

### Neuere Entwicklungen bei der Delphi-Methode: Literaturbericht II

Häder, Michael; Häder, Sabine

Veröffentlichungsversion / Published Version

Biographie / biography

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Häder, M., & Häder, S. (1998). *Neuere Entwicklungen bei der Delphi-Methode: Literaturbericht II*. (ZUMA-Arbeitsbericht, 1998/05). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-200515>

#### 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.

ZUMA-Arbeitsbericht 98/05

**Neuere Entwicklungen bei der Delphi-Methode  
Literaturbericht II**

*Michael Häder, Sabine Häder*

September 1998

ZUMA  
Quadrat B2,1  
Postfach 12 21 55  
D-68072 Mannheim

Telefon: (0621) 1246 –136 oder -282  
Telefax: (0621) 1246 - 100  
E-mail: haeder@zuma-mannheim.de  
sabine.haeder@zuma-mannheim.de

## **Gliederung**

0. Einleitung	3
1. Definitionen von Delphi als Indikatoren für die sich ausdifferenzierenden Erwartungen an diese Methode	4
2. Anwendungen von Delphi	7
2.1 Einsatzgebiete	7
2.2 Einteilungsversuche	9
2.3 Klassisches Design und Modifikationen	10
3. Evaluation der Delphi-Methode	12
3.1 Zur Kritik an Delphi	12
3.2 Allgemeine Evaluationsbemühungen	13
3.3 Vergleich von Delphi mit anderen Ansätzen	15
3.4 Zusammenfassung	16
4. Designaspekte	19
4.1 Zahl der Runden	19
4.2 Anonymität	21
4.3 Expertenauswahl	22
4.4 Umfang der Expertengruppe	24
4.5 Panelmortalität bei Delphi	25
4.6 Feedback und Monitoring-Team	26
4.7 Kosten- und Zeitaufwand	27
4.8 Computerunterstützte Delphi-Anwendung	27
5. Kognitionspsychologische Aspekte für die Begründung der Funktionsweise von Delphi	30
5.1 Vorgabe bestimmter Cues	31
5.2 Tendenz zur Meinungsänderung und psychologische Dispositionen	32
5.3 Kognitionspsychologische Grundlagen der Urteilsbildung	33
5.4 Kognition und Feedback	35
6. Zusammenfassung	37
Literatur	39
Weiterführende Literatur	44

## 0. Einleitung

Seit Fertigstellung unserer ersten Literaturstudie (Häder/Häder 1994a)<sup>1</sup> können wir verschiedene neue Entwicklungstendenzen, zahlreiche weitere Anwendungen und ein insgesamt gestiegenes Interesse an der Delphi-Technik konstatieren. Beleg für die gewachsene Aufgeschlossenheit gegenüber Delphi in Deutschland sind zahlreiche Anfragen zu Beratungen über Delphi-Projekte an das ZUMA, das rege Interesse an einem 1996 bei ZUMA organisierten Workshop zur Delphi-Methode sowie die große Nachfrage nach den ZUMA-Arbeitsberichten zu diesem Thema.

Im internationalen Maßstab sprechen eine Reihe neuer Publikationen während der letzten drei Jahre für das anhaltende Interesse an dieser Methode. Dabei handelt es sich im wesentlichen einerseits um Berichte über praktische Einsätze von Delphi und andererseits um methodische Tests, die der weiteren Evaluation der Methode dienen. Diese Aufsätze sollen im folgenden referiert werden. Außerdem werden unsere eigenen Bemühungen um die Erklärung der Funktionsweise von Delphi in diese Übersicht einfließen.

Allgemein haben u.E. drei Aspekte zum anhaltenden bzw. sogar gewachsenen Interesse an der Delphi-Methode geführt:

- Entscheidungen in Wirtschaft und Wissenschaft müssen von immer mehr Experten vorbereitet und getragen werden.

Experten mit Universalwissen und Universalverantwortungsbereitschaft, die eventuell als Einzelpersonen dazu in der Lage wären, die entsprechenden Entscheidungen zu treffen<sup>2</sup>, werden seltener. An deren Stelle tritt eine Anzahl hoch spezialisierter Fachleute für einzelne Gebiete (Expertengremien).

- Entscheidungen dieser Art werden immer komplizierter und komplexer, ihr Zeithorizont unter Umständen wesentlich weiter.

So stellt gegenwärtig beispielsweise die Entwicklung und der Bau des Hochgeschwindigkeitszuges Transrapid ein nicht unumstrittenes Projekt dar, dessen tatsächlicher Nutzen sich Schätzungen zufolge erst in etwa 15 bis 25 Jahren herausstellen wird.

---

<sup>1</sup> Wir verzichten darauf, die Ergebnisse dieser Darstellung an dieser Stelle zu wiederholen und wenden uns mit diesem Arbeitsbericht stattdessen stärker an Interessenten, die bereits über Grundkenntnisse zu Delphi verfügen.

<sup>2</sup> Ein Beispiel ist der 1997 verstorbene Universalgelehrte Manfred von Ardenne.

Projekte, zu denen Beschlüsse mit ähnlich weitreichenden Konsequenzen getroffen werden müssen, dürfte es vor allem im Bildungs- und Energiebereich geben.

- Entscheidungen sind mit immer höheren Kosten verbunden.

Während die Kosten für Investitionen steigen, werden gleichzeitig die zur Verfügung stehenden Mittel auf vielen Gebieten knapper. So geht es beispielsweise beim Transrapid um geschätzte Kosten von 10.8 Milliarden DM<sup>3</sup>.

Dies alles bedeutet: Auf unsicherem Wissen basierende Entscheidungen von Expertengruppen, die weitreichende Konsequenzen haben, müssen qualifiziert werden. Das Ergebnis dieses Berichts bereits vorwegnehmend kann eingeschätzt werden, daß Delphi in diesem Zusammenhang einen Beitrag leisten kann.

## **1. Definitionen von Delphi als Indikatoren für die sich ausdifferenzierenden Erwartungen an diese Methode**

Die Bestimmungen der Delphi-Methode haben seit den 70er Jahren eine deutliche Diversifikation erfahren. Diese dokumentiert zugleich die sich ändernden, d.h. sowohl sich ausdifferenzierenden als auch diffuser werdenden Erwartungen an die in den Sozialwissenschaften immer noch relativ selten eingesetzte Methode. Geht man davon aus, daß sich in den unterschiedlichen Definitionsvorschlägen die Intentionen der Autoren bzw. Anwender ausdrücken, dann lassen sich aus der jeweiligen Charakterisierung von Delphi bereits auch bestimmte Erwartungen an diese Methode ableiten.

Eine Reihe von Autoren sieht die Delphi-Technik ganz allgemein als Instrument zur verbesserten Erfassung von Gruppenmeinungen bzw. zur Steuerung der Gruppenkommunikation an (vgl. z.B. Linstone/Turoff 1975:3). Die folgenden Beispiele demonstrieren diese Sichtweise:

- „The Delphi technique is a questionnaire method for *organizing and sharing opinion* through feedback.“ (Bardecki 1984:281, vgl. auch Dalkey/Helmer 1963.)
- „An alternative means of accessing expert opinion and *evaluating incomplete information* is the Delphi technique, a systematic procedure for soliciting the advice of a number of experts, and forging a consensus from that advice.“ (Richey et al. 1985:136)

---

<sup>3</sup> Diese Vermutung wurde in einem Artikel der „Berliner Zeitung“ vom 26. April 1997 geäußert. Vom Gesamtvolumen der Investition werden ca. 6.1 Milliarden DM für den Fahrweg und 3.7 Milliarden DM für das Betriebssystem veranschlagt. Es soll an dieser Stelle dahingestellt bleiben, ob auch aufgrund einer Delphi-Studie der Bau des Transrapid empfohlen worden wäre.

- Murry/Hammons (1995) definieren unter Rückgriff auf die Bestimmung von Delbecq/van de Ven/Gustafson (1975) Delphi als „eine Methode für die systematische *Sammlung von Urteilen* zu einem bestimmten Gegenstand mit Hilfe einer sorgfältig designten, wiederholten Fragebogenaktion, in die zusammengefaßte Informationen und Feedback über die Meinungen der anderen Teilnehmer eingestreut sind.
- Delphi sei „an accepted method of achieving *consensus among experts*,“ wobei zugleich das Hauptproblem von Delphi in der Bestimmung eben dieses „Konsens“ bestehen soll, meint Duffield (1993:227).
- Delphi „can be used for any purpose for which a committee or decision-making group is appropriate“, ist die Auffassung von Erffmeyer et al. (1986:121).

In anderen Charakterisierungen der Methode spielen stärker inhaltliche Aspekte, insbesondere das potentielle Leistungsvermögen von Delphi als qualitativer Ansatz für bestimmte Problemlösungen (vor allem für Prognosen) eine größere Rolle.

- Delphi ist ein „excellent forum for reaching a consensus about complex conceptual problems in the field of environmental assessment (and, presumably, in other multidisciplinary fields)“ (Richey et al. 1985:145).
- Die Delphi-Technik als „cornerstone of future research“ ist nach Ono/Wedemeyer (1994:290) inzwischen das am meisten genutzte Verfahren für Zukunftsvorhersagen, sie ist ein effizienter und effektiver *Gruppenkommunikationsprozeß*, der viele psychologische Nachteile von Gruppenkommunikationen vermeidet und den Experten auf ihren jeweiligen Gebieten systematisch Urteile entlockt.

Mit dem Verweis auf die psychologischen Aspekte von Delphi und einem Rückgriff auf die Entscheidungstheorie wird in einigen Definitionen ein weiterer Aspekt angesprochen:

- So definieren Häder/Häder (1995:12) beispielsweise die Delphi-Methode wie folgt: „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.“
- Die folgende Bestimmung von Kaynak et al. (1994:19) geht in eine ähnliche Richtung: „It has been indicated in such studies that the Delphi technique is suitable to use when *dealing with uncertainties* in an area of imperfect knowledge“. Die Autoren

behaupten sogar: „If basic data are sparse or lacking, there may be no alternative to the Delphi technique.“

- „Today ... a review of the literature indicates that it (das Delphi-Verfahren, die Verf.) is considered a reliable qualitative research method with potential for use in *problem solving*, decision making, and group consensus reaching in a wide variety of areas“ schreiben Murry et al. (1995:425) unter Bezugnahme auf Arbeiten von Cochran (1983), Judd (1972), Linstone/Turoff (1975) und Uhl (1983).
- Auf die Fruchtbarkeit eines psychologischen Ansatzes in der weiteren Diskussion um Delphi verweist - seine Testergebnisse zusammenfassend - auch Bardecki (1984:291). Dies betrifft z.B. die Untersuchung kognitiver Dissonanzen und des ego-involvements für die Teilnahmebereitschaft und den Grad der Neigung zu Meinungsänderungen bei Delphi-Befragungen.
- „Thus, according to Coates, Delphi is a technique of ‘last resort’, to be used when no adequate models exist upon which some statistical prediction or judgment might be based“ (Coates 1975, zit. nach Rowe 1991:236).
- Die Funktion von „Delphi as a judgment aiding/enhancing tool“ betont schließlich Rowe (1991:236).

Sackman wertete bereits 1975 über 150 Delphi-Studien aus und stellte daraufhin fest, daß es keine anerkannte Arbeitsdefinition für diese Methode gebe (zit. nach Goodman 1987:731). Dies bestätigte sich in der Analyse von Delphi-Definitionen in unserem ersten Literaturbericht (Häder/Häder 1994a) und gilt offenbar noch immer, wie die obige Zusammenstellung gezeigt hat. Den zitierten Beschreibungen von Delphi ist aber immerhin gemeinsam, daß keine Quelle Zweifel daran läßt, daß es sich bei Delphi inzwischen um eine etablierte Methode handelt. Es zeigt sich weiterhin, daß eine Entwicklung in der Charakterisierung von Delphi stattgefunden hat: Während zunächst die gruppenspezifischen Merkmale für die Bestimmung des Wesens von Delphi zentral waren, betonen neuere Bestimmungen stärker den Problemlösungscharakter bzw. das Umgehen mit Ungewißheit bei der Charakteristik der Delphi-Methode.

Zusammenfassend ist festzustellen: Die Grundidee von Delphi besteht darin, in mehreren Wellen Expertenmeinungen zur Problemlösung zu nutzen und sich in diesem Prozeß eines anonymen Feedbacks zu bedienen. Über diese allgemeine Interpretation

besteht Konsens, wie auch in den zitierten Äußerungen kaum Polemik bei der Beschreibung des Wesens von Delphi zu finden ist.

Auf eine Ausnahme muß in diesem Zusammenhang jedoch verwiesen werden: Nicht einleuchtend ist die Darstellung von Delphi als einer „Allround-Methode“, die sich immer dann als alternatives Notfallinstrument anwenden läßt, wenn andere Verfahren nicht weiterhelfen. Solche Auffassungen haben sich bei den Bestimmungen etwa von Erffmeyer et al. (1986), Kaynak et al. (1994) und Coates (1975) gefunden. Evaluationsbemühungen haben jedoch ergeben, daß die Validität und auch die Zuverlässigkeit von Delphi im Zusammenhang stehen u.a. mit den zu beurteilenden Sachverhalten. Das heißt: Auch der Einsatz von Delphi muß an bestimmte Voraussetzungen geknüpft werden, welche es nicht erlauben, von einer beliebig nutzbaren Universalmethode zu sprechen, wie dies bei einigen der zitierten Autoren der Fall ist. Wie noch zu zeigen sein wird, handelt es sich hier um eine Ursache für weitere Mißverständnisse.

Weiterhin mitunter vorgetragene - jedoch weniger fundamentale - Bedenken gegenüber der Delphi-Methode betreffen deren Beliebbarkeit bei der Design-Gestaltung und deren mangelnde Strukturiertheit. Auf diese Bemerkungen wird in den Abschnitten 3 und 4 eingegangen.

## **2. Anwendungen von Delphi**

### **2.1 Einsatzgebiete**

Die Anwendbarkeit der Delphi Methode erstreckt sich über ein breites Spektrum von Gebieten der Gesellschaft und des öffentlichen Lebens. Im weiteren wird anhand vorliegender Veröffentlichungen versucht, die in der letzten Zeit bevorzugten Einsatzgebiete von Delphi zu systematisieren, um Anregungen für die Anwendung von Delphi-Studien zu vermitteln.

- **Bildungswesen**

Richey et al. (1985:137) verweisen auf die erfolgreiche Nutzung der Delphi-Technik bei der Evaluation von Bildungsinhalten. Auch Zoski (1989) hebt die Eignung von Delphi zur Erkundung von „educational research needs“ hervor. Schließlich zählen auch Ono/Wedemeyer (1994:290) das Bildungswesen zu den typischen Einsatzgebieten von Delphi-Studien. Auf zahlreiche Literaturquellen, die die Anwendung der Delphi-Technik in der Hochschulforschung beschreiben, verweisen Murry/Hammons (1995:425, vgl. auch Bronner et al. 1991)



- Tourismus

Ono/Wedemeyer (1994:290) nennen als weiteres typisches Anwendungsgebiete von Delphi auch Vorhersagen auf dem Gebiet Tourismus. Den besonderen Erfolg von Delphi-Studien im Tourismusbereich betonen ebenfalls Kaynak et al. (1994:19). Demnach ist Delphi „one qualitative technique which had been used with substantial success in the past for predicting tourism demand and potential.“ Race/Planek (1992) schreiben: „Since that time, (gemeint ist das Jahr 1950; d. Verf.) the Delphi has been used successfully in numerous other applications, such as: forecasting of leisure trends (Chai 1977), wellness program planning and evaluation (Mullen/Gold 1988).“

- Betriebswirtschaftliche Anwendungen

Über eine erfolgreiche Anwendung von Delphi im Bereich regionale Entwicklung der Telekommunikation berichten Ono/Wedemeyer (1994). Die Delphi-Methode wurde weiterhin für Vorhersagen des Wachstums im Bereich Büroautomatisierung (Spiegel, 1987:134) sowie im Direktmarketing-Bereich (Annen 1997:66) genutzt. Delphi wurde schließlich auch zur Analyse von „organizational change“ (Brown/Heller 1981:141ff.) angewendet. Delphi ist außerdem laut Richey et al. (1985: 145) ein „excellent forum for reaching a consensus about complex conceptual problems in the field of environmental assessment (and, presumably, in other multidisciplinary fields)“.

- Politik

Das Delphi-Design wird nicht zuletzt für sogenannte „political risk analysis“ genutzt. Diese sollen das zukünftige, politisch bestimmte Geschäftsklima eines Landes im Zusammenhang mit der zu erwartenden internationalen Situation vorhersagen. Dazu werden Expertengruppen in einem Delphi-Design um entsprechende Urteile gebeten. (Drobnick 1984:315).

- Gesundheitswesen

Schließlich sind auffallend viele Artikel zu Anwendungen der Delphi-Methode im „Journal of Advanced Nursing“ erschienen. Die besondere Eignung von Delphi im Krankenhausbereich erörtern z.B. Williams/Webb (1994:181) unter Bezugnahme auf Reid (1988). Die hier vorgetragene Argumentation leuchtet ein, da mit Hilfe des Delphi-Designs berücksichtigt wird, daß der Einfluß von Meinungsführern neutralisiert werden kann, was aufgrund der strengen Hierarchien im Gesundheitswesen als besonders sinnvoll erscheint (vgl. auch Goodman 1987, Duffild 1993). Auch Ono/Wedemeyer (1994) heben das Einsatzgebiet Gesundheitswesen hervor.

Zusammenfassend läßt sich feststellen: Entsprechend der Charakterisierung von Delphi als einer Methode zur Gewinnung von Wissen über Sachverhalte, zu denen lediglich unsicheres oder unvollständiges Wissen existiert, handelt es sich um ein Instrument, welches nicht auf einen bestimmten Anwendungsbereich beschränkt ist.<sup>4</sup> Dies widerspiegelt sich in der obigen Übersicht.<sup>5</sup> Weiterhin ist jedoch zu konstatieren, daß Delphi bevorzugt für Aussagen über jene Bereiche benutzt wird, die kostenaufwendige Investitionen (z.B. in der Telekommunikation) erfordern bzw. zu denen schwer einzuschätzende Erfolgskriterien (z.B. im Bildungswesen und im Umweltschutz) vorliegen. Wenn - wie in den eben zitierten Darstellungen erfolgt - von einem erfolgreichen Einsatz von Delphi gesprochen wird, so bleibt bei zahlreichen Quellen jedoch unklar, ob diese Erfolgsmeldungen aufgrund sorgfältiger, gezielter, methodischer Vergleichsstudien zustande kommen, oder ob es sich eher um eine Legitimation der mit Delphi gewonnenen inhaltlichen Forschungsergebnisse handelt.

## **2.2 Einteilungsversuche**

Das breite Spektrum von Einsatzmöglichkeiten für die Delphi-Methode hat dazu geführt, daß verschiedene Autoren Versuche unternommen haben, die zahlreichen Varianten und Modifikationen des Standard-Designs zu kategorisieren, um gegebenenfalls Anwendungsempfehlungen aussprechen zu können. Im folgenden Abschnitt sollen diese Versuche zitiert werden, wenn auch schon eingangs darauf hingewiesen werden muß, daß diese Typologien mit Mängeln behaftet sind.

Von den zahlreichen Varianten und Modifikationen des ursprünglichen Delphi-Ansatzes sollen hier die folgenden vorgestellt werden<sup>6</sup> :

Laut Kenis (1995:1) sind zwei Arten von Delphi zu unterscheiden:

1. das klassische Vorhersage-Delphi
2. das Strategie-Delphi<sup>7</sup> (wobei letzteres wegen zahlreicher weiterer Untervarianten schwieriger zu beschreiben sei).

---

<sup>4</sup> Bei ZUMA wurden beispielsweise Delphi-Studien betreut, die die Entwicklung des Handwerks bis zum Jahr 2010, die Stadtentwicklung Sydneys im Zusammenhang mit der Vorbereitung der olympischen Spiele, die Evaluation von Maßnahmen zum Denkmalschutz im Rahmen der Stadtentwicklung, mentale-affektive Repräsentation und Traumarekonstruktion sowie die Entwicklung von Technologiefeldern betrafen.

<sup>5</sup> Weitere Einsatzgebiete wurden bereits an anderer Stelle zusammengetragen (Häder/Häder 1994).

<sup>6</sup> Auf einen vollständigen Überblick muß wegen der großen Anzahl und der mitunter zugleich nur geringen methodischen Bedeutung verzichtet werden.

Nach Strauss/Zeigler (1975) gibt es dagegen drei Typen von Delphi:

1. Numerische Delphi - sie dienen der genauen Festlegung eines minimalen Range für die Schätzung oder Vorhersage eines Problems (der am meisten verbreitete Typ)
2. Strategie Delphi - sie suchen Antworten auf gegenwärtige oder zukünftige soziale und politische Probleme (vgl. auch Turoff 1970) und
3. Historische Delphi - sie werden für die Erklärung vergangener Entscheidungen genutzt. Ihnen gilt die besondere Aufmerksamkeit der genannten Autoren.

Seeger (1979:20ff.) verwendet schließlich die möglichen Ziele des Delphi-Ansatzes zur Klassifikation der Methode:

1. Zielfindungs-Delphi
2. Problemfindungs-Delphi
3. Maßnahmen- und Strategieplanungs-Delphi
4. Ideenbewertungs- und Ideenfindungs-Delphi

Für eine Beurteilung der genannten Einteilungsversuche sollte nach dem mit ihrem Einsatz verbundenen Informationsgewinn gefragt werden. Dieser erscheint uns als nicht besonders hoch. Solange die Delphi-Technik nicht beispielsweise für bestimmte Anwendungsbereiche detaillierter ausgearbeitet wird, ist es lediglich von heuristischer Bedeutung, Delphi auf diese Weise zu kategorisieren. Erst wenn eine Einteilung aufgrund eindeutig relevanter Kriterien vorgenommen werden kann, ermöglicht dies praktische Rückschlüsse auf den Gebrauch des Verfahrens. Bis dahin besteht die Gefahr eines relativ unsystematischen Vorgehens bei den Einteilungsversuchen. Erweiterungen scheinen hier mit großer Beliebigkeit vorgenommen werden zu können.

### **2.3 Klassisches Design und Modifikationen**

In unserer Literaturstudie von 1994 hatten wir aufgrund der Analyse zahlreicher Quellen zur Delphi-Methode Merkmale des klassischen Designs herausgearbeitet. Diese Zusammenstellung ist offenbar - wie die Durchsicht neuerer Literatur ergab - nach wie vor gültig. Deshalb soll sie hier nochmals dargestellt werden (vgl. Häder/Häder 1994a):

- Verwendung eines formalisierten Fragebogens
- Befragung von Experten
- Anonymität der Einzelantworten

---

<sup>7</sup> Der Ausdruck „Policy Delphi“ wird mit Strategie-Delphi übersetzt.

- Ermittlung einer statistischen Gruppenantwort
- Information der Teilnehmer über die (statistische und verbale) Gruppenantwort
- (mehrfache) Wiederholung der Befragung

Zugleich ist aber zu konstatieren - und dies scheint typisch für die Anwendung der Delphi-Methode zu sein - daß in der letzten Zeit wiederum zahlreiche Varianten dieses Vorgehens eingesetzt wurden. Das heißt, daß bei Anwendungen von Delphi z.T. nur auf ausgewählte Elemente des klassischen Designs zurückgegriffen wird, während andere modifiziert oder ausgelassen werden. Begründet wird dieses Vorgehen häufig mit einer erwarteten Zeit- und/oder Kostenersparnis.

Zum Variantenreichtum von Delphi tragen z.B. bei:

- unterschiedliche Ansichten über die erforderliche Zahl und die Auswahl der Experten
- eine unterschiedliche Auslegung der erforderlichen Anzahl der Wellen, sie liegt den vorliegenden Quellen zufolge zwischen zwei und zehn (Woudenberg 1991:133)
- die unterschiedliche Gestaltung des Feedback, z.B. die Nutzung von Graphiken und verbalen Begründungen einzelner Expertenurteile
- die Erfragung von self-ratings der Experten über deren Kompetenz und entsprechende Gewichtung der Urteile
- der relativ beliebige Einsatz unterschiedlicher Aufgabentypen, z.B. Rating- (z.B. „Wie wünschenswert ist eine Erfindung?“), Ranking- (z.B. „In welcher Reihenfolge stehen Strategieziele?“) und Schätzaufgaben (z.B. „In welchem Jahr treten die Ereignisse x, y, z auf?“) sowie verbal zu beantwortende Aufgaben
- unterschiedliche Ansichten über Abbruch- bzw. Konsenskriterien

Der Variantenreichtum von Delphi wird in der Literatur als Nachteil und Vorteil zugleich angesehen. Nachteilig wirke sich der fehlende theoretische Hintergrund nahezu aller Modifikationen von Delphi aus. Als Vorteil gilt dagegen die Möglichkeit, Delphi relativ einfach an zahlreiche Problemstellungen anpassen zu können. Aus der Sicht des Methodikers, der für einen sorgfältigen Umgang mit den sozialwissenschaftlichen Instrumenten eintritt, überwiegen derzeit jedoch die negativen Folgen einer solchen Beliebigkeit. Da es über die Auswirkungen methodischer Variationen bei Delphi kaum gezielte Arbeiten gibt, bleibt unklar, zu welchen Folgen eine Abwandlung des ebenfalls noch nicht hinreichend erforschten klassischen Designs führt. Damit wird zugleich berechtigte Kritik provoziert, welche Delphi mangelnde Reliabilität und Validität unter-

stellt und teilweise auch empirisch nachweist. Solange es an Wissen über die Folgen von Modifikationen mangelt, sollte unseres Erachtens mit dem großen vermeintlichen Vorteil von Delphi - der scheinbar einfachen Adaption des Verfahrens - sorgsamer umgegangen werden.

### **3. Evaluation der Delphi-Methode**

#### **3.1 Zur Kritik an Delphi**

Grundsätzliche Zweifel am prinzipiellen Funktionieren des Delphi-Ansatzes finden sich in der neueren Literatur nicht mehr. Vor allem die inzwischen wahrgenommene Zuverlässigkeit von Delphi-Prognosen und erfolgreiche Anwendungen bei sehr unterschiedlichen Problemstellungen sind hierfür wahrscheinlich wesentliche Gründe. Diese mittlerweile nachweisbaren Erfolge sind allerdings auch darauf zurückzuführen, daß der Delphi-Methode zunächst von verschiedenen Kritikern mit starken Zweifeln, die zum Teil bis zur generellen Ablehnung dieser Methode reichten, begegnet wurde. Die von den Kritikern in diesem Zusammenhang angemahnten methodischen Forschungen und Evaluationsbemühungen und die zu dieser Thematik durchgeführte Kontroverse in der Literatur haben die Entwicklung von Delphi mit Sicherheit insgesamt positiv beeinflußt.

Die umfassendste Kritik an der Delphi-Methode stammt wahrscheinlich von Sackman (1974, hier zitiert nach Strauß/Zeigler 1975:28). In seinem Aufsatz fordert er die Anwender von Delphi auf, professionelle Standards zu entwickeln und Forschung zur Funktionsweise von Delphi zu betreiben, um die empirischen Ergebnisse von Delphi-Studien abzusichern, weil:

1. der Delphi-Anspruch, valide Expertenmeinungen zu generieren, wissenschaftlich unhaltbar und übertrieben sei
2. die Überlegenheit von Delphi gegenüber Einzelschätzungen und gegenüber face-to-face Gruppen nicht ausreichend belegt wäre
3. der bei Delphi erzeugte Meinungskonsens trüge
4. die Fragen zumeist verschwommen formuliert würden
5. die in einer Delphi-Studie gefundenen Antworten zum Teil nicht eindeutig ausfielen
6. Delphi-Urteile im Vergleich zu systematischen Studien über bestimmte Ereignisse flüchtig seien
7. die Anonymität der Experten unerklärliche Unwägbarkeiten erzeuge

8. Delphi nicht Mainstream sei, sondern einen besonderen interdisziplinären Präzedenzfall darstelle.

Kritik am konventionellen Delphi üben auch Beckers/de Boer (1977:27), Binsbergen/de Boer (1988:17), Daniël/Duijzer (1988:89), Binsbergen/de Boer/Maassen (1988:36) und Maassen/van Vught (1984:10f.). Sie weisen auf folgende Punkte hin:

- a) Zu viele Delphi-Studien würden unprofessionell durchgeführt, sie bedienten sich eines schlechten Fragebogendesigns und würden nur ungenügend Pretests unterzogen.
- b) Zu oft seien mehrdeutige Fragestellungen zur Aggregation von Meinungen genutzt worden.
- c) Es bestünden Mängel in bezug auf die theoretischen Grundlagen des Delphi-Ansatzes, dessen Gültigkeit und Zuverlässigkeit seien fragwürdig.
- d) Die Expertenauswahl sei ein besonders anfälliger Aspekt bei Delphi-Befragungen.
- e) Die Delphi-Methode ignoriere Erkenntnisse aus projektiven Techniken, der Psychometrie und von experimentellen Designs.
- f) Bei Delphi würde lediglich Homogenität in den Urteilen unterstützt und damit Abweichungen bestraft.

Aufgrund dieser kritischen Sicht auf die Leistungsfähigkeit der Delphi-Methode - einige der genannten Aspekte sind jedoch delphiunspezifisch und gelten generell für den Umgang mit sozialwissenschaftlich Methoden - bei verschiedenen Autoren hat es eine Vielzahl von Bemühungen um die Evaluation und die Standardisierung von Delphi gegeben. Einige sollen im folgenden vorgestellt werden<sup>8</sup>.

### **3.2 Allgemeine Evaluationsbemühungen**

Im folgenden werden zunächst Experimente dargestellt, die sich der Evaluation der Delphi-Methode insgesamt widmen. Die Leistungsfähigkeit des Delphi-Ansatzes wird dabei jeweils aufgrund des Vergleichs zwischen den durch die Expertengruppe geschätzten Ergebnissen und den „wahren Werten“ (die selbstverständlich lediglich dem Monitoring Team bekannt sind) beurteilt.

Ono/Wedemeyer (1994:291) beschreiben einen Ansatz, in dem eine Delphi-Studie von 1976 überprüft wurde. Diese sollte eine bestimmte Entwicklung im US-Bundesstaat

---

<sup>8</sup> Zu weiteren Bemühungen um die Evaluation der Delphi-Methode vgl. Häder/Häder (1994a:25ff.)

Hawaii für zwei verschiedene Zeiträume - erstens bis 1991 (15 Jahre) und zweitens bis 2006 (30 Jahre) - prognostizieren, wobei Trend- und Ereignisfragen gestellt wurden. 1991 erfolgte eine Überprüfung der Ergebnisse. Fünf Ereignisse wurden korrekt vorhergesagt, vier nicht. Die Schätzung der Trends gelang hingegen besser. Die nicht korrekte Vorhersage von Ereignissen hat nach Ansicht der Autoren jedoch Gründe, welche nicht direkt mit der Delphi-Methode zusammenhängen, z.B. die überraschende Energiekrise, die in den 70er Jahren eintrat. Außerdem waren die Experten teilweise uneinig darüber, ob 1991 ein bestimmtes Ereignis bereits stattgefunden habe oder nicht, dies deshalb, weil die Ereignisdefinition in der Delphi-Befragung unterschiedlich verstanden worden war. Die Schlußfolgerung war insgesamt ein „support to the hypothesis that Delphi technique is a valid technique for long-range forecasting.“

Über zwei Experimente mit der Delphi-Methode berichten Häder/Häder (1994b) und Häder/Häder/Ziegler (1995). Dabei wurden Antwortverteilungen für unterschiedlich skalierte Fragen aus einer bestimmten Bevölkerungsbefragung geschätzt, die den Teilnehmern nicht bekannt waren. Insgesamt wurden 192 Aufgaben bearbeitet, wobei sich zeigte, daß die Delphi-Methode insgesamt nur in sieben Prozent der Fälle völlig versagte (Häder 1996:27). In diesem Zusammenhang ist interessant, daß offenbar bestimmte Aufgabentypen von den Testteilnehmern besonders gut und andere weniger gut geschätzt werden konnten. Dies ist als ein Hinweis darauf zu interpretieren, daß Delphi lediglich unter noch näher zu fixierenden Voraussetzungen valide Resultate zu erbringen vermag.

Unterschiede im Erfolg beim Einsatz von Delphi in Abhängigkeit vom zu prognostizierenden Gegenstand bzw. Gebiet sind bei einer anderen Studie ebenfalls aufgefallen. Hier wurden Prognosen über Entwicklungen auf bestimmten Gebieten des öffentlichen Lebens den tatsächlich eingetretenen Innovationen gegenübergestellt. Das Ergebnis lautet: „Die Unterschiede zwischen den einzelnen Gebieten sind allerdings erheblich. In den Bereichen Medizin und Informationstechnik sowie in der Landwirtschaft und sogar im Umweltschutz, des weiteren beim Fachgebiet Kultur und Technik sowie in der Stadtentwicklung überwiegen die eher richtigen Schätzungen. Hierbei ist allerdings zu beachten, daß der Anteil der vollrealisierten Innovationen im Bereich des Umweltschut-

zes klein und bei der Stadtentwicklung Null ist. ... Besonders ungünstig war die Vorschau für Energie und Verkehr.“ (BMFT 1993:75).<sup>9</sup>

### **3.3 Vergleich von Delphi mit anderen Ansätzen**

Kritiker merkten in bezug auf die Leistungsfähigkeit der Delphi-Methode wiederholt an, daß den Vorteilen von Delphi gegenüber Gruppendiskussionen bzw. Brain storming („1. das Vorhandensein dominierender Persönlichkeiten in der Gruppe, 2. das Vorhandensein irrelevanter Informationen und Kommunikation, 3. Das Vorhandensein eines Gruppenzwanges zur Konformität“; Albach 1970:18) gravierende Nachteile gegenüberstünden (z.B. Konsensdruck durch Feedback und ein starres Befragungsschema). Delphi würde deshalb keineswegs zu besseren Ergebnissen führen als beispielsweise Gruppendiskussionen. In verschiedenen Experimenten wurde deshalb die Leistungsfähigkeit von Delphi im Vergleich zur Gruppendiskussion und ähnlichen Ansätzen untersucht. Parenté/Anderson-Parenté (1987) kommen z.B. aufgrund eines Tests zum Ergebnis: „Nevertheless, the general trend is toward more valid judgments over iteration“. Ihr Experiment zeigt, daß Delphi bessere kurzfristige Vorhersagen erbringt als dies im Ergebnis von Gruppenbefragungen der Fall ist. Bei Almanachfragen ergibt sich jedoch das entgegengesetzte Resultat (Brockhoff 1975): Hier sind Gruppendiskussionen leistungsfähiger.

Riggs berichtet über das folgende Experiment: Die Spielergebnisse von Footballmannschaften mit ähnlichem Punktespektrum sollten vier Wochen im voraus vorhergesagt werden. Zwei Spielergebnisse waren zu schätzen. Das erste Spiel war intensiv von den Medien kommentiert worden (woraus die Vermutung resultiert, daß über dieses Spiel relativ viele Informationen vorhanden waren), das andere Spiel nicht. Testteilnehmer waren Studenten. Eine Hälfte der gesplitteten Gruppe diskutierte die wahrscheinlichen Resultate face-to-face, die andere Hälfte nahm an einem Delphi in zwei Runden (= 2 x 8 Gruppen) teil. „The result therefore appears to support the contention that Delphi is superior to the conference method for long-range forecasting in both high and low information environments.“ (Riggs 1983:93).

Die Ergebnisse dieser Experimente zusammenfassend ist einzuschätzen, daß Delphi sich gegenüber anderen Verfahren, insbesondere der Gruppendiskussion, nicht in jedem

---

<sup>9</sup> Weitere Evaluationsbemühungen sind aus der deutschen Sozialforschung nicht bekannt. Hier wird die Delphi-Methode im Unterschied zu den USA offenbar mit einem erheblichen Zutrauen und zugleich mit



Fall als überlegene Technik erweist. Deshalb wird teilweise der Schwerpunkt der Argumentation von den Delphi-Befürwortern auf die ökonomischen Aspekte bei der Anwendung des Verfahrens gelenkt: Delphi sei kostengünstig. (Fischer 1981, zit. nach Rowe 1991:239). Tatsächlich wird Delphi häufig genau dann eingesetzt, wenn es mit vertretbarem finanziellen und zeitlichen Aufwand nicht möglich ist, Experten in face-to-face Gruppen zu rekrutieren. Damit stellt sich die Frage nach einer Entscheidung zwischen Delphi oder Gruppendiskussion ohnehin oft nicht. Häder (1996:30) weist in diesem Zusammenhang darauf hin, daß in einer Delphi-Studie (Cuhls et al. 1995) von 2300 ursprünglich kontaktierten Experten weniger als 0.5 Prozent über eine ausreichende Sachkenntnis verfügten, um einen bestimmten Fragekomplex zu beantworten. „Es erscheint zweifelhaft, daß es gelänge, eine solche exklusive Expertengruppe anders als über eine Delphi-Befragung zu rekrutieren.“

Insgesamt erscheint uns die Diskussion Delphi versus Gruppendiskussion noch relativ wenig fortgeschritten. Es fehlen Argumente, die eine Einschätzung darüber erlauben, unter welchen Bedingungen die beiden Methoden zum gewünschten Erfolg führen. Die Besonderheiten des jeweiligen Forschungsgegenstandes, die Struktur der Expertengruppe und die vorliegenden Randbedingungen sind weitere Aspekte, die für das Gelingen einer Expertenbefragung wesentlich sein dürften und die bei der Diskussion um die Entscheidung zwischen einem Delphi oder einer Gruppendiskussion u.E. stärker berücksichtigt werden sollten. Der gegenwärtig in der Literatur vorgestellte Erkenntnisstand ermöglicht keine befriedigende Antwort auf die Frage Delphi oder Gruppendiskussion bzw. Brain Storming. Praktische Gesichtspunkte stehen deshalb bei der Entscheidung oftmals zu Recht im Mittelpunkt.

### **3.4 Zusammenfassung**

Seit dem Beginn des Einsatzes der Delphi-Methode gibt es Versuche, die mit diesem Verfahren gewonnenen Ergebnisse zu beurteilen. Über einige Ansätze und über die daran geknüpften Konzepte wurde hier berichtet. Zusammenfassende Beschreibungen der Leistungsfähigkeit von Delphi kommen nun zu sehr unterschiedlichen Schlüssen.

Eher indifferent äußern sich die folgenden Autoren: „Die Delphi-Methode ist kein zuverlässiges Allheilmittel zur Lösung von Problemlösungsaufgaben. Sie versagt aber auch nicht vollständig“ (Brockhoff 1979:166f.). Es leuchtet zwar ein, wenn dem Delphi-

---

fast blindem Vertrauen angewandt.

Ansatz vorgehalten wird: „So-called expert opinions are generally less satisfactory than hard facts“ (Kaynak et al. 1994:20), besonders hilfreich ist dieser Hinweis u.E. jedoch nicht. Auch die folgende Feststellung fand sich in der Literatur: „... after all, there is no reason why Delphi *should not* be used for aiding forecasts of the near future or assessing present trends for which suitable data may be lacking. (Rowe 1991:241)“

Ein insgesamt negatives Fazit seiner sekundäranalytischen Bemühungen um die Evaluation von quantitativen Delphi-Ansätzen zieht Woudenberg (1991). Der Autor referiert zahlreiche Tests zur Bestimmung der Über- oder Unterlegenheit von Delphi-Studien beispielsweise gegenüber einer face-to-face Gruppenkommunikation. Er schätzt, daß das Delphi-Verfahren etwa nur in der Hälfte der von ihm beschriebenen Fälle überlegen ist. Da ihm diese Quote zu gering erscheint, trifft er ein kritisches Urteil über die Delphi-Methode. Bei dieser Einschätzung unterläuft Woudenberg jedoch u. E. ein Fehler: Die intervenierenden Variablen für den erfolgreichen Einsatz von Delphi werden vom Autor nicht betrachtet. Er hinterfragt nicht, *warum* Delphi in ca. 50% der Fälle nicht funktioniert, d.h. unter welchen Bedingungen Erfolg ausbleibt. Mit anderen Worten: Ein bestimmtes Werkzeug, welches immer dann versagt, wenn man es zu einem nicht angemessenen Zweck einsetzt, ist deshalb noch kein prinzipiell untaugliches Instrument. So wurde bereits angedeutet, daß Delphi mitunter in Fällen angewandt wird, in denen ein alternativer Ansatz erfolgversprechender wäre. In der Einleitung wurden z.B. Beschreibungen von Delphi zitiert, die diese Technik als eine universell nutzbare Notfallvariante verstehen. Woudenberg nimmt offenbar diese Beschreibungen ernst und folgt diesen unscharfen Definitionen. Anstatt zu fragen, unter welchen Voraussetzungen Delphi funktioniert, geht er von der (u. E. falschen) Annahme aus, daß es sich tatsächlich um eine Universalmethode handele, die letztlich immer funktionieren müsse. So gelangt er zu seiner kritischen Schlußfolgerung über die Möglichkeiten eines unscharf definierten Delphi-Ansatzes. Es wird damit nochmals deutlich, daß eine Auffassung wie die von Race/Planek, das Fehlen von Konventionen bei Delphi sei Vor- und Nachteil zugleich (1992:173), wenig hilfreich ist.

Die Zusammenstellung von Ergebnissen zur Überprüfung der Leistungsfähigkeit von Delphi ließe sich fortsetzen. Es muß jedoch auch auf Kritik an der Evaluationsmethodik selbst verwiesen werden: So werden in Laborexperimenten häufig Studenten als Experten eingesetzt. Dies ist mitunter ein durchaus adäquates Vorgehen, wenn rein methodische Fragen wie z.B. die Eignung verschiedener Skalentypen für Delphi erkundet wer-

den sollen (vgl. Häder/Häder/Ziegler 1995). Geht es jedoch darum, mit Hilfe von Wissensfragen die Leistungsfähigkeit der Delphi-Methode zu testen, wäre eine Auswahl der Teilnehmer unter dem Gesichtspunkt notwendig, daß möglichst unterschiedliche (für die Problemstellung relevante) Wissensprofile verfügbar gemacht werden. Dies ist aber bei Studenten in speziellen Seminaren zumeist nicht der Fall. Damit stellt dann auch die Rückkopplung der Ergebnisse keine Weitergabe zusätzlicher Information an die Befragten dar. So ist z.B. das Ausgangswissen bei der Einschätzung von Almanachfragen, wie etwa nach den Äpfelpreisen in einem US-Bundesland zu einen bestimmten Zeitpunkt (vgl. Dalkey 1969), bei Studenten weitgehend ähnlich.<sup>10</sup> Es müssen also anspruchsvollere und dem Problem jeweils adäquate Bemühungen zur Evaluation von Delphi angestellt werden, ein Rückgriff auf die häufig benutzten Almanachfragen und auf Labor-Delphis mit Studenten reicht nicht aus.

In diesem Zusammenhang ist auch die Diskussion um die Entwicklung von Kriterien, nach denen der Erfolg von Delphi beurteilt werden kann, als weiterer Aspekt zu nennen. Für Aufgabenstellungen im Rahmen von Evaluationen der Delphi-Technik, bei denen von Experten Schätzaufgaben gelöst werden (z.B. Almanachfragen), während die „wahren Werte“ dem Monitoring Team bekannt sind und damit für die Beurteilung der Güte der Schätzungen zur Verfügung stehen, schlägt Häder (1996:8ff) eine Lösung vor. Er behandelt u.a. die folgenden in Frage kommenden Aspekte:

1. die Verringerung des Fehlers, berechnet als Abstand zwischen dem wahren Wert und dem Median der Schätzungen in der letzten Runde gegenüber den Schätzungen in der ersten Runde
2. die Treffgenauigkeit, bestimmt aus der Lage des wahren Werts innerhalb oder außerhalb des Ranges der Schätzungen der letzten Runde - denkbar wäre auch die Anwendung eines anderen Streuungsmaßes
3. den Vergleich der Schätzungen der besten Experten mit der Gruppenleistung,
4. den Vergleich der Orientierungen der Experten bei ihren erneuten Schätzungen am wahren Wert bzw. am rückgemeldeten Gruppendurchschnitt

Dabei wird gezeigt, daß es, wendet man alle diese Kriterien an einem identischen Beispiel an, zu völlig gegensätzlichen Schlußfolgerungen über Erfolg oder Mißerfolg von Delphi kommen kann. Die Lösung sieht Häder in der Entwicklung von syntheti-

---

<sup>10</sup> Nicht zuletzt ist die Eignung von Almanachfragen zur Evaluation der Delphi-Methode (vgl. z.B. Dalkey

schen Erfolgstypen, welche auf den Kriterien Verringerung des Fehlers und Treffgenauigkeit aufbauen (1996:25ff.).

## **4. Designaspekte**

### **4.1 Zahl der Runden**

Bisher gibt es keinen Standard dafür, wie viele Runden bei einer Delphi-Befragung notwendig bzw. sinnvoll sind. Die im folgenden beschriebenen Experimente führen tatsächlich auch zu sehr unterschiedlichen Empfehlungen:

Bei einem Test wurden 72 Studenten (36 männliche und 36 weibliche) zufällig in 4-Personengruppen mit gleichem Geschlecht eingeteilt. Die Aufgabe im Rahmen des NASA Moon-Survival-Problem bestand darin, die Bedeutung von 15 Ausrüstungsgegenständen nach ihrer Wichtigkeit zu ranken, um auf dem Mond nach einem Unfall überleben zu können. Die NASA hatte dazu konkrete Vorstellungen, die als „wahre Werte“ angenommen wurden, so daß die Ergebnisse des Tests beurteilt werden konnten. Die Untersuchung wurde über sechs Delphi-Runden durchgeführt, wobei die Qualität der Ergebnisse für jede Runde ermittelt wurde. Die Ergebnisse der einzelnen Runden zeigen: „The quality scores remained stable for the succession from Round 4 to 5 and from Round 5 to 6.“ Nach vier Befragungswellen wurde also Stabilität erreicht. Die Schlußfolgerung aus diesem Test ist nach Meinung der Autoren, daß eine häufig bei Delphi zu beobachtende Begrenzung auf zwei Runden nicht hinreichend für die Erzielung eines stabilen Ergebnisses sei. Die Zusammensetzung der Gruppe, die Art des zu lösenden Problems sowie die Ausgestaltung des Feedback seien Einflüsse, die ebenfalls Einfluß auf die Anzahl der Runden haben. Dazu sind jedoch nach Meinung der Autoren weitere Forschungen erforderlich. (Erffmeyer et al. 1986:127, Erffmeyer/Lane 1984). Dieses Experiment kritisiert Woudenberg (1991), indem er vor allem das Kriterium der NASA für die richtige Auswahl der Ausrüstungsgegenstände (also die „wahren Werte“) in Frage stellt.

Andere Autoren haben in Experimenten jedoch nicht die gleichen Erfahrungen mit der notwendigen Zahl der Runden gemacht: „The efficiency of the technique was demonstrated by the fact that only two rounds were needed to attain opinion convergence.“ (Kaynak et al. 1994). Murry/Hammons gehen ebenfalls von der Notwen-

---

1969a, Becker 1974, Brown et al. 1969, Brown/Helmer 1964) insgesamt umstritten (Rowe 1991).

digkeit von lediglich zwei quantifizierenden Wellen aus (1995:429). Analog zu weiteren Autoren (Lanford 1972, Häder/Häder 1994b, Häder/Häder/Ziegler 1995) stellen sie fest, daß die größten Veränderungen zwischen der ersten und zweiten Runde auftreten. So wird kaum über Delphi-Studien außerhalb eines Labors berichtet, die über mehr als drei Wellen gelaufen sind.

Als Optimum wird eine minimale Anzahl von Runden bei einem akzeptablen Maß an erzielter Genauigkeit angesehen, wobei hinsichtlich des Kriteriums für Genauigkeit unterschiedliche Auffassungen existieren. So werden sowohl die Erreichung von Stabilität (z. B. in den oben dargestellten Experimenten) als auch von Konsens (vgl. z.B. Richey et al. 1985 und Duffield 1993) als Kriterien für den Abbruch der Befragung angesehen. Insofern verbindet sich mit der Diskussion um die erforderliche Anzahl der Runden die essentielle Frage, worin überhaupt das Ziel einer mehrfach wiederholten Befragung bestehe. Unsere Erachtens ist besonders das Konsenskriterium als allgemeines Ziel von Delphi und damit zugleich als das Abbruchkriterium der Expertenbefragung kritisch zu hinterfragen, da Konsens nicht prinzipiell mit „richtig“ gleichgesetzt werden kann. Der Erfolg einer Delphi-Erhebung sollte deshalb nicht generell aufgrund einer geringen finalen Streuung der Expertenurteile beurteilt werden. So ist es durchaus denkbar, daß auch die Ermittlung bestehender Divergenzen in den Expertenmeinungen einen Erfolg bedeuten können. Sinnvoller scheint es deshalb, jenen Autoren zu folgen, die die Stabilität der Urteile als Kriterium für den Abbruch der Delphi-Studie gelten lassen wollen.

Diese Diskussion mag allerdings dem Anwender, der Delphi außerhalb des Labors einsetzt, unter Umständen wenig hilfreich erscheinen. Die Kontroverse beachtet die praktischen Grenzen, die vor allem aus der Aufrechterhaltung der Motivation der Experten zur fortgesetzten Teilnahme sowie aus der Einhaltung eines zumeist vorgegebenen Zeitlimits resultieren, faktisch nicht. Abstrahiert man von diesen Problemen, dann muß es bei Delphi darum gehen, das bei den Experten vorhandene Wissen maximal zu aktivieren. Sollte sich bereits nach der ersten Welle zeigen, daß zu bestimmte Fragen Konsens besteht, so ist es wenig sinnvoll, diese im Frageprogramm zu belassen. Eine schrittweise Reduktion des Frageprogramms auf von den Teilnehmern divergent geschätzte Aspekte mag ein Kompromiß bei der Entscheidung über die Anzahl der erforderlichen Wellen darstellen.

Interessant ist in diesem Zusammenhang der Hinweis aus der Literatur, für die Aussonderung bestimmter Items, über deren Einschätzung Konsens besteht bzw. die als trivial angesehen werden, den Scree-Test einzusetzen. Dieses Vorgehen ist vor allem dann zu erwägen, wenn in einer vorausgegangenen qualitativen Welle, in der die Experten offen nach bestimmten Sachverhalten gefragt worden waren, eine besonders große Zahl von Aspekten genannt wurde. Das Aussondern von Items stellt damit eine Reduktion der Informationsfülle und der Menge der notwendigen Entscheidungen dar, was nicht zuletzt auch im Interesse der Aufrechterhaltung der Teilnahmemotivation der Experten liegen dürfte. Außerdem besteht in der Unterscheidung zwischen trivialen und nicht trivialen Items nach Abschluß der Delphi-Befragung ebenfalls ein wichtiges Anliegen bei der Interpretation der Ergebnisse. Mit diesen Zielen wurde der Scree Test von Zoski/Jurs eingesetzt, die seine Anwendung wie folgt beschreiben (1990:215): „This was accomplished through the examination of a graph of characteristic roots or eigenvalues plotted along the ordinate, with the factors plotted along the abscissa. The first roots show a cliff and the rest denote the rubble at the bottom.“ Die theoretischen Grundlagen dieses Vorgehens sind u.a. bei Cattell (1966) dargestellt. Weitere Diskussionen zum Einsatz des Scree-Tests bei Delphi-Befragungen finden sich bei Altshuld/Thomas (1991), Zoski/Jurs (1991) sowie Race/Planek (1992). Insgesamt zeigt sich, daß der Scree-Test bei Delphi-Studien, die sich einer relativ homogenen Fragestellung widmen und bei denen es zugleich darum geht, eine größere Anzahl von Aspekten zu beurteilen, gute Möglichkeiten für die Reduzierung der Zahl der zu beurteilenden Items bietet.

#### **4.2 Anonymität**

Die Anonymität der Experten untereinander ist ein wesentliches methodisches Design-Element von Delphi, d.h. der Delphi-Ansatz bezieht seine Legitimität nicht zuletzt aus der Anonymität der Befragungssituation. Ein Zweifel daran, daß Anonymität zu einer verbesserten Urteilsfindung beiträgt, bedeutet damit letztlich ein Zweifeln am Funktionieren des Delphi-Designs.

Gerade die Anonymität der Befragung wird jedoch mitunter als Nachteil interpretiert, da in anonymen Situationen die Teilnehmer nicht für die Urteile verantwortlich gemacht werden können, die sie abgegeben haben (Goodman 1987:730). In diesem Zusammen-

hang wird nicht ausgeschlossen, daß es aufgrund der Anonymität zu übereilten, nur unzureichend durchdachten Einschätzungen kommen könne (ebenda:732).

Immerhin lassen sich Argumente finden, die den in dieser Vermutung befürchteten Trend - so es ihn tatsächlich gibt - neutralisieren bzw. ihn überlagern: Erstens trägt Anonymität dazu bei, eine Meinungsführerschaft in der Expertengruppe zu verhindern. Zweitens führt eine anonyme Erhebungssituation dazu, die Beteiligten vor dem eventuell bei einer Meinungsänderung zu befürchtenden Prestigeverluste zu schützen. Wenn drittens Experten mit extrem abweichenden Urteilen um verbale Begründungen gebeten werden, stellt dies einen weiteren Mechanismus dar, um oberflächliche Schätzungen zu verhindern. Schließlich ist zu vermuten, daß viertens durch eine anonyme Erhebungssituation die Bereitschaft zur Beteiligung an einer Delphi-Befragung, bei der es darum geht, unter Unsicherheit ein Urteil abzugeben, erhöht wird. Kenis (1995) stellt dar, wie insgesamt 15 Nachteile von nicht anonymen face-to-face Gruppen bei einem anonymen Delphi-Ansatz umgangen werden.

### **4.3 Expertenauswahl**

Daß es sich bei der Auswahl der Experten für eine Delphi-Erhebung um eines der methodischen Hauptprobleme handelt, ist unbestritten (Brockhoff 1979:167). Dementsprechend sind diesem Aspekt zahlreiche Experimente gewidmet.

Mitunter werden relativ aufwendige Prozeduren beschrieben, die benutzt wurden, um die geeigneten Experten zu finden. Richey et al. (1985:139) schlagen z.B. ähnlich wie Murry/Hammons (1995:428) vor, vor Beginn der Delphi-Studie mit den potentiellen Experten Kontakt aufzunehmen, um diese standardisiert nach ihrem Expertenwissen zu befragen. In diesem Pre-Delphi-Survey (abgekürzt: PDS) wird außerdem nach dem den potentiellen Teilnehmern zur Verfügung stehenden Zeitfonds gefragt. Auf der Grundlage des Ergebnisses dieser Befragung wird eine Auswahl der Experten getroffen und erst anschließend mit dem eigentlichen Delphi begonnen. Die Vorteile einer solchen Strategie sind offenkundig: Es wird möglich, die Expertise der Expertengruppen optimal zu steuern, beispielsweise mit Hilfe einer ausgewogenen Einbeziehung von Experten, die sich auf jeweils bestimmten Gebieten Kompetenz zuschreiben. Allerdings gehen die Autoren nicht auf die ebenfalls offensichtlichen Nachteile einer solchen Strategie, die Motivation der Experten zur Teilnahme an zusätzlichen Befragungsrunden sowie den beträchtlich höheren Zeit- und Kostenaufwand, ein.

Auch Rowe plädiert für besondere Anstrengungen bei der Auswahl der Experten. Sein Vorschlag sind Tests vor Beginn der Delphi-Studie, um die Expertise der designierten Experten zu bestimmen und dann aufgrund dieser Ergebnisse die finale Gruppe gezielter zusammensetzen zu können (1991:242).

Spiegel berichtet über ein Design, welches ebenfalls eine Überprüfung der Expertise der Experten einschloß. Es wurde ermittelt, inwieweit die Teilnehmer dazu in der Lage seien, bekannte Werte zu schätzen (1987:135). Aus diesem Ergebnis wurde dann auf das Ausmaß der Expertise der Experten geschlossen. Diese Darstellung enthält jedoch wiederum keine Hinweise darauf, wie der Eindruck einer Prüfungssituation bei den teilnehmenden Experten neutralisiert werden kann und welcher Zusammenhang tatsächlich zwischen dem Ausmaß der Expertise und dem konkret ermittelten Faktenwissen besteht.

Zur Expertenauswahl gibt auch Duffield (1993:228) eine Anregung: Es sollte Wert darauf gelegt werden, daß die jeweiligen Experten über den notwendigen Einfluß verfügen, um die gefundenen Ergebnisse später auch in die Praxis umsetzen zu können.

Diskussionsstoff bietet mitunter die Frage, ob bei Delphi-Studien eine Zufallsauswahl der Experten erfolgen sollte. Reid (1988) untersucht 13 veröffentlichte Delphi-Studien im Gesundheitsbereich, wobei das Expertengremium nur in einem Fall mit Hilfe einer Zufallsstichprobe rekrutiert worden sei (zit. nach Williams/Webb 1994:182). „It would therefore seem to be more appropriate to recruit individuals who have knowledge of a particular topic and who are consequently willing to engage in discussion upon it without the potentially misleading title of ‘expert’“ (Goodman 1987:732). Häufiger wird bei Laborexperimenten berichtet, daß der Teilnehmerkreis über einer Zufallsauswahl zusammengestellt wurde (Zoski/Jurs 1990:214).

Da bei Delphi-Studien in der Regel nicht das Ziel besteht, die Ergebnisse derart hochzurechnen, daß sie für die Gesamtheit aller Experten zum jeweiligen Thema der Studie repräsentativ sind, erscheint eine gezielte Auswahl der Experten sinnvoller. Zudem setzen Zufallsauswahlen zunächst Kenntnisse über die Struktur der Grundgesamtheit voraus. Bei Delphi ist aber eine Bestimmung der Grundgesamtheit der potentiell zu einem Themengebiet befragbaren (kompetenten) Experten faktisch unmöglich, so daß kein Auswahlrahmen konstruiert werden kann. Aber selbst wenn ein derartiger Frame zur Verfügung stünde, würde bei einer Zufallsauswahl der Experten der Nonresponse eine größere Rolle spielen als bei einer bewußten Auswahl, so daß das



Expertengremium u.U. nicht in der gewünschten Struktur zusammengesetzt wäre. Insofern sollte sich die Frage nach Zufallsauswahlen der Experten bei Delphi nicht stellen.

#### **4.4 Umfang der Expertengruppe**

Wie verschieden die Erfahrungen in bezug auf den optimalen Umfang der Expertengruppe bei Delphi sind, zeigt die folgende Übersicht (vgl. auch Murrey/Hammons 1995):

- Eine Panel-Größe von drei Personen gilt als zu gering (Woudenberg 1991).
- Ein Minimum stellen 10 teilnehmende Personen dar (Parenté/Anderson-Parenté 1987).
- Größere Panels reduzieren den Fehler bei Delphi (Cochran 1983).
- Ein Maximum stellen 30 gut ausgesuchte Experten dar (Delbecq et al. 1975).
- Ein Maximum stellen 25 Experten dar (Brooks 1979).
- Eine Panelgröße von 1685 Experten gilt als die größte, von der bisher berichtet worden ist (Williams/Webb 1994:181).
- Es sollte keine Obergrenze für die Zahl der Teilnehmer geben, solange ein für das zu bearbeitende Thema angemessen strukturiertes Panel existiert.

Auch über ein Experiment zur notwendigen Panelgröße wird in der Literatur berichtet: In zwei unterschiedlich großen Panels ( $n=16$  und  $n=34$ ) wurden die gleichen Fragen bearbeitet. Die Ergebnisse der beiden Panels stimmten zu 92.86% überein. Diese Ähnlichkeit der Ergebnisse in beiden Gruppen bestätigt zunächst die Reliabilität von Delphi. Da kleinere Panels ( $n=16$ ) leichter zu organisieren sind und die Panelgröße in diesem Fall keinen Einfluß auf das Ergebnis hatte, gebe es aufgrund dieses Experiments keinen Grund, große Panels zu benutzen, schlußfolgert Duffield (1993:236).

Vor einer Verallgemeinerung dieses Ergebnisses ist u.E. jedoch zu warnen. Vermutlich spielt die Komplexität der Fragestellung für die Bestimmung der Größe der Expertengruppe eine wesentliche Rolle. Lediglich für den Fall, daß die Komplexität des Problems mit relativ wenigen und dafür um so kompetenteren Experten befriedigend abgedeckt werden kann, mag der geschilderte Befund zutreffen. Demgegenüber muß die Expertise der Experten und damit wahrscheinlich der Umfang der Expertengruppe bei komplexeren Studien, wenn z.B. mehrere Probleme innerhalb einer Delphi-Befragung zugleich bearbeitet oder noch unbekanntere Sachverhalte beurteilt werden sollen, größer

sein. Bei einer Problemstellung mit weniger Dimensionen kann die Rekrutierung der Experten gezielter erfolgen und damit eine optimale Teilnehmerzahl ermittelt werden.

#### **4.5 Panelmortalität bei Delphi**

Die Experten bei Delphi über einen längeren Zeitraum zur Mitarbeit zu motivieren, ist ein Problem, dem große Aufmerksamkeit geschenkt werden muß, da es in direktem Zusammenhang zum voraussichtlichen Erfolg der Studie steht. Nur wenn das Wissenspotential von der ersten bis zur letzten Runde möglichst vollständig zur Verfügung steht und entsprechend in die Rückmeldungen einfließt, ist mit einem optimalen Ergebnis zu rechnen.

Denkbar wäre erstens, daß sich die Aussteiger bei einem Delphi-Panel durch bestimmte gleiche Merkmale auszeichnen, z.B. zu den besonders unentschlossenen oder passiven Experten zählen. Diese Vermutung konnte jedoch in einem Experiment nicht bestätigt werden. Einer anderen Hypothese zufolge bedeuten stark abweichende Meinungen zugleich ein starkes ego-involvement und führen dazu, nicht vorzeitig abzubrechen, sondern die anderen Teilnehmer von der eigenen Meinung überzeugen zu wollen.

Die Ergebnisse eines Experiments zeigen schließlich, daß für die Ausfälle zwischen der ersten und zweiten Runde stärkere Nonkonformitätshaltungen typisch sind und daß demgegenüber kognitive Dissonanzen weniger bedeutend für die Entscheidung zum Ausstieg aus einer Delphi-Befragung sind. Unterstrichen werden soll die Schlußfolgerung aus diesem Experiment, daß psychologische Ansätze dazu beitragen können, die bei einer Delphi-Studie ablaufenden Prozesse besser zu verstehen. (Bardecki 1984:289ff.)

Williams/Webb (1993:184) empfehlen es zu beobachten, ob „the range of expert opinion is adequately represented in successive rounds“. Vor allem dann, wenn die Expertengruppe gezielt aus Angehörigen beispielsweise verschiedener Berufsgruppen rekrutiert wurde, erscheint es u. E. tatsächlich sinnvoll und notwendig zu kontrollieren, wie sich die gewollte Struktur der Teilnehmergruppe aufgrund von Abbrüchen verändert.

Reid (1988) „recommends a follow-up of non-respondents.“ Dieser Vorschlag hat Bedeutung für methodische Studien zu Delphi, dürfte jedoch in der Regel über die Möglichkeiten eines konkreten Projekts hinausgehen.

#### **4.6 Feedback und Monitoring-Team**

Die Rückinformation der Teilnehmer ist Grundbestandteil einer Delphi-Befragung. Die dabei benutzten Strategien (Rückmeldung von Mittelwerten und geeignet erscheinenden Streuungsmaßen, graphischen Darstellungen, Tabellen, verbalen Äußerungen usw.) sind vielfältig, es existieren bisher keine Standards für die Gestaltung des Feedbacks.

In verschiedenen Studien wurden z.B. neben den üblichen Maßen wie Mittelwerten und Streuung der Antworten auch verbale Kommentare rückgemeldet (Richey et al. 1985:142 Ono/Wedemeyer 1994:293). Jedoch gilt: „There is no way of concluding whether supporting comments or mean scores were more influential in effecting a change of opinion“ (Duffield 1993). Auch Riggs (1983:91) plädiert dafür, daß Experten mit besonders großen Abweichungen darum gebeten werden sollen, die Gründe ihrer extremen Ansichten anzugeben (1983:90; vgl. auch Bernstein/Cetron 1969). Rowe kritisiert an diesem Vorgehen, daß wenn lediglich Teilnehmer zu Worte kommen, deren Schätzungen außerhalb bestimmter Grenzen liegen, dies eine Beschränkung des Informationsgehaltes der Rückmeldung und damit eine Begrenzung der Verbesserung der Urteile bedeuten könne. Seines Erachtens sollte allen Teilnehmern die Chance zu verbalen Äußerungen gegeben werden (1991:244).

Mit den Möglichkeiten einer bewußten Manipulation von Ergebnissen einer Delphi-Studie über die Rückmeldung befaßt sich Nelson (1978). Er befürchtet, daß aus dem Wunsch nach einem bestimmten Ergebnis, z.B. in bezug auf eine prognostizierte Entwicklung, mit der dann eigene wirtschaftliche Interessen verknüpft sei könnten, Delphi mißbraucht werden könne. Dem Autor gelingt bei einem Experiment der Nachweis, daß durch wiederholte massive Eingriffe bei der Rückmeldung die Ergebnisse von Delphi-Studien in eine vorher bestimmte Richtung verändert werden können. Seine Schlußfolgerung lautet deshalb, daß das Monitoring-Team keine eigenen Interessen an bestimmten Ergebnissen einer Delphi-Studie haben sollte. Außerdem spricht das Ergebnis dafür, in Delphi-Studien den Experten *stets* die Möglichkeit für verbale Kommentare einzuräumen, diese erschwerten bewußte Manipulationen durch das Monitoring-Team (1978:55). Die von Nelson beschriebenen Experimente sind auch aus kognitionspsychologischer Sicht interessant und werden im Abschnitt 5.4 noch näher besprochen.

Auch in einer anderen Quelle wird die Auswahl des Monitoring-Teams als Problem bei Delphi-Studien thematisiert. Richey et al. (1985:137) sehen hier sogar eine beson-

dere Schwäche von Delphi. Dies erscheint jedoch als kaum plausibel, da alle sozialwissenschaftliche Studien subjektiven Einflüssen unterworfen sind.

#### **4.7 Kosten- und Zeitaufwand**

Einige wenige Veröffentlichungen befassen sich auch mit den bei einer Delphi-Studie entstehenden Kosten. Richey et al. (1985:143) stellen dazu Delphi wiederum anderen Verfahren gegenüber. Sie gelangen zu der Auffassung, daß Delphi Reisekosten spare und deshalb relativ billig sei. Die Nutzung interaktiver PC würde dagegen fünf- bis zehnfach höhere Kosten verursachen als die Verwendung von Papier und Bleistift.

Bei praktischen Anwendungen der Methode spielt auch der erforderliche Zeitfonds eine wichtige Rolle. Hier gehen die berichteten Erfahrungen weit auseinander. Sie reichen von sogenannten real-time Delphis, bei denen in einer Workshopsituation die Ergebnisse unmittelbar zurückgemeldet werden (Goodman 1987:731) bis zu - in Abhängigkeit von der Anzahl der vorgesehenen Befragungswellen - zwei Jahren (Cho et al. 1991:282). Geschka gibt als Mindestgröße sieben Monate an (1977:36).

Richey et al. (1985:141) haben den für die Experten entstehenden Zeitaufwand gemessen, er lag bei einer Delphi-Runde zwischen 6 und 10 Stunden, ohne daß er den Teilnehmern zu hoch erschien. Hier wird deutlich, wie wichtig eine ausreichende Motivation der Experten für den Erfolg von Delphi ist.

Die Ermittlung des erforderlichen Zeitaufwandes bei einem Drei-Wellen-Delphi zeigte, daß dieser sich im Mittel von der ersten Welle (25.4 Minuten) bis zur dritten Welle (16.2 Minuten) deutlich reduzierte. Zugrunde lag hier eine Selbsteinschätzung der beteiligten Experten (Häder/Häder 1994b).

Insgesamt ist es nicht besonders sinnvoll, globale Aussagen zum Zeit- und Kostenaufwand bei Delphi anzustellen, da hierfür zu viele Details im Delphi-Design eine Rolle spielen. Dies zeigen nicht zuletzt die genannten Studien.

#### **4.8 Computerunterstützte Delphi-Anwendung**

Gegenwärtig spielen die Möglichkeiten der Computerunterstützung von Delphi - einschließlich der Verwendung des Intranets - eine wichtige Rolle bei der Weiterentwicklung des Verfahrens. Bereits 1975 erwähnten Linstone/Turoff die Vorteile der Nutzung von Computern für Delphi-Befragungen. In bezug auf das Konzept des „real-time Delphis“ unter Einsatz von Computern erwarteten sie vor allem aufgrund damals noch fehlender Computervernetzung jedoch keine schnelle Umsetzung. Dies dürfte inzwi-

schen - insbesondere bei Delphi-Studien innerhalb von Firmen oder Institutionen - kein Problem mehr sein.

In bezug auf die Nutzung des Internet als Medium für Delphi-Befragungen werden derzeit erste Erfahrungen gesammelt<sup>11</sup>. Da gerade Experten häufiger über einen Zugang zum Internet verfügen dürften, sind hier interessante Ergebnisse zu erwarten.

Im folgenden sollen nun einige spezielle Verfahrensweisen beschrieben werden, bei denen sich der Computerunterstützung bedient wird.

- Group Decision Support System

Ein Grund für die relativ späte Entwicklung von Systemen wie dem Group Decision Support System (GDSS) war die bisherige starke Orientierung der Computerisierung auf den traditionellen Delphi-Ansatz für die Vorhersage und damit die Rückmeldung von Median und Streuung. Für die Verarbeitung der Antworten auf offene Fragen waren dagegen lange Zeit Computer nur eingeschränkt geeignet. Diese Situation hat sich mittlerweile verändert, so daß inzwischen einige Anwendungen von GDSS existieren.

Das relativ komplizierte Vorgehen bei GDSS wird in verschiedenen Quellen wie folgt beschrieben (vgl. Kenis/Bollaert 1992, D'Hondt/Kenis 1992 und Kenis/Verhaegen 1993):

Erster Schritt: In einer aus drei Runden bestehenden Sitzung wird zunächst eine offene Frage auf den Server gelegt, die die Experten verbal zu beantworten haben. Alle Antworten werden rückgemeldet, zusätzlich eventuelle Kommentare. Die Aufgabe besteht danach für die Teilnehmer darin, auf einer fünfstufigen Skala Zustimmung bzw. Ablehnung zu allen Antworten anzugeben. Schließlich werden eine graphische Verteilung der Antworten sowie zusätzlich die abgegebenen Kommentare rückgemeldet. Die Experten können nun bei Bedarf ihre Antworten verändern.

Zweiter Schritt: Die Experten werden aufgefordert, ihre Antwort auf die offene Frage aus der ersten Sitzung nochmals mit dem neu gewonnenen Wissen zu überdenken. Diese Antworten werden wiederum rückgemeldet und von den Experten beurteilt sowie gegebenenfalls mit weiteren Bemerkungen versehen. Dabei wird im Ergebnis ein Ranking der Antworten erstellt.

---

<sup>11</sup> Vgl.: <http://www.uni-sb.de/philfak/fb6/krause/trauma>

Dritter Schritt: Das Ergebnis des Rankings aus dem zweiten Schritt wird den Experten zur Kenntnis gegeben und von diesen erneut beurteilt und wiederum gegebenenfalls mit Kommentaren versehen.

Die prinzipiellen Vorteile derartiger Computer-Delphi insbesondere für Strategieentwicklungen betreffen vor allem die Zeit- und Arbeitersparnis. Als Nachteil muß allerdings die für die Teilnehmer entstehende Wartezeit bis zum Ende einer Runde in Kauf genommen werden. Besonders Kenis (1995) setzt sich für die Nutzung des GDSS ein, wobei die Bereitstellung der notwendigen Hard- und Software eingeplant werden müsse. Für die Gestaltung des Frageprogramms wird eine einfache Reihen- und Wahlabfolge empfohlen, um den Experten den Überblick zu erleichtern.

- Nutzung von Datenbanken

Eine andere Variante des Einsatzes von Computern bei Delphi-Befragungen schildert Brockhoff (1979). Er benutzte den Computer, um die Experten über Datenbanken mit speziellen Informationen zu versorgen. Sein Design sah vor, Experten mittelfristige Zinsentwicklungen schätzen zu lassen. Dafür konnten sie aus verschiedenen Datenbanken die ihnen wichtig erscheinenden Informationen abrufen.

- Cross-Impact-Analysen

Eine Variante zur Verbesserung der Qualität von Vorhersagen mit der Delphi-Methode beschreibt Riggs (1983). Er nutzte Cross-Impact-Analysen, da bei Prognosen davon ausgegangen werden muß, daß bestimmte Ereignisse in der Zukunft zusammenhängend auftreten bzw. ausbleiben. Ein computergestützter iterativer Prozeß sorgt deshalb bei den Cross-Impact-Analysen dafür, daß die Vorhersage dieser „interrelated events“ konsistent ist und widersprüchliche Vorhersagen entsprechend ausgesondert werden (1983:90f.).

- Semi-Markov-Konzept

Ein weiterer interessanter Ansatz ist schließlich die Umsetzung des Semi-Markov-Konzepts für die Analyse der Daten aus Delphi-Befragungen. Cho et al. haben ein entsprechendes Vorgehen entwickelt, um dem Problem von Delphi-Prognosen zu begegnen, daß „it does not consider the significance of relationships and interactions between developments.“ (1991:274). So beeinflusse die (erfolgreiche) Lösung bestimmter technologischer Entwicklungen die (erfolgreiche) Lösung anderer Entwicklungen bzw. verzögere oder verhindere diese. Zur adäquaten Berücksichtigung dieser Beziehungen zwischen den Ereignissen analysierten sie die Prognosewerte für

das Eintreten bestimmter technologischer Entwicklungen mit Semi-Markov-Modellen: „The semi-Markov model describes well the sequential technology development process in a specific field of technology having uncertainty. This model characterizes the sequential technology development process by capturing its two probabilistic characteristics, namely, the uncertain sequence of developments and the uncertain time length between successive developments.“ (Cho et al. 1991:274) Dies kann in einem Semi-Markov-Modell analog verstanden werden als: 1. das Niveau einer Entwicklung und 2. die Verweildauer auf diesem Niveau. Die Annahme bei Markov-Modellen ist nun, daß der Fortgang einer Entwicklung davon abhängt, welches Niveau bereits erreicht ist, wobei angenommen wird, daß die Entwicklung von einem Niveau zum anderen fortschreitet und daß die einzelnen Entwicklungen voneinander abhängen.

Zur Umsetzung dieses Konzepts wurde das Computerprogramm IDEAS entwickelt. Dieses erstellt ein Modell, welches die in einem Entwicklungsprozeß auftretenden Zeitabstände zwischen einzelnen Ereignissen beschreibt. Als geeignetes Beispiel kann die Aufeinanderfolge der Entwicklung von Computerchips mit stets größer werdender Speicherkapazität angenommen werden.

Insgesamt wird in der Literatur die Rolle der Computerunterstützung bei Delphi-Befragungen eher zurückhaltend diskutiert. Es wird zwar gewürdigt, daß Inter- und Intranetze eine wesentliche Erleichterung der Kommunikation unter den Experten darstellen, wesentliche Impulse für die Weiterentwicklung von Delphi seien aber von diesem eher technischen Aspekt nicht zu erwarten.

## **5. Kognitionspsychologische Aspekte für die Begründung der Funktionsweise von Delphi**

Die Bearbeitung kognitiver Aspekte im Rahmen der methodischen Forschungen zur Delphi-Methode erscheint momentan - auch angesichts beachtlicher Fortschritte in der Kognitionspsychologie - als besonders lohnenswert, da der bislang noch nahezu unerforschte Prozeß der Meinungsbildung bei den Experten inzwischen prinzipiell kognitionspsychologisch erklärbar sein dürfte. Wenn es gelänge, theoretisch zu begründen, *wie* es den Experten gelingt, *richtige* Urteile über Sachverhalte zu fällen, über die ihnen zunächst nur unvollständiges Wissen vorliegt, so hätte dies sowohl für die Legitimation von Delphi generell als auch für die Entwicklung und Beurteilung konkreter Delphi-

Designs wesentliche Implikationen. Auch die im Abschnitt 3 geschilderten teilweise sehr unterschiedlichen Erfahrungen bei der Anwendung und Evaluation von Delphi könnten sich kognitionspsychologisch erklären lassen. Auf den Mangel an kognitionspsychologischem Wissen über Delphi kann schließlich sogar ein wesentlicher Teil der gegenüber der Delphi-Methode vorgetragenen Kritik zurückgeführt werden. Deshalb handelt es sich hier u.E. um den gegenwärtigen Hauptforschungsbedarf zur Darstellung der Funktionsweise (und damit zur theoretischen Aufwertung) der Delphi-Methode. Einige Arbeiten, die sich dieses zum Ziel gesetzt haben, sollen im folgenden relativ ausführlich vorgestellt werden.

### **5.1 Vorgabe bestimmter cues**

Obwohl nicht explizit als kognitionspsychologisches Experiment konzipiert, soll zunächst ein Versuch beschrieben werden, in dem mit Hilfe der Vorgabe bestimmter Orientierungen die Expertenurteile beeinflusst werden sollten. Als Experten für die Delphi-Studie wurden bekannte Universitätsprofessoren gewonnen. Diese wurden gebeten, sich gedanklich in jeweils einen politischen Philosophen (Plato, Aristoteles, Hobbes, Machiavelli, Swift, Burke, Rousseau, Locke, Marx oder Freud) hineinzusetzen. Auf diese Weise wurden zehn Panels, jeweils eines für jeden Philosophen, mit je sechs Experten gebildet. Die Aufgabe in der ersten Runde bestand nun darin, 42 Statements mit aktuellen Problemen der westlichen Gesellschaft (z.B. Kriminalität, Menschenrechte, Wohlfahrt) zu beurteilen. Diese waren mit einer dreiteiligen Frage verknüpft: 1. Was wäre die Sicht des jeweiligen politischen Philosophen auf das Problem? 2. Wie würde er in seiner Zeit darauf reagiert haben? 3. Wie würde er heute darauf reagieren? In der zweiten Runde wurden anonym zu beantwortende multiple-choice Fragen für jedes der zehn Panels gestellt, wobei die Ergebnisse aus der ersten Runde zugrunde gelegt wurden. Im Ergebnis wurden verschiedene Auffassungen zur Lösung zeitgenössischer Probleme entwickelt, die auf dem philosophischen Erbe basieren (Strauss/Zeigler 1975:256ff.).

Dieses Experiment ist für Delphi insofern interessant, daß jeder Experte vor der kognitiven Aufgabe steht, für seine Schätzung Gedankenstützen, sogenannte Cues, zu entwickeln, die ihm bei der Problemlösung helfen. (Gigerenzer et al. 1991, Häder/Häder 1995). Die Verfügbarkeit hochwertiger Cues zeichnet letztlich den Experten aus. Durch die Vorgabe bestimmter philosophischer Denkrichtungen bei der Problemlösung in die-



sem Experiment werden diese Cues vom Monitoring-Team explizit gesteuert. Interessant wäre zu beobachten, welche Qualität die vorgeschlagenen Problemlösungen haben, inwieweit es den Experten tatsächlich gelungen ist, sich in die historischen Personen bzw. in ihre Theoriegebäude hineinzusetzen und welche weiteren Cues auf diese Weise von den Experten aktiviert oder unterdrückt worden sind. Leider finden sich dazu in der entsprechenden Literatur keine konkreten Hinweise.

## **5.2 Tendenz zur Meinungsänderung und psychologische Dispositionen**

Die psychologischen Aspekte des Urteilsverhaltens der Experten in den einzelnen Wellen werden von verschiedenen Autoren untersucht. Insbesondere die Frage, welcher Mechanismus eine Meinungsänderung der Experten in die richtige Richtung bewirkt, stellt eine Herausforderung dar. So wird unterschieden zwischen den „Swingers“, die ihre Antworten in einer Folgewelle ändern, da sie beispielsweise in ihrer Haltung unsicher sind und zugleich in ihren Urteilen am weitesten vom wahren Wert entfernt liegen, und den Experten, die am besten geschätzt haben und die ihre Antworten nicht ändern, da sie sich in ihren Einschätzungen sicher(er) fühlen. Das Funktionieren von Delphi setzt damit voraus, daß die Experten über Wissen zur Sicherheit ihres Urteils verfügen (Rowe 1991:238).

In einem Aufsatz (Bardecki 1984) wird ein Experiment zum Zusammenhang zwischen dem Grad an Dogmatismus eines Individuums und dessen Neigung zur Meinungsänderung beschrieben. Die ursprüngliche These, derzufolge dogmatische Menschen weniger zur Meinungsänderung neigen, konnte hier nicht vollständig bestätigt werden (vgl. Mulgrave/Ducanis 1975). Trotzdem wird in der Folgezeit von Delphi-Studien berichtet (Nelson 1978:46), die vor Beginn der eigentlichen Befragung mit Hilfe der Skala von Berger (1967) bzw. Rockeach (1969) den Grad an Dogmatismus bei den beteiligten Experten überprüfen.

Der Beziehung zwischen der Stärke der Dissonanz der Experten gegenüber der Gruppe sowie der Glaubwürdigkeit der Expertengruppe auf der einen Seite und der Tendenz zur Meinungsänderung auf der anderen Seite ist - unter Rückgriff auf die Literatur zur Einstellungsänderung - ein anderes Experiment gewidmet. Ausgangspunkt ist die Überlegung, daß die Rückmeldung als externer psychologischer Anker fungiert, der einen bestimmten Einfluß auf die Urteilsbildung haben kann. Nun existieren drei Möglichkeiten:

1. der Anker wird ignoriert,
2. das Urteil wird vom Anker weg verändert, z.B. um die Gesamtmeinung näher an die eigene zu bringen (Kontrasthaltung) oder
3. die Meinung wird zum Anker hin verändert (Assimilationshaltung).

Es wird vermutet, daß je größer die Dissonanz mit der Gruppe ist, desto stärker auch der Druck zur Assimilation ausfällt. Die Stärke der Dissonanz hängt nun wiederum von verschiedenen Aspekten ab:

1. von der eigenen Urteilssicherheit,
2. von der Gruppengröße,
3. von der Bedeutung des Gegenstandes,
4. von der Glaubwürdigkeit der (nicht zustimmenden) Personen und
5. von der Stärke der Ablehnung.

Aronson et al. (1963) behaupten, daß die Meinungsänderung vor allem eine Funktion der Glaubwürdigkeit der anderen Experten und der Größe des eigenen Abstandes vom Gruppenergebnis sei. Diese Annahme wurde für Nichtexperten, die demzufolge im Rahmen einer Delphi-Befragung entsprechend weniger glaubwürdige Urteile abgeben müßten, überprüft. Dazu wurde mit 300 kanadischen Land- und Waldbesitzern eine Delphi-Studie veranstaltet. Das Ergebnis zeigte zunächst, daß es bei Teilnehmern mit einem nur geringen Abstand vom Mittelwert ebenfalls eine Tendenz zur Konformität gab. Jedoch ergab sich weiterhin, daß diese Tendenz bei Teilnehmern, die in ihren Urteilen einen größeren Abstand zur Gruppenmeinung hatten (mehr als zwei Skaleneinheiten bei einer 7stufigen Skala), nicht mehr auftrat. Dies spricht für die Hypothese, daß tatsächlich die mangelnde Glaubwürdigkeit der anderen Testteilnehmer dazu beiträgt, daß keine Tendenz zum Konsens auftritt (Bardecki 1987:282ff.). Die Ergebnisse dieses Experimentes unterstützen nicht zuletzt die Bedeutung der Expertenauswahl bei Delphi-Studien. Außerdem ergänzen sie kognitionspsychologische Experimente zur Funktionsweise von Delphi.

### **5.3 Kognitionspsychologische Grundlagen der Urteilsbildung**

Eine ausführlichere kognitionspsychologische Argumentation zur Bildung der Expertenurteile in der ersten und in den folgenden Befragungswellen liefern Häder/Häder (1995). Sie versuchen die Frage zu beantworten, warum angenommen werden kann, daß mehrere Wiederholungen der Urteilsabgabe durch Experten bei

Delphi zu qualitativ höherwertigen Schätzungen führen. Die Autoren gehen grundsätzlich davon aus, daß es sich bei den Expertenschätzungen innerhalb einer Delphi-Studie um Urteilen unter Unsicherheit handele. Bei solchen Urteilen „hängt der Erfolg entscheidend davon ab, wie gut Informationssuche, Informationsbewertung und Informationsverknüpfung gelingen“ (Zimolong/Rohrman 1988:625).

Zur kognitiven Erklärung der Bildung der Expertenurteile in der ersten Runde wird das *Informationsverarbeitungsparadigma* herangezogen. Danach können generell - also auch in Alltagssituationen - unvollständige Informationen aus dem Wissensbestand des Individuums ergänzt werden. Dazu nehmen die Individuen keine passiv-photographische Abbildung der Umwelt vor, sondern konstruieren diese intern aktiv. Schließlich kommt es zur zyklischen Rückkopplung zwischen Wahrnehmung, Gedächtnis und Informationssuche.

Die *Theorienklasse mentaler Modelle* bietet dann einen weiteren Zugang zum näheren Verständnis des Urteilens unter Unsicherheit. Für den Erfolg einer Schätzung bei unsicherem Wissen ist entscheidend, ob für das Urteil ein lokales mentales Modell (lokales MM) oder ein probabilistisches mentales Modell (PMM) genutzt 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 (vgl. Gigerenzer et al. 1991).

In den Urteilen der folgenden Wellen spielen dann sowohl Lernprozesse, wie sie etwa Zimolong/Rohrman (1988:628) beschreiben als auch die Rückmeldungen eine jeweils spezifische Rolle. So stellt Schwarz (1991, vgl. auch Hippler/Schwarz/Noelle-Neumann 1989, Schwarz/Sudman 1992 und Schwarz/Strack/Hippler 1990) auf der Grundlage von Arbeiten zur flexiblen Konstruktion mentaler Repräsentationen (Barsalou 1987, 1989) zur Urteilsbildung innerhalb einer Befragung fest, daß die Untersuchungspersonen hierbei auf verschiedene Arten von Wissen zurückgreifen. Bestimmtes Wissen sei den Befragten permanent verfügbar, andere Informationen dagegen nur temporär. Wichtig für die Diskussion um die kognitionspsychologischen Grundlagen von Delphi-Studien ist nun, 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 (Bodenhausen/Wyer 1987). „Das Urteil beruht daher nur auf der Teilmenge potentiell relevanter Information, die zum Urteils-

zeitpunkt verfügbar ist - sei dies chronisch oder temporär“ (Schwarz 1991:72). Die den Teilnehmern einer Delphi-Befragung dann nach jeder Welle zurückgemeldeten Informationen bewirken schließlich, daß diese den „Suchprozeß“ in ihrem Gedächtnis nochmals aufnehmen und nach weiteren, für den jeweiligen Sachverhalt relevanten Informationen fahnden. Kognitionspsychologisch bewirkt die rückgemeldete Gruppenantwort damit einen gewissen Kontexteffekt und trägt letztlich in indirekter Weise zur Verbesserung des abgegebenen Urteils bei. Zugleich ist die Rückinformation aber auch als eine 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 als abzuweisen eingeschätzt wird.

Bei verschiedenen Experimenten zur Unterscheidung der Wirkung des Feedbacks einerseits und des wiederholten Urteilens andererseits stellte sich heraus, daß allein durch die Wiederholung von Schätzungen (ohne Feedback) validere Ergebnisse zustande kommen. „As long as panelists have *some* idea as to where the true answer should lie, then this mechanism would explain improvements in accuracy over rounds“ (Rowe 1991:247).

Überlegungen wie die zuletzt dargestellten leisten einen wichtigen Beitrag zum Verständnis des Funktionierens von Delphi und zu den Möglichkeiten der weiteren Entwicklung des Delphi-Designs. Sie deuten an, wie der Eindruck einer gewissen Beliebigkeit bzw. Zufälligkeit, mit der die Ergebnisse von Delphi zustande kommen, reduziert werden könnte.

#### **5.4 Kognition und Feedback**

Als erste untersuchten Cyphert und Gant (1970) das Problem der Manipulation von Delphi über ein verändertes Feedback in einem Experiment. Eine als wenig wichtig eingestufte Vorgabe wurde manipuliert und den Experten nach der ersten Welle im Feedback als angeblich höher bewertet rückgemeldet. Tatsächlich folgten die Experten dieser Manipulation, indem sie das entsprechende Item in der nächsten Runde als bedeutsamer einschätzten, trotzdem gelangte es nach Abschluß aller Delphi-Runden nicht unter die zehn wichtigsten Vorgaben.

Über ein Experiment zum Zusammenhang von Meinungsänderung und manipuliertem Feedback berichten auch Scheibe et al. (1975). Sie gaben den bei einer Delphi-

Befragung teilnehmenden Experten nach der ersten Runde eine falsche Rückmeldung und beobachteten dann, ob sich die Experten stärker an dieser manipulierten Rückmeldung orientierten oder ob sie dieses gefälschte Feedback ignorieren und sich (trotzdem) dem wahren Wert annähern würden. Das Ergebnis spricht für die Stärke der Delphi-Methode: Die Experten ließen sich in ihren Urteilen zwar zunächst von der (manipulierten) Rückmeldung beeinflussen, bewegten sich jedoch in der folgenden Welle in ihren Schätzungen wieder stärker auf den wahren Wert zu als nach der ersten Welle. Dieses häufig zitierte Experiment zeigt, daß bei den Experten im Spannungsfeld zwischen Konsens und Gruppenzwang auf der einen Seite und der Orientierung an leistungsfähigen individuellen Cues auf der anderen Seite, letzteres dominieren kann und eine Annäherung an den wahren Wert durchaus möglich ist.

Welty berichtet über eine Manipulation des Feedback während der gesamten Studie, also nicht nur - wie in den beiden bisher berichteten Beispielen - bei einer Rückmeldung. Die von ihm in verschiedenen Veröffentlichungen dargestellten Ergebnisse sind jedoch widersprüchlich (1971, 1972, 1973) und können deshalb hier nicht schlüssig wiedergegeben werden.

Schließlich versuchte Nelson (1978) in einem Experiment, die Möglichkeiten und Folgen massiver, über alle Runden erfolgter Manipulation von Delphi-Befragungen zu beleuchten. Dabei wurden ein fakten- sowie ein werteermittelndes Delphi betrachtet. Es stellte sich heraus, daß die Manipulation des Feedback zu einer starken Instabilität der Ergebnisse führte, wobei diese Tendenz beim faktenermittelnden Delphi noch stärker war als beim werteermittelnden.

Der kognitionspsychologisch interessante Hintergrund der Manipulation des Feedback ist die Vermutung, die Teilnehmer an einer Delphi-Befragung würden sich bei ihren Schätzungen in den Folgerunden *lediglich* am Feedback und nicht bzw. zu wenig am wahren Wert orientieren. Nun überrascht es u.E. tatsächlich nicht, wenn sich die Schätzungen in der auf eine Manipulation folgenden Runde nicht in Richtung auf den wahren Wert bewegen. Die Rückmeldung wird als weitere Orientierungshilfe bei der Beurteilung von Sachverhalten benutzt, zu denen bisher nur unvollständiges Wissen vorlag. Damit kann nicht erwartet werden, daß ein manipulierte Feedback von den Experten völlig ignoriert wird - es sei denn, die Experten würden sich gegenseitig als nicht kompetent einschätzen. Die Ergebnisse der Experimente haben aber gezeigt, daß das Feedback bei den Teilnehmern lediglich *eine* zusätzliche Erkenntnisstütze darstellt,

welche mit anderen Cues in Verbindung gebracht wird. Damit kann immerhin nicht davon ausgegangen werden, daß ab der zweiten Runde eine Bewegung der Gruppenmeinung ausschließlich in Richtung auf die (manipulierte) Rückmeldung erfolgt.

## 6. Zusammenfassung

Zunächst ist interessant, daß immer noch keine Veröffentlichung zur Delphi-Methode ohne eine - zumindest kurze - Beschreibung des Grundanliegens und des Vorgehens von Delphi auszukommen meint. Dies mag als Indiz dafür gelten, daß die Fachwelt noch immer relativ uninformiert über diese Methode ist. Damit kann trotz der stärkeren Wahrnehmung von Delphi (gemessen an der Zahl der Publikationen in der letzten Zeit) noch nicht davon gesprochen werden, daß es sich um ein allgemein be- und anerkanntes sozialwissenschaftliches Verfahren handelt.<sup>12</sup>

Zwei generelle Folgerungen sind nun aus den Darstellungen in diesem Bericht zu ziehen:

1. Der *Anwender der Delphi-Methode* muß sich darüber im klaren sein, daß er ein Verfahren nutzt, dessen Möglichkeiten und Grenzen noch nicht befriedigend erforscht sind. Beispielen, die eine erfolgreiche Anwendung von Delphi bestätigen, stehen andere gegenüber, die nicht zu dem erwarteten Ergebnis geführt haben. Eine behutsame Nutzung, d.h. der Verzicht auf methodisch unbegründete Modifikationen von Delphi und zurückhaltende Interpretationen - ein bestimmter Anteil der Ergebnisse wird sehr wahrscheinlich *nicht richtig geschätzt* worden sein - sind Voraussetzungen, über die sich ein potentieller Anwender der Delphi-Methode im Klaren sein muß.
2. Das Ziel *des Methodikers*, der sich mit Arbeiten zur weiteren Legitimierung von Delphi befaßt, muß es vor allem sein, die Voraussetzungen für einen erfolgreichen Einsatz der Methode genauer zu beschreiben. „We conclude that Delphi does have potential as judgement-aiding technique - but improvements are needed, and the basis of these improvements can only be determined by a more thorough consideration of the mechanics of judgements change within groups and a greater understanding of the factors that influence the validity of statistical/nominal groups. (Rowe 1991:249)“  
Als weiterhin *ungelöst bzw. problematisch* stellen sich nach wie vor die folgenden Aspekte von Delphi dar:

---

<sup>12</sup> Diese Einschätzung mag aus betriebswirtschaftlicher Sicht anders ausfallen.

- Die scheinbare methodische Einfachheit bzw. Beliebigkeit des Verfahrens lädt zu einem unprofessionellen Gebrauch von Delphi ein.
- Die Unschärfe in der Zielbestimmung des Einsatzes der Delphi-Methode, ihr Ruf, ein universelles Notfallinstrument zu sein, führt zu einer undifferenzierten Anspruchshaltung an Delphi und zu entsprechender Kritik, wenn diese zumeist überhöhten Ansprüche nicht eingelöst werden können.
- Die Evaluationsbemühungen um Delphi haben zumeist den Charakter von in Labor-situationen veranstalteten Fallstudien und berücksichtigten bisher zu wenig die Frage nach den generalisierbaren Einsatzvoraussetzungen dieses Verfahrens.

Damit hat sich die Situation seit unserem ersten Literaturbericht insofern verändert, daß inzwischen eine Hinwendung zu kognitionspsychologischen Überlegungen zur theoretischen Begründung der Leistungsfähigkeit von Delphi zu konstatieren ist. Dies dient zweifellos der Akzeptanz dieser Methode in der Fachwelt. Methodische Fragen des Designs werden aber nach wie vor relativ beliebig gehandhabt, experimentell abgesicherte Standards existieren im Prinzip noch immer nicht. Hinsichtlich dieser Aspekte der Delphi-Methode sind weitere Forschungsarbeiten unbedingt notwendig, um ihre Anwendbarkeit hinreichend zu legitimieren und den wachsenden Bedarf nach Unterstützung von Entscheidungen unter Unsicherheit zu befriedigen.

Abschließend wollen wir einige Forschungsprojekten zur Delphi-Methode skizzieren, von denen wir weiteren Aufschluß über das Funktionieren der Methode erhoffen.

Kurzfristig:

1. Die zukünftige erfolgreiche Anwendung der Delphi-Methode dürfte in Zusammenhang mit der kognitionspsychologischen Begründung von Expertenurteilen stehen.
2. Darauf aufbauend stellt die stärkere Hinwendung zur Inter- und Intranet-Nutzung eine reizvolle technische Bereicherung und Perspektive der Delphi-Technik dar.

Mittelfristig:

3. Es müssen methodisch gesicherte Standards erstellt und auf Foren mit Nutzern und Entwicklern von Delphi diskutiert werden.

Längerfristig:

4. Die Evaluation der Delphi-Methode kann keineswegs als abgeschlossen betrachtet werden, weitere Bemühungen sind erforderlich.

## Literatur

- Albach, H. (1970): Informationsgewinnung durch strukturierte Gruppenbefragung. In: Zeitschrift für Betriebswirtschaft, 40. Jahrgang, Ergänzungsheft, S. 11-26.
- Altschuld, J.W./ Thomas, Ph.M. (1991): Considerations in the Application of a Modified Scree test for Delphi Survey Data. In: Evaluation Review, vol. 15, No. 2, April 1991, S. 179-188.
- Annen, E. (1997): Neue Herausforderungen. In: Marketing & Kommunikation, Heft 4, S. 66-67.
- Arosen, E./ Turner, J./ Craswell, J. (1963): Communication credibility and communication discrepancy as determinates of opinion change. In: Journal of Abnormal Social Psychology 67, S. 31-36.
- Bardecki, M.J. (1984): Participants' Response to the Delphi Method: An Attitudinal Perspective. In: Technological Forecasting and Social Change 25, S. 281-292.
- Barsalou, L.W. (1987): The instability of graded structure: Implications for the nature of concepts. In: Neisser, U. (Hrsg.) Concepts and conceptual development: Ecological and intellectual factors in categorization, Cambridge University Press., S. 101-140.
- Barsalou, L.W. (1989): Intraconcept similarity and its implications for interconcept similarity. In: Vosniadou, S./ Ortony, A. (Hrsg.) Similarity and analogical reasoning, Cambridge: Cambridge University Press, S. 76-121.
- Beckers, T./ de Boers, C. (1977): De Delphi-methode: technocratie versus democratie? In: Planning, methodiek en toepassing, 2, S. 26-27 (zit. nach Kenis 1995).
- Berger, V.F. (1967): Effects of Repression and Acquiescence Response Set on Scales of Authoritarianism and Dogmatism. Unpublished masters theses, University of Oklahoma, (zit. nach Nelson 1978).
- Bernstein, G./ Cetron, M. (1969): SEER: A Delphi Approach applied to Information Processing. In: Technological Forecasting 1, S. 33-54.
- Binsbergen, P.A./ de Boer, H.F. (1988): Hoger Onderwijs in de jaren '90. CSHOB Twente, S. 16-39 (zit. nach Kenis 1995).
- Binsbergen, P.A./ de Boer, H.F./ Maassen, P.A.M. (1988): Delphi: van orakel naar methode? In: Delphi: methode of Mode, SISWO pub. 327, Amsterdam, S. 35-46 (zit. nach Kenis 1995).
- BMFT (Hrsg.) (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. In: Hippler, H.-J./ Schwarz, N./ Sudman, S. (Hrsg.) Social information processing and survey methodology, New York: Springer, S. 6-41.
- Brockhoff, K. (1979): Delphi-Prognosen im Computerdiallog. Experimentelle Erprobung und Auswertung kurzfristiger Prognosen, Tübingen: J. C. Mohr.
- Brooks, K.W. (1979): Delphi Technique: Expanding Applications. In: North Central Association Quarterly 53, no 3, S. 377-385.



- Brockhoff, K.W. (1975): The Performance of Forecasting Groups in Computer Dialogue and Face to Face Discussions. In: Linstone, H./ Turoff, M. (Hrsg.) The Delphi Method: Techniques and Applications. Addison-Wesley, London.
- Bronner, R. et al. (1991): Anforderungen an Spitzen-Führungskräfte: Ergebnisse einer Delphi-Studie. In: Zeitschrift für Betriebswirtschaft, Nr. 11, S. 1227-1242.
- Brown, A./ Heller, F. (1981): Usefulness of Group Feedback Analysis as a Research Method: Its Application to a Questionnaire Study. In: Human Relations, Volume 34, No. 2, S. 141-156.
- Cattell, R.B. (1966): The scree test for the number of factors. In: Multivariate Behavioral Research 1, No. 2, S. 245-276.
- Cho, Yun Yeong/ Jeong, Gi Ho/ Kim, Soung Hie (1991): A Delphi Technological Forecasting Approach Using a Semi-Markov Concept. In: Technological Forecasting and Social Change, No 40, S. 273-287.
- Chai, D.X. (1977): Future of leisure: A Delphi application. In: Research Quarterly, No 48, S. 518-524.
- Cuhls, K./ Breiner, S./ Grupp, H. (1995): Delphi-Bericht 1995 zur Entwicklung von Wissenschaft und Technik. Endbericht an das Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF), Fraunhofer-Institut für Systemtechnik und Innovationsforschung, Karlsruhe.
- Coates, J.F. (1975): In Defense of Delphi: A Review of Delphi Assessment, Expert Opinion, Forecasting and Group Process by H. Sackman. In: Technological Forecasting and Social Change, No 7, S. 193-194.
- Cochran, S.W. (1983): The Delphi Method: Formulating and Refining Group Judgements. In: Journal of Human Sciences 2, No. 2 (November 1983) S. 111-117.
- Cyphert, F.R./ Gant W.L. (1970): The Delphi Technique: A Tool for Collecting Opinions. In: Teacher Education. J. E. Teacher Ed. 11 (No 3) S. 417 - 425.
- D'Hondt, T./Kenis, D. (1992): A Client-server Architecture for Groupwork: Initial findings on the suitability of state-of-the-art computing environments for cross-disciplinary group dynamics. In: Proceedings of Apple European University Consortium Bruges, Apple Belgium, Brussels, S. 6-9.
- Dalkey, N./ Helmer, O. (1962/3): An Experimental Application of the Delphi Method to the Use of Experts. In: Management Science, No 9, S. 458-467.
- Dalkey, N. (1969): The Delphi Method: An Experimental Study of Group Opinion. RAND RM 5888-PR, June.
- Daniéls, J.J.M.C./ Duijzer, G. (1988): Discussie, Siswo symposiumverlag no. 327, Amsterdam, S. 81-95 (zit. nach Kenis 1995).
- Delbecq, A.L./ Van de Ven, A./ Gustafson, D. (1975): Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes. Glenview, Ill: Scott, Foresman and Company.
- Drobnick, R. (1984): Political Risk Analysis for Canada and Mexico. In: Technological Forecasting and Social Change, No 26, S. 315-353.
- Duffield, Ch. (1993): The Delphi technique: a comparison of results obtaining using two expert panels. In: International Journal of Nursing Studies, 30, No 3, S. 227-237.

- Erffmeyer, R.C./ Lane, I.M. (1984): Quality and Acceptance of an Evaluative Task: The Effects of Four Group Decision-Making Formats. In: *Group and Organization Studies*, No 9, S. 509-529.
- Erffmeyer, R.C./ Erffmeyer, E.S./ Lane, I.M. (1986): The Delphi Technique: An Empirical Evaluation of the Optimal Number of Rounds. In: *Group and Organization Studies*, 11, No 1-2, S. 120-128.
- Fischer, G.W. (1981): When Oracle Fail: A Comparison of Four Procedures for Aggregating Subjective Probability Forecasts. In: *Organizational Behavior and Human Performance*, No 28, S. 96-110.
- Geschka, H. (1977): Delphi. In: Bruckmann, G. (Hrsg.) *Langfristige Prognosen, Möglichkeiten und Methoden der Langfristprognostik komplexer Systeme*, Physica-Verlag Würzburg Wien.
- Gigerenzer, G./ Hoffrage, U./ Kleinböling, H. (1991): Probabilistic Mental Models: A Brunswikian Theory of Confidence. In: *Psychological Review*, Vol. 98, No. 4, S. 506-528.
- Goodman, C.M. (1987): The Delphi technique: a critique. In: *Journal of Advanced Nursing*, No 12, S. 729-734.
- Häder, M./ Häder, S. (1994a): Die Grundlagen der Delphi-Methode: Ein Literaturbericht. ZUMA-Arbeitsbericht 94/02, Mannheim.
- Häder, M./Häder, S. (1994b): Ergebnisse einer experimentellen Studie zur Delphi-Methode. ZUMA-Arbeitsbericht 94/05, Mannheim.
- Häder, M./ Häder, S./ Ziegler, A. (1995): Punkt- versus Verteilungsschätzung: Ergebnisse eines Tests zur Validierung der Delphi-Methode, ZUMA Arbeitsbericht 95/05, Mannheim.
- Häder, M./ Häder, S. (1995): Delphi und Kognitionspsychologie. Ein Zugang zur theoretischen Fundierung der Delphi-Methode. In: *ZUMA-Nachrichten* 37, Mannheim.
- Häder, M. (1996): Zur Evaluation der Delphi-Technik. Eine Ergebnisübersicht, ZUMA-Arbeitsbericht 96/02, Mannheim.
- 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.
- Judd, R.C. (1972): Forecasting to Consensus Gathering, Delphi Grows Up to College Needs. In: *College and University Business* 53, No 1, S. 35-38.
- Kaynak, E./ Bloom, J./ Leibold, M. (1994): Using the Delphi Technique to Predict Future Tourism Potential. In: *Marketing Intelligence & Planning* 12, No 7, S. 18-29.
- Kenis, D./ Bollaert, L. (1992): MacPolicy: a Group Decision Support System. In: *Revue des systèmes de décision*, vol. 1, Nr. 2-3, Brussel, S. 305-318.
- Kenis, D./ Verhaegen, L. (1993): The MacPolicy project: Developing a Group Decision Support System based on the Delphi method. In: Bots, P.W.G./ Sol, H.G./ Traunmüller, R. (Hrsg.) *Decision support in public administration*, Elsevier Science Publishers, Amsterdam, S. 159-170.

- Kenis, D. (1995): *Improving Group Decisions, Designing and Resting Techniques for Group Decision Support Systems Applying Delphi Principles*, Universiteit Utrecht.
- Lanford, H.W. (1972): *Technological Forecasting Methodologies: A Syntheses*. New York: American Management Association, Inc.
- Linstone, H.A./ Turoff, M. (Hrsg.) (1975): *The Delphi Method: Techniques and Applications*. Reading, Mass: Addison-Wesley Company.
- Maassen, P.A.M./ van Vught, F.A.A. (1984): Dupliek. Naar de toekomst van het hogeronderwijs? In: *Tijdschrift voor Hoger Onderwijs*, 3, nr. 1, S. 34-36 (zit. nach Kenis 1995).
- Mullen, K.D./ Gold, R.S. (1988): Wellness construct delineation: A Delphi study. In: *Health Education Research: Theory & Practice*, No 3, S. 353-366.
- Mulgrave, N.W./ Ducanis, A.J (1975): Propensity to change responses in a Delphi round as a function of dogmatism. In: Linstone, H.A./ Turoff, M. (Hrsg.) *The Delphi Method: Techniques and Applications*. Addison-Wesley, Mass.
- Murry, J.W.Jr./ Hammons, J.O. (1995): Delphi: A Versatile Methodology for Conducting Qualitative Research. In: *The Review of Higher Education*, Summer 1995, Volume 18, No. 4, S. 424-436.
- Nelson, B.W. (1978): Statistical Manipulation of Delphi Statements: Its Success and Effects on Convergence and Stability. In: *Technological Forecasting and Social Change*, No 12, S. 41-60.
- Ono, R./ Wedemeyer, D.J. (1994): Assessing The Validity of The Delphi Technique. In *Futures* 26 (No 3), S. 289-304.
- Parenté, F.J./ Anderson-Parenté, J.K. (1987): *Delphi Inquiry Systems*. In: Wright, G./ Ayton P. (Hrsg.) *Judgmental Forecasting*, New York: John Wiley & Sons.
- Race, K.E.H./ Planek, Th.W. (1992): Modified Scree Test, Further Considerations on Its Application to Delphi Study Data. In: *Evaluation Review*, vol. 16, No. 2, April, S. 171-183.
- Reid, N.G. (1988): The Delphi technique: its contribution to the evaluation of professional practice. In: Ellis, R. (Hrsg.) *Professional Competence and Quality Assurance in the Caring Professions*, Chapman and Hall, New York.
- Richey, J.S./ Mar, B.W./ Horner, R.R. (1985): The Delphi Technique in Environmental Assessment. In: *Journal of Environmental Management*, 21, No 2, S. 135-159.
- Riggs, W.E. (1983): The Delphi Technique, an Experimental Evaluation. In: *Technological Forecasting and Social Change* 23, S. 89-94.
- Rokeach, M. (1969): *Beliefs, Attitudes, and Values: A Theory of Organization and Change*. Philadelphia: Jossey-Boss.
- Rowe, G./ Wright, G./ Bolger, F. (1991): Delphi. A reevaluation of Research and Theory. In: *Technological Forecasting and Social Change* 39, S. 235-251.
- Sackman, H. (1975): *A Delphi Critique*. Lexington Books, Lexington, Massachusetts.
- Scheibe, M./ Skutsch, M./ Schofer, J. (1975): Experiments in Delphi methodology. In: Linstone, H.A./ Turoff, M. (Hrsg.) *The Delphi Method: Techniques and Applications*. Addison-Wesley, Mass.

- Schwarz, N. (1991): Assimilation und Kontrast in der Urteilsbildung: Implikationen für Fragereihenfolgeeffekte. In: ZUMA-Nachrichten 29, S. 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 Forschungsgebiets. 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. Freiburg: HochschulVerlag, Diss.
- Spiegel, B.L. (1987): Evaluating a proposed telecommunications service: some research methods used. In: Sociological Practice, 6, No 2, S. 125-148.
- Strauss, H.J./ Zeigler, L.H. (1975): The Delphi Technique and Its Uses In Social Science Research. In: The Journal of Creative Behavior, vol. 9, no 4.
- Turoff, M. (1970): The design of a policy Delphi. In: Technological Forecasting and Social Change, No 2.
- Uhl, N.P. (1983): Using the Delphi Technique in Institutional Planning. In: New Directions for Institutional Research, 37, S. 81-94.
- Welty, G.A. (1971a): A Critique of Some Long-range Forecasting Developments. In: Bulletin of the International Statistical Institute 54, S. 403 - 408.
- Welty, G.A. (1971b): A Critique of the Delphi-Technique. In: Proceedings of the American Statistical Association. Washington D.C., S. 377 - 382.
- Welty, G.A. (1973): Platon and Delphi. In: Future 5 (No 3) S.281- 286.
- Welty, G.A. (1972a): Problems of Selecting Experts for Delphi Exercises. In: Acad. Manage. Journal 15 (No 1), S. 121-124.
- Welty, G.A. (1972b): A Selection and Deflection of Expertise in Delphi Exercises. Joint ORSA-TIMS Meeting. Atlantic City.
- Williams, P.L./ Webb, Ch. (1994): The Delphi technique: a methodological discussion. In: Journal of Advanced Nursing, 19, S. 180-186.
- Woudenberg, F. (1991): An Evaluation of Delphi. In: Technological Forecasting and Social Change 40, S. 131-150.
- Zimmolong, B./ Rohrman, B. (1988): Entscheidungshilfetechnologie. In: Frey, D./ Graf Hoyos, C./ Stahlberg, D. (Hrsg.) Angewandte Psychologie. Weinheim: Psychologie Verlags Union.
- Zoski, K.W. (1989): Research needs in educational technology for the 1990er: A delphi study. Ph.D. diss, University of Toledo (zit. nach Zoski/Jurs 1990).
- Zoski, K./ Jurs, St. (1990): Priority Determination in Surveys, An Application of the Scree Test. In: Evaluation Review, vol. 14, No. 2, April, S. 214-219.
- Zoski, K./ Jurs, St. (1991): Applications for the Modified Scree Test Revisited. In: Evaluation Review, vol. 15, No. 2, April, S. 189-190.

## Weiterführende Literatur

- Becker, D. (1974). Analyse der Delphi-Methode und Ansätze zu ihrer optimalen Gestaltung. Inaugural-Dissertation zur Erlangung der Würde eines Doktors der Wirtschaftswissenschaften der Universität Mannheim.
- Berghofer, D.E. (1971): An application of the Delphi technique to educational planning. In: *The Canadian Administrator*, 10, S. 25-28.
- Brown, B./ Helmer, O. (1964): Improving the Reliability of Estimates Obtained from a Consensus of Experts, RAND Corporation, P-2986.
- Brown, B./ Cochran, S./ Dalkey, N. (1969): The Delphi Method II: Structure of Experiments, RAND Corporation, RM-5957-PR.
- Brown, B./ Cochran, S./ Dalkey, N. (1969): The Delphi Method III: Use of Self-Ratings to improve Group Estimates. RAND Corporation RM 6115-PR, November.
- Busch, H. (1972): Praktische Erfahrungen mit der Delphi-Methode. In: *Analysen und Prognosen*, S. 23-25.
- Cyphert, F.R./ Gant, W.L. (1971): The Delphi Technique: A case study. *Phi Delta Kappan*, S. 272 - 273.
- Gewald, K. (1972): Die Delphi-Methode als Instrument technischer Prognosen - Praktische Erfahrungen. In: Blohm, H./ Steinbuch, K. (Hrsg.) *Technische Prognosen in der Praxis. Methoden Beispiele Probleme*, Düsseldorf, S. 13- 18.
- Helmer, O. (1977): Delphi and Causal Cross-Impact Analysis. *Problems in Future Research*. In: *Future*, 9, S. 17-31.
- Hennings, U./ Hüber, R.P.O./ Stanke, F. (1972): Delphi-Befragungen über zukünftige Einsatzmöglichkeiten für Roboter. In: Blohm, H./ Steinbuch, K. (Hrsg.) *Technische Prognosen in der Praxis. Methoden Beispiele Probleme*, Düsseldorf.
- Hitch, P.J./ Murgatroyd, J.D. (1983): Professional communications in cancer care: a Delphi survey of hospital nurses. In: *Journal of Advanced Nursing* 8, S. 413-422.
- Kaufmann, H.-J. (1972): Methoden der technologischen Vorschau im Dienste der Forschungsplanung industrieller Unternehmen unter besonderer Berücksichtigung der Delphi-Methode, Inaugural-Dissertation, Universität Mannheim.
- Kaynak, E./ Macauley, J.A. (1984): The Delphi Technique in the Measurement of Tourism Market Potential. In: *Tourism Management*, June, S. 87-101.
- Klages, H. (1971): Brainstorming, Delphi I und Delphi II. In: *Methoden der Prioritätsbestimmung III*, Schriftenreihe Forschungsplanung, Bonn.
- 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.
- Köhler, H. (1978): Zur Prognosegenauigkeit der Delphi-Methode, zugleich eine kritische Besprechung von H. Albach „Informationsgewinnung durch strukturierte Gruppenbefragung“. In: *Zeitschrift für Betriebswirtschaft*, Nr. 48, S. 53-60.
- König, W./ Poblitzki, A. von/ Heinzl, A. (1997): Die zentralen Forschungsgegenstände der Wirtschaftsinformatik inden nächsten zehn Jahren. (<http://www.wiwi.uni-frankfurt.de/ansgar/d1/d1html>)
- Martino, J.P. (1970a): The Lognormality of Delphi Estimates. In: *Technological Forecasting* 1 (No. 4), S. 355-358.

- Martino, J.P. (1970b): The Precision of Delphi-Forecasting. In: The Futurist.
- Martino, J.P. (1972): Technological forecasting for decisionmaking. American Elsevier Publishing company, New York.
- Makridakis, Sp./ Wheelwright, St.C. (1978): The Delphi Approach. In: Forecasting, Methods and Applications, John Wiley & Sons, Santa Barbara New York Toronto.
- Overburry, R.E. (1969): Technological Forecasting a Criticism of the Delphi Technique. In: Long Range Planning, vol. 1, No 4, June, S. 76-77.
- Parenté, F.J./ Anderson, J.K./ Myers, P./ O'Brien,T. (1984): An Examination of Factors Contributing to Delphi Accuracy. In: Journal of Forecasting 3 (No. 2), S. 173-182.
- Rowlands, D.G. (1969): Technological Forecasting and the Delphi-Technique: A Reply. In: Long-Range-Planning, Vol.2, No.2, S. 78-79.
- Salancik, J.R./ Wenger, W./ Helfer, E. (1971): The Construction of Delphi Event Statements. In: Technological Forecasting and Social Change 3 (No. 1), S. 65-73.
- Salancik, J.R. (1973): Assimilation of Aggregated Inputs into Delphi Forecasts: A Regression Analyses. In: Technological Forecasting and Social Change 5 (No. 3), S. 243 - 248.
- Saliger, E./ Kunz, C. (1981): Zum Nachweis der Effizienz der Delphi-Methode. In: Zeitschrift für Betriebswirtschaft 51, S. 470-480.
- Wechsler, W. (1978): Delphi-Methode - Gestaltung und Potential für betriebliche Prognoseprozesse. München.