen probability cues auf den erfragten Sachverhalt hindeuten, (besonders) fachmännisch einzuschätzen. Die zur Lösung der Aufgaben herangezogenen probability cues dürften in beiden Gruppen identisch sein, da formal jeweils nach den gleichen Sachverhalten gefragt wurde (Einstellung zur deutschen Wiedervereinigung sowie allgemeine Lebensziele von Jugendlichen). Wenn nun Unterschiede bei der Aufgabenbewältigung auftreten, so dürften diese auf die Art der gestellten Aufgabe und auf den damit jeweils verbundenen unterschiedlichen Schwierigkeitsgrad („target variable“) zu-

rückzuführen sein. Der Test wird zeigen, ob es möglich ist, für beide Aufgabentypen geeignete PMM zu bilden, oder ob nur eine Aufgabenart lösbar ist. Die daraus für die Gestaltung von Delphi-Befragungen resultierenden Implikationen liegen nahe.

4.2 Testaufbau und Ergebnisübersicht

Den beiden Gruppen von Untersuchungspersonen (jeweils N=16) wurden die beiden folgenden Indikatoren der Shell-Jugendstudie zur Beurteilung vorgelegt:

1. „Wie stehst Du zur Vereinigung von ehemaliger DDR und alter Bundesrepublik von heute aus gesehen?“

(1) sehr dafür; (2) eher dafür; (3) unentschieden; (4) eher dagegen; (5) sehr dagegen.

2. „Die Menschen sind ja sehr unterschiedlich, wenn es um ihre Lebensziele geht. Manche sind sehr anspruchsvoll und ehrgeizig, andere finden diese weniger gut oder wichtig. Wie ist das bei Dir?“

(1) nicht so anspruchsvoll und ehrgeizig (5) sehr anspruchsvoll und ehrgeizig.

Die Aufgabe bestand darin, im Rahmen eines Drei-Wellen-Delphi jeweils für Gruppe 1 als Punktschätzung und für Gruppe 2 als Verteilungsschätzung zu beurteilen, wie die 18-29jährigen westdeutschen Befragten in der Studie geantwortet haben.

Damit stehen die „wahren Werte“ aus der Shell-Studie zur Beurteilung der Güte der Schätzungen zur Verfügung. Diese soll insbesondere anhand der folgenden Kriterien vorgenommen werden (vgl. Häder/Häder/Ziegler 1995):

- Überdecken die Interdezilbereiche der Schätzungen jeweils die wahren Werte?
- Haben sich die Gruppenergebnisse den „wahren Werten“ angenähert?
- Haben sich die individuellen Schätzfehler pro Indikator von der ersten zur finalen Runde verringert?

In Tabelle 1 sind zunächst als Übersicht die Schätzergebnisse für die beiden Fragen der Gruppe 1 (Mittelwertschätzungen) dargestellt. Die Untersuchungspersonen waren offenbar in der Lage, die Mittelwerte der Verteilungen korrekt zu schätzen. Zwar weichen die finalen Punktschätzungen jeweils um circa 0,3 von den „wahren Werten“ der Shell-Studie ab. Doch liegen die Werte der Shell-Studie sogar für alle drei Wellen in den jeweiligen *Interquartilsbereichen* (restriktiver als Interdezilbereiche) aller drei Wellen. Damit kann die Qualität der Schätzungen aufgrund des ersten Kriteriums als sehr befriedigend eingeschätzt werden. Etwas differenzierter fällt die Beurteilung der Punktschätzungen aus, wenn die Veränderung der Distanzen der mittleren Schätzungen an die „wahren Werte“ von Welle zu Welle betrachtet werden. Nur im Fall des „Wiedervereinigungsindikators“

gelang eine Annäherung an den wahren Wert um 0,1 Skalenpunkte. Bei dem Indikator zur Beschreibung von Lebenszielen entfernte sich dagegen die finale Schätzung gegenüber der ersten Schätzung um 0,1 Skalenpunkte vom „wahren Wert“. Da insgesamt aber die ersten Schätzungen schon relativ genau waren, sollte diese geringfügige Abweichung nicht überbewertet werden.

Tabelle 1: Mittelwerte, Interdezilbereiche und Interquartilsbereiche der Schätzungen von Gruppe 1

<i>Frage 1 (Wiedervereinigung)</i>			
	Punktschätzung	Interdezilbereich	Interquartilsbereich
1. Welle	2.9	2.0 - 3.8	2.0 - 3.5
2. Welle	2.8	2.1 - 3.4	2.4 - 3.1
3. Welle	2.8	2.2 - 3.5	2.3 - 3.3
Shell	2.5		
<i>Frage 2 (Lebensziele)</i>			
	Punktschätzung	Interdezilbereich	Interquartilsbereich
1. Welle	3.5	3.0 - 3.4	3.1 - 3.9
2. Welle	3.5	3.0 - 3.9	3.0 - 3.8
3. Welle	3.4	3.0 - 3.9	3.1 - 3.7
Shell	3.7		

Dagegen ist deutlich erkennbar (vgl. Tabelle 2), daß die Untersuchungspersonen der zweiten Gruppe nicht in der Lage waren, die Verteilungen korrekt zu schätzen. So werden nur fünf von zehn Mittelwerten der Shell-Studie von den Interdezilbereichen der Schätzungen der dritten Welle überdeckt, die geschätzten und die wahren Verteilungsformen stimmen nicht annähernd überein.

Ganz offenbar waren die Untersuchungspersonen mit dieser relativ schwierigen Aufgabe überfordert. Dies mag u.a. auf die mangelnde Vertrautheit der Studenten mit statistischen Verteilungen zurückzuführen sein. So wurden z.B. die Schiefen der Verteilungen nicht richtig beurteilt: Die beiden finalen Verteilungen der Schätzungen sind sich relativ ähnlich, sie erinnern an Normalverteilungen, während von den Originalverteilungen die eine rechts- und die andere linksschief ist.

Abschließend werden nun die Schätzergebnisse der einzelnen Experten für die beiden Fragen betrachtet. Dabei soll geprüft werden, wie sich die individuellen Schätzfehler der beiden Gruppen jeweils für beide Fragen von Welle zu Welle entwickelt haben (3. Kriterium).

Tabelle 2: Mittelwerte (MW) und Interdezilbereiche (IDB) der Schätzungen von Gruppe 2

	1. Welle		2. Welle		3. Welle		Shell
	MW	IDB	MW	IDB	MW	IDB	
Frage 1							
sehr dafür	17	5-40	11	5-20	11	5-20	21
eher dafür	26	20-40	24	15-40	28	20-30	32
unentschieden	30	10-40	32	20-50	32	20-40	24
eher dagegen	16	5-30	22	15-30	20	15-25	18
sehr dagegen	11	5-20	11	5-20	9	5-15	5
Frage 2							
n.so ehrgeizig	13	5-25	11	5-15	11	10-15	1
	17	10-20	21	15-30	21	10-30	8
	25	10-50	34	30-45	32	25-40	32
	25	8-50	22	15-30	23	15-30	42
sehr ehrgeizig	20	5-40	12	10-15	13	10-15	17

Die Verringerungen der Fehler bei den Mittelwertschätzungen von Welle zu Welle bei fast allen Untersuchungspersonen sprechen für die hohe Qualität der abgegebenen Schätzungen: Bei Frage 1 haben sich nur zwei Experten von der ersten zur dritten Welle verschlechtert, bei Frage 2 hat sich der χ^2 -Wert von lediglich drei der sechzehn Versuchsteilnehmer in der dritten gegenüber der ersten Welle erhöht (vgl. Tabelle 3)⁸⁾. Bei beiden Punktschätzungen kann eine Verringerung des mittleren Schätzfehlers von der ersten zur letzten Runde konstatiert werden.

Tabelle 4 zeigt, daß sich die mittleren Schätzergebnisse der Gruppe bei der Verteilungsschätzung sowohl bei Frage 1 als auch bei Frage 2 über die Wellen trotz der insgesamt schlechten Schätzung verbessert haben. Hinsichtlich der individuellen Ergebnisse kann insgesamt auch eine positive Einschätzung getroffen werden: Vier Verschlechterungen von Welle 1 zur Welle 3 bei Frage 1 stehen zehn Verbesserungen gegenüber. Bei Frage 2 ist dieses Verhältnis etwas ungünstiger, dort beträgt es sechs zu neun.

Insgesamt sind damit die Punktschätzungen erfolgreicher verlaufen als die Verteilungsschätzungen. Dies bestätigt die Vermutung, daß bei gleicher inhaltlicher Aufgabenstellung (trotzdem) nur eine der beiden Aufgabenarten lösbar ist. Offenbar unterscheidet sich der Schwierigkeitsgrad beider Aufgabentypen wesentlich.

Tabelle 3: Individuelle χ^2 -Werte für Frage 1 und 2, Gruppe 1 (Mittelwertschätzung)*)

Frage 1, Gruppe 1					Frage 2, Gruppe 1				
ID	Welle 1	Welle 2	Welle 3	Diff.	ID	Welle 1	Welle 2	Welle 3	Diff.
1	0.1	0.1	0.3	-0.2	1	0.1	0.1	0.1	0.0
2	0.7	0.1	0.1	0.6	2	0.0	0.0	0.0	0.0
3	0.7	0.3	0.3	0.4	3	0.0	0.0	0.0	0.0
4	0.4	0.1	0.4	0.0	4	0.4	0.1	0.1	0.3
5	0.3	0.2	0.0	0.3	5	0.0	0.0	0.0	0.0
6	0.1	0.0	0.0	0.1	6	0.1	0.1	0.1	0.0
7	0.2	0.7	0.2	0.0	7	0.0	0.0	0.0	0.0
8	0.1	0.1	0.1	0.0	8	0.0	0.1	0.1	-0.1
9	0.3	0.2	0.1	0.2	9	0.0	0.0	0.0	0.0
10	0.1	0.3	0.1	0.0	10	0.0	0.0	0.0	0.0
11	0.1	0.7	0.1	0.0	11	0.0	0.1	0.1	-0.1
12	0.7	0.0	0.0	0.7	12	0.0	0.0	0.3	-0.3
13	0.1	0.1	0.1	0.0	13	0.1	0.0	0.0	0.1
14	0.3	0.0	0.5	-0.2	14	0.1	0.0	0.0	0.1
15	0.1	0.1	0.1	0.0	15	0.1	0.1	0.1	0.0
16	0.1	0.0	0.0	0.1	16	0.1	0.0	0.0	0.1
Sum.	4.4	2.4	2.4		Sum.	1.0	0.6	0.9	

Tabelle 4: Individuelle χ^2 -Werte für Frage 1 und 2, Gruppe 2 (5 Antwortkategorien)*)

Frage 1, Gruppe 2					Frage 2, Gruppe 2				
ID	Welle 1	Welle 2	Welle 3	Diff.	ID	Welle 1	Welle 2	Wel.3	Diff.
1	46.2	74.0	---	---	1	105.4	105.4	---	---
2	35.4	45.1	27.8	7.6	2	64.1	117.5	115.4	-51.3
3	11.3	11.7	34.5	-23.2	3	33.8	106.2	196.9	-163.1
4	41.4	16.5	12.6	28.8	4	395.6	125.0	115.4	280.2
5	26.2	27.9	24.8	1.4	5	92.6	106.2	90.9	1.7
6	39.2	14.2	3.6	35.6	6	43.5	111.2	242.9	-199.4
7	33.1	23.3	17.6	15.5	7	227.4	161.8	89.8	137.6
8	8.8	22.9	12.6	-3.8	8	64.1	214.2	228.7	-164.6
9	45.1	34.7	---	---	9	85.6	40.6	118.6	-33.0
10	11.3	57.0	34.5	-23.2	10	612.7	246.7	127.0	485.7
11	57.7	51.4	3.6	54.1	11	419.3	159.2	276.6	152.7
12	30.9	18.2	21.2	9.7	12	1707.3	149.7	99.9	1607.4
13	57.0	29.0	6.1	50.9	13	395.6	109.0	106.2	289.4
14	12.9	12.6	16.4	-3.5	14	9.7	105.4	105.4	-95.7
15	143.2	48.5	22.4	120.8	15	400.6	381.8	246.7	153.9
16	74.2	26.5	23.2	51.0	16	38.4	64.2	90.9	-52.5
Sum.	673.9	513.6	260.7		Sum.	4695.5	2304.1	2251.	

*) In Tabelle 3 und 4 wurden die Werte der Shell-Studie als erwartete Werte angenommen.

Bemerkenswert ist weiterhin, daß selbst bei den beiden Aufgaben zur Verteilungsschätzung eine Verringerung der mittleren Schätzfehler der Gruppe festgestellt werden konnte. Dieses Ergebnis deutet insgesamt darauf hin, daß die eingangs erwähnten Denkprozesse, bei denen im Laufe der Aufgabenbearbeitung immer mehr Kontextwissen für die Problemlösung einbezogen wird, was sich in der Konstruktion immer hochwertigerer mentaler Modelle niederschlägt, tatsächlich zu einer höheren Qualität der finalen Schätzung führen als dies bei einer nur einmaligen Urteilsabgabe zu erwarten ist.

4.3 Zusammenfassung

Mit diesem Test liegt ein weiterer empirischer Hinweis dafür vor, daß Delphi zu Denkvorgängen während der Aufgabenbearbeitung (hervorgerufen durch wiederholte mentale Auseinandersetzung mit der inhaltlichen Aufgabenstellung sowie den Rückmeldungen) führt, die positiven Einfluß auf die Qualität der Ergebnisse haben.

Wie gut allerdings die Schätzungen der letzten Welle absolut sind, hängt wesentlich von der generellen Eignung der Experten ab, d.h. von ihrer Sachkompetenz schon zu Beginn der Delphi-Studie⁹⁾. Insofern muß für diesen Test eine Einschränkung gemacht werden: Die Kompetenz der Untersuchungspersonen gründete sich lediglich darauf, selbst Angehörige der Subpopulation zu sein, deren Befragungsergebnisse aus der Shell-Studie zu beurteilen waren. Dagegen gaben die Studenten in der Zusatzbefragung zur Ermittlung der von ihnen benutzten kognitiven Orientierungshilfen an, daß Erfahrungen mit ähnlichen Fragestellungen sowie mit statistischen Verteilungen nur eine geringe Rolle bei den abgegebenen Schätzungen gespielt haben¹⁰⁾. Als „probability cues“ nutzten sie dagegen vor allem (übereinstimmend in beiden Gruppen) erstens die Erfahrungen mit anderen Jugendlichen, zweitens haben sie sich „spontan durch das Gefühl“ leiten lassen und drittens Stereotype über Jugendliche (vgl. Häder/Häder/Ziegler 1995) verwendet.

Das Testergebnis erlaubt weiterhin Rückschlüsse auf das Vorhandensein einer Fähigkeit, Tendenzen der öffentlichen Meinung wahrzunehmen. Einen empirisch begründeten Hinweis auf die menschliche Fähigkeit, nicht nur Meinungen anderer wahrzunehmen, sondern diese sogar „statistisch“ zu schätzen, findet sich bei Noelle-Neumann (1989).¹¹⁾ Sie versteht den Menschen zunächst als ein soziales Wesen und schlußfolgert daraus, daß er permanent mit einer „Isolationsfurcht“ konfrontiert werde. Diese hat die Autorin mittels empirischer Experimente nachgewiesen (vgl. Noelle-Neumann 1989: 59ff.; neuere Experimente werden auf S. 303ff. beschrieben). Zur Vermeidung dieser Isolation beobachten sich die sozialen Wesen untereinander und lernen so, die Meinungen der anderen (insbesondere Mehrheitsmeinungen) zu erkennen. Diese Wahrnehmung bewirke soziale Inte-

gration und sei so eine Voraussetzung für das Funktionieren einer Gesellschaft. „Die menschlichen Gesellschaften versichern sich auf diese Weise ihres Zusammenhalts und einer ausreichenden Bereitschaft des einzelnen zum Kompromiß“ (ebenda: 167).

Noelle-Neumann geht in ihrer Argumentation noch weiter, sie spricht nicht nur von der Entdeckung einer neuen menschlichen Fähigkeit, der „Wahrnehmung von Meinungsklima“ (ebenda: 27), sondern kennzeichnet diese Begabung sogar als „quasistatistisch“ (ebenda: 31).

Das Ergebnis unseres Tests erlaubt eine Präzisierung dieser Aussagen. So hat sich nur bei einem Aufgabentyp - bei der Schätzung von Mittelwerten - gezeigt, daß die befragten Studenten dazu in der Lage sind, Mehrheitsmeinungen richtig wiederzugeben. Differenziertere Auskünfte über relativ feine Abstufungen dieses Urteils gelingen ihnen jedoch nicht. Die kompliziertere Schätzung von einzelnen Skalenpunkten übersteigt zumindest die Kompetenz der an diesem Test beteiligten Studenten. So wird tatsächlich nur ein *Meinungsklima* wahrgenommen und nicht etwa detaillierte Meinungsdifferenzierungen. Für die Gestaltung von Delphi-Befragungen gilt es - die Ergebnisse dieses Tests verallgemeinernd - zu vermerken, daß differenziertere, über eine Wiedergabe des Meinungsklimas (auf dem Niveau von Mittelwerten) hinausgehende Urteile zumindest von Experten ohne statistische Grundkenntnisse nicht erwartet werden können.

In unseren Ausgangsüberlegungen hatten wir die bei Delphi ablaufenden kognitiven Prozesse als Urteilen unter Unsicherheit charakterisiert. Die Schätzungen, die im hier vorgestellten Test von den Untersuchungspersonen abzugeben waren, sind typische Beispiele für derartige Denkprozesse, die tatsächlich auch mit (mehr oder weniger) Erfolg bewältigt worden sind. Es konnten damit Parallelen hergestellt werden zwischen der von Gigerenzer et al. beschriebenen Bildung mentaler Modelle und den kognitiven Prozessen, die bei der Bearbeitung einer Delphi-Studie zur Schätzung von Ergebnissen einer Bevölkerungsbefragung zu leisten sind. Es hat sich jedoch auch gezeigt, daß weitere intervenierende Bedingungen zu beachten sind. So tritt das erwartete Ergebnis nicht unter allen Umständen ein. Die von den Testteilnehmern im Rahmen ihrer mentalen Modelle aktivierten probability cues reichen lediglich, um über Mittelwerte zuverlässig Auskunft zu geben. Die für die Schätzung von Verteilungen erforderlichen cues, wie z.B. statistische Kenntnisse, das Wissen über die bei Bevölkerungsbefragungen zu erwartenden empirischen Untersuchungsergebnisse usw. fehlen jedoch. Schließlich besteht ein wesentliches Ergebnis dieser Überlegungen darin, daß mit der Theorie mentaler Modelle eine taugliche kognitionspsychologische Grundlage für die Aufdeckung, Erklärung und Vorhersage von bei Delphi-Befragungen ablaufenden Prozessen gefunden worden ist.

5. Ausblick

Mit dem im vorliegenden Aufsatz dargestellten Theorieansatz zur Erklärung der bei Delphi-Studien ablaufenden kognitiven Prozesse und der empirischen Bearbeitung einiger Aspekte dieser theoretischen Überlegungen wurde das Ziel verfolgt, einen Beitrag zu leisten, um diese Methode für die Anwendung auch in den deutschen Sozialwissenschaften aufzubereiten und letztlich zu legitimieren.

Das derzeit gebräuchliche Standardrepertoire an Methoden der empirischen Sozialforschung läßt durchaus Bedarf an weiteren Techniken für die Lösung spezieller Probleme offen. Gerade für Themenstellungen, die sich einer Bearbeitung durch direkte Bevölkerungsbefragungen bzw. Befragungen spezieller Subpopulationen entziehen, da sie

- besonders heikle Fragestellungen beinhalten und/oder
 - bei den Befragten nicht aktuell präsent sind (z.B. da sie in der Vergangenheit liegen)¹²⁾ und/oder
 - zu kompliziert bzw. zu komplex für die Bearbeitung in Form einer Bevölkerungsbefragung sind und/oder
 - sich auf Zielpopulationen richten, die schwer bzw. nicht erreichbar sind¹³⁾,
 - für die aber eine Abbildung der Meinungen bzw. Einstellungen aus forschungstheoretischen Erwägungen dennoch unverzichtbar ist,
- könnte der Delphi-Ansatz erfolgreich eingesetzt werden.

Dabei - und dies sei unterstrichen - will die im Test beschriebene Form der Delphi-Nutzung keinesfalls Befragungen ersetzen. Sie stellt aber bei Problemen wie den genannten eine potentiell mögliche Lösung dar und macht sie damit der empirischen Analyse zugänglich.

Anmerkungen

*) Wir möchten uns herzlich bei Herrn Rolf Steyer und Herrn Ullrich Hoffrage für die Hinweise zu diesem Manuskript bedanken.

1) Die Angaben in der Literatur darüber, ob es sich um ein Hunde- oder Pferderennen handelte, widersprechen sich (vgl. Woundenberg 1991: 132; Seeger 1979: 57).

2) Eine zusammenfassende und kritische Diskussion solcher Validierungsansätze hat Woundenberg (1991) vorgelegt.

3) Almanachfragen „das sind Fragen, deren Gegenstand sich auf gegenwärtige oder vergangene soziale Realitäten beziehen und deren Antworten bekannt sind“ (Seeger 1979: 61).

4) Seeger stellt dazu fest: " ... sicher auch die Beliebtheit der Anwendung haben (neben anderen Faktoren, M.H./S.H.) es verursacht, daß in der Fachöffentlichkeit recht große Verwirrung über Zweck, Ziel und Wirkung der Methode bereitgemacht hat und bislang einem umfangreichen Verständnis der Methode im Weg gestanden hat (1979: 44)."

5) Zur Begründung dieser Spezifizierungen vgl. Abschnitt 3.

6) Es handelte sich um 32, allerdings mit empirischer Sozialforschung kaum vertraute Medizinstudenten der Universität Marburg. Der Test wurde im November und Dezember 1994 unter Klausurbedingungen in drei Wellen sowie einer Zusatzbefragung zu benutzten Orientierungshilfen und über die Motivation der Teilnehmer veranstaltet.

7) Eine Referenzklasse, wie sie in dem von Gigerenzer et al. (1991) geschilderten Experiment (vgl. Abschnitt 3.) gebildet werden könnte, würde alle deutschen Städte umfassen, zu denen auch Heidelberg und Solingen gehören. Als Probability cues können nun solche Annahmen dienen, die - resultierend aus der Aufgabenstellung, die größere der beiden Städte herauszufinden - Hinweise auf die jeweilige Einwohnerzahl erlauben. So kann angenommen werden, daß: von größeren Städten eher eine Mannschaft in der Bundesliga spielt, größere Städte kürzere Kraftfahrzeugkennziffern besitzen, wenn es sich bei einer der beiden Städte um die Hauptstadt eines Bundeslandes handelt, diese dann die größere ist usw. In der Theorie der PMM wird weiterhin darauf verwiesen, daß Probability cues die Eigenschaft haben, (lediglich) mit einer bestimmten Wahrscheinlichkeit auf die target variable, den interessierenden Sachverhalt, hinzudeuten. Gigerenzer et al. (1991: 509) stellten beispielsweise fest, daß die Benutzung des Fußballmannschafts-probability cues in der Bundesligaspielsaison 1988/89 mit einer relativen Häufigkeit von 91 Prozent zu einer richtigen Antwort bei der Frage nach der größeren Einwohnerzahl einer westdeutschen Stadt führen würde.

8) Der χ^2 -Wert wird hier als Maß benutzt, das die Distanz zwischen tatsächlicher und geschätzter Antwortverteilung zusammengefaßt darstellt.

9) In einem Test mit Experten von ZUMA konnte z.B. gezeigt werden, daß die Schätzungen vierstufiger Antwortverteilungen durchaus in hoher Qualität absolviert werden können.

10) Da in dieser Beziehung also nicht von einer optimalen Zusammensetzung des Expertengremiums gesprochen werden kann, haben wir es in diesem Testbericht vorgezogen, die Studenten als „Testteilnehmer“ bzw. „Untersuchungspersonen“ zu bezeichnen. Gleichwohl war die Testanordnung geeignet, einige Aspekte prinzipiell bei Delphi ablaufender Denkprozesse zu untersuchen.

11) Die Fähigkeit von Experten, Auskunft über die Meinungen anderer Menschen geben zu können, hat auch Reuband (1990) in einem anderen Zusammenhang empirisch nachgewiesen. Er stellt methodische Forschungen zu Interviewerfälschungen vor, die ebenfalls Relevanz für die Schätzung der Ergebnisse von Bevölkerungsbefragungen mit Hilfe der Delphi-Methode besitzen dürften und kommt anhand empirischer Tests zu dem Ergebnis, daß Studenten dazu in der Lage sind, die Ergebnisse von ALLBUS-Befragungen (allerdings lediglich in einer Befragungswelle) zu schätzen. Reuband stellt fest: „Vergleicht man die Randverteilungen der gefälschten Interviews mit der Randverteilung der jeweiligen Referenzumfrage, so ergeben sich unerwartet große Ähnlichkeiten“ (Reuband 1979: 714).

12) Es würde sogar eine wesentliche Bereicherung des sozialwissenschaftlichen Methodenarsenals darstellen, wenn es gelänge - eine ausreichende Validität vorausgesetzt - auch retrospektiv die Ergebnisse von Bevölkerungsbefragungen zu schätzen. Hier mag ein Verweis auf die sich im Rahmen der gegenwärtigen Forschungen zur sozialen Transformation in Ostdeutschland erschließenden Möglichkeiten - z.B. Vergleichsdaten zu erheben bzw. zu schätzen - genügen.

13) Dies könnte beispielsweise im Rahmen sozialwissenschaftlicher Forschungen zur AIDS-Prävention gelten. Hier wurde - zwar nicht mit Hilfe der Delphi-Methode - „der Zugang über Prostituierte gewählt und weibliche Prostituierte als Expertinnen über ihre Freier befragt“ (Markert 1994: 369). Eine sinnvolle Nutzung des Delphi-Ansatzes könnte hier sicherlich zu einer weiteren Qualifizierung der Ergebnisse beitragen.

Literatur

Albach, H., 1970: Informationsgewinnung durch strukturierte Gruppenbefragung. Die Delphi-Methode. Zeitschrift für Betriebswirtschaft 40/1970 (Ergänzungsheft): 11-26.

Barsalou, L. W., 1987: The instability of graded structure: Implications for the nature of concepts, S. 101-140 in: U. Neisser (Hrsg.), Concepts and conceptual development: Ecological and intellectual factors in categorization, Cambridge: Cambridge University Press.

Barsalou, L. W., 1989: Intraconcept similarity and its implications for interconcept similarity, S. 76-121 in: S. Vosniadou/A. Ortony (Hrsg.), Similarity and analogical reasoning, Cambridge: Cambridge University Press.

Becker, D., 1974: Analyse der Delphi-Methode und Ansätze zu ihrer optimalen Gestaltung. Dissertation, Universität Mannheim.

- BMFT, 1993: Deutscher Delphi-Bericht zur Entwicklung von Wissenschaft und Technik, im Auftrag des Bundesministeriums für Forschung und Technologie (BMFT), Bonn.
- Bodenhausen, G. V./Wyer, R. S., 1987: Social cognition and social reality: Information acquisition and use in the laboratory and the real world. S.6-41 in: H.-J. Hippler/N. Schwarz/S. Sudman (Hrsg.), Social information processing and survey methodology, New York: Springer.
- Brockhoff, K., 1979: Delphi-Prognosen im Computer-Dialog. Experimentelle Erprobung und Auswertung kurzfristiger Prognosen. Tübingen: Paul Siebeck.
- Brunswik, E., 1943: Organismic achievement and environmental probability. Psychological Review, 50.
- Coates, J. F., 1975: In Defense of Delphi: A Review of Delphi Assessment, Expert Opinion, Forecasting and Group Process by H. Sackman, Technological Forecasting and Social Change 7: 193-194.
- Conrad, E.-M., 1993: Gedächtnis und Wissensrepräsentation: Aspekte der Abbildungsleistung kognitionspsychologischer und filmsemiotischer Modelle; ein Impuls zum Paradigmenwechsel. Hildesheim: Olms, Serien: Philosophische Texte und Studien; 37.
- Dalkey, N. C., 1969: The Delphi Method: An Experimental Study of Group Opinion. RAND RM 5888-PR, June.
- Dutke, St., 1994: Mentale Modelle: Konstrukte des Wissens und Verstehens. Göttingen/Stuttgart: Verlag für angewandte Psychologie.
- Frey, D./Irle, M., 1985: Theorien der Sozialpsychologie. Band 3. Bern u.a.: Hans Huber.
- Geschka, H., 1977: Delphi. In: Bruckmann, G. (Hrsg.), Langfristige Prognosen. Möglichkeiten und Methoden der Langfristprognostik komplexer Systeme. Würzburg/Wien.
- Gigerenzer, G./Murray, D. J., 1987: Cognition as Intuitive Statistics. Hillsdale/London: Lawrence Erlbaum Associates.
- Gigerenzer, G./Hoffrage, U./Kleinböling, H., 1991: Probabilistic Mental Models: A Brunswikian Theory of Confidence. In : Psychological Review. Vol. 98. No. 4:506-528.
- Gordon, T. J./Helmer, O., 1964: Report on a Long Range Forecasting Study. Rand Paper P-2982. Santa Monica, Cal: Rand Corporation.
- Häder, M./Häder, S., 1994a: Die Grundlagen der Delphi-Methode. Ein Literaturbericht. Mannheim, ZUMA-Arbeitsbericht Nr. 94/02.
- Häder, M./Häder, S., 1994b: Ergebnisse einer experimentellen Studie zur Delphi-Methode. Mannheim, ZUMA-Arbeitsbericht Nr. 94/05.

- Häder, M./Häder, S./Ziegler, A., 1995: Punkt- vs. Verteilungsschätzungen: Ergebnisse eines Tests zur Validierung der Delphi-Methode. Mannheim, ZUMA-Arbeitsbericht Nr. 95/05.
- Hasher, L./Goldstein, D./Toppino, T., 1977: Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16.
- Helmer, O./Rescher, N., 1960: On the Epistemology of the Inexact Sciences. Project Rand Report R-353.
- Hippler, H.-J./Schwarz, N./Noelle-Neumann, E., 1989: Response Order Effects in Dichotomous Questions: The Impact of Administration Mode. ZUMA-Arbeitsbericht 89/17. Mannheim.
- Howell, W. C./Burnett, S., 1978: Uncertainty measurement: A cognitive taxonomy, in: *Organizational Behavior and Human Performance*, 22.
- Jugend '92, Maschinenlesbares Codebuch, ZA-Nr. 2323, Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln.
- Kaufmann, H.-J., 1972: Methoden der technologischen Vorschau im Dienste der Forschungsplanung industrieller Unternehmen unter besonderer Berücksichtigung der Delphi-Methode. Dissertation, Universität Mannheim.
- Köhler, G., 1992: Methodik und Problematik einer mehrstufigen Expertenbefragung, In: J. H. P. Hoffmeyer-Zlotnik (Hrsg.), *Analyse verbaler Daten. Über den Umgang mit qualitativen Daten*. Opladen: Westdeutscher Verlag.
- Linestone, H. A./Turoff, M. (Hrsg.), 1975: *The Delphi Method*. London u.a.: Addison-Wesley.
- Markert, St., 1994: Risikoverhalten von Freiern. S. 369-375 in: W. Heckmann/M. A. Koch (Hrsg.), *Sexualverhalten in Zeiten von AIDS*. Berlin: edition sigma.
- Nisbett, R. E./Ross, L., 1980: *Human inferences: Strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice Hall.
- Noelle-Neumann, E., 1989: *Öffentliche Meinung. Die Entdeckung der Schweigespirale*, Frankfurt, M. Berlin: Ullstein.
- Reuband, K. H., 1990: Interviews, die keine sind. "Erfolge" und "Mißerfolge" beim Fälschen von Interviews. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 42: 706-733.
- Rowe, G./Wright, G./Bolger, F., 1991: Delphi, A Reevaluation of Research and Theory, *Technological Forecasting and Social Change* 39: 235-251.

- Sackman, H., 1975: Delphi Critique. Expert Opinion, Forecasting, and Group Process. Lexington/Toronto/London: Lexington Books.
- Saliger, E./Kunz, C., 1981: Zum Nachweis der Effizienz der Delphi-Methode. Zeitschrift für Betriebswirtschaft 51: 470-480.
- Scane, R., 1987: A historical perspective. In: Gardiner, M. M./Christie, B. (Hrsg.): Applying cognitive psychology to user interface design. Chichester: Wiley.
- Schwarz, N., 1991: Assimilation und Kontrast in der Urteilsbildung: Implikationen für Fragereihenfolgeeffekte. ZUMA-Nachrichten 29: 70 - 86.
- Schwarz, N./Sudman, S., 1992: Context Effects in social and psychological research. New York: Springer.
- Schwarz, N./Strack, N./Hippler, H.-J., 1990: Kognitionspsychologie und Umfrageforschung: Themen und Befunde eines interdisziplinären Forschungsgebietes. ZUMA-Arbeitsbericht 90/07. Mannheim.
- Seeger, Th., 1979: Die Delphi-Methode. Expertenbefragungen zwischen Prognose und Gruppenmeinungsbildungsprozessen; überprüft am Beispiel von Delphi-Befragungen im Gegenstandsbereich Information und Dokumentation. Diss., Freiburg: HochschulVerlag.
- Stahlberg, D./Osnabrügge, G./Frey, D., 1985: Die Theorie des Selbstwertschutzes. In: Frey, D./Irlle, M. (Hrsg.): Theorien der Sozialpsychologie. Band 3. Bern u.a.: Hans Huber.
- Strack, F., 1985: Urteilsheuristiken. In: Frey, D./Irlle, M. (Hrsg.): Theorien der Sozialpsychologie. Band 3. Bern u.a.: Hans Huber.
- Wechsler, W., 1978: Delphi-Methode - Gestaltung und Potential für betriebliche Prognoseprozesse, München.
- Weidenmann, B., 1988: Psychologische Prozesse beim Verstehen von Bildern. Bern u.a.: Hans Huber.
- Wimmer, H./Perner, J., 1979: Kognitionspsychologie. Eine Einführung. Stuttgart: Kohlhammer.
- Woundenberg, F., 1991: An Evaluation of Delphi, Technological Forecasting and Social Change 40: 131-150.
- Zacks, R. T./Hasher, L./Sanft, H., 1982: Automatic encoding of event frequency: Further findings. Journal of Experimental Psychology, 35.
- Zimolong, B./Rohrman, B., 1988: Entscheidungshilfetechnologien. In: Frey, D./Graf Hoyos, C./Stahlberg, D. (Hrsg.), Angewandte Psychologie. Weinheim: Psychologie Verlags Union.