

Punkt- vs. Verteilungsschätzungen: Ergebnisse eines Tests zur Validierung der Delphi-Methode

Häder, Michael; Häder, Sabine; Ziegler, Andreas

Veröffentlichungsversion / Published Version
Forschungsbericht / research report

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

Empfohlene Zitierung / Suggested Citation:

Häder, M., Häder, S., & Ziegler, A. (1995). *Punkt- vs. Verteilungsschätzungen: Ergebnisse eines Tests zur Validierung der Delphi-Methode*. (ZUMA-Arbeitsbericht, 1995/05). Mannheim: Zentrum für Umfragen, Methoden und Analysen - ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-70247>

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 95/05

**Punkt- vs. Verteilungsschätzungen: Ergebnisse eines Tests zur
Validierung der Delphi-Methode**

Michael Häder, Sabine Häder, Andreas Ziegler

August 1995

ZUMA
Quadrat B2,1
Postfach 12 21 55
68072 Mannheim

Telefon: (0621) 12 46 - 281
Telefax: (0621) 12 46 - 100
E-mail: haeder@zuma-mannheim.de
zieglera@mail.uni-marburg.de

Gliederung

1. Beiträge zu den theoretischen Grundlagen der Delphi-Methode	2
1.1. Zum gegenwärtigen Stand der Ausarbeitung der Methodik	2
1.2. Zur Theorie der bei Delphi ablaufenden Denkprozesse.....	3
1.2.1. Urteilsbildung aus kognitionspsychologischer Sicht.....	4
1.2.2. Theorie mentaler Modelle.....	5
1.2.3. Die allgemein-menschliche Fähigkeit, Meinungen zu schätzen.....	7
1.3. Ziel dieses Experiments	9
2. Darstellung des Testdesigns.....	10
3. Ergebnisübersicht.....	11
3.1. Übersicht über die Ergebnisse der Schätzung vierstufiger Antwortverteilungen	12
3.2. Übersicht über die Ergebnisse der Schätzung zweistufiger Antwortverteilungen (Alternativantworten).....	15
3.3. Übersicht über die Ergebnisse der Schätzung von Mittelwerten	18
3.4. Ergebnisse der Zusatzbefragung	21
3.5. Zusammenfassung	24
4. Punkt- vs. Verteilungsschätzung: Ein methodischer Test	25
4.1. Kognitionspsychologischer Rahmen des Tests	26
4.2. Zur Homogenität der Splitt-Gruppen.....	28
4.3. Übersicht zu den Ergebnissen des Splitts	34
4.4. Zusammenfassung	40
5. Ausblick.....	41
Literatur.....	43
Anlagenverzeichnis.....	48

1. Beiträge zu den theoretischen Grundlagen der Delphi-Methode

1.1. Zum gegenwärtigen Stand der Ausarbeitung der Methodik

Der erste Verweis auf die Benutzung eines Delphi-Ansatzes stammt aus dem Jahr 1948. Damals soll diese Technik dazu benutzt worden sein, um die Ergebnisse eines Hunde- oder Pferderennens¹ vorauszusagen. In den folgenden Jahren wurde die Delphi-Methode von der RAND Corporation in 14 Experimenten für militärische Zwecke eingesetzt, beispielsweise um mögliche Ziele sowjetischer Angriffe auf die USA zu schätzen. Wissenschaftliche Ergebnisse dieser Studien sind jedoch nicht publiziert worden (vgl. Linstone/ Turoff 1975: 10). Bekannt wurde die Delphi-Methode schließlich durch einen 1964 ebenfalls von der RAND Corporation erarbeiteten "Report on a Long-Range Forecasting Study" (Gordon/ Helmer 1964). Das Ziel dieser Studie bestand in der langfristigen Vorhersage wissenschaftlicher und technischer Entwicklungen, wobei eine Zeitspanne von zehn bis fünfzig Jahren als langfristig definiert wurde. Diese Untersuchung wird in der deutschsprachigen Literatur oft als die erste bezeichnet (vgl. Albach 1970; Geschka 1977; Saliger/ Kunz 1981). Seit den 70er Jahren fand dann eine Ausbreitung der Delphi-Methode auch in Westeuropa - einschließlich Deutschland - statt. Die Anlagen dieser Folgestudien waren zunächst stark vom erwähnten RAND Corporation-Projekt beeinflusst (vgl. Helmer/ Rescher 1960). Insbesondere in der Betriebswirtschaft kam es zur Rezeption der Delphi-Methode für Prognosezwecke. Die Einsatzgebiete wurden aber schnell vielfältiger, so daß Seeger bereits 1979 schätzte, es habe innerhalb der 15jährigen Anwendungsdauer ca. 1500 Delphi-Untersuchungen verschiedenster Art gegeben (vgl. Seeger 1979: 32). Dennoch wird die Diskussion um die Leistungsfähigkeit dieser Methode auch Jahrzehnte nach ihrer Implementation in das Methodenarsenal noch kontrovers geführt. So mangelt es an einer gesicherten Beschreibung der Voraussetzungen und der Grenzen für die Verwendung der Delphi-Methode. Die Nutzung von Begriffen wie "ideales Verfahren" (BMFT 1993: 5) und "Allheilmittel" (Brockhoff 1979: 166) zur Charakterisierung der Möglichkeiten dieser empirischen Methode deutet

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

hingegen schon formal auf bestehende Ratlosigkeit bzw. auch auf Verwirrung beim Umgang mit der Delphi-Methode hin.²

In einer Literaturstudie (vgl. Häder/ Häder 1994a: 3) konnte gezeigt werden, daß die Sozialwissenschaften in Deutschland - anders als in den USA und auch im Unterschied zu Japan - bisher vergleichsweise sehr wenig von der Delphi-Methode Gebrauch gemacht haben. Ein Indiz für die Zurückhaltung der Profession gegenüber der Delphi-Methode ist beispielsweise der Umfang, mit dem diese Technik in der aktuellen Studienliteratur im Fach „Methoden der empirischen Sozialforschung“ behandelt wird. Nur in zwei von zehn daraufhin untersuchten Fachlehrbüchern findet sie überhaupt Erwähnung - und dies lediglich auf wenigen Zeilen. Hierfür mögen verschiedene Ursachen eine Rolle spielen. Nach unserer Auffassung sind vor allem die mangelnde Validierung der Ergebnisse von Delphi-Studien sowie - damit im Zusammenhang stehend - bestehende große Unklarheiten über die tatsächliche Leistungsfähigkeit dieser Methode im sozialwissenschaftlichen Forschungsprozeß dafür verantwortlich.

Um die beschriebenen Defizite abzubauen, ist ein umfangreiches Forschungsprogramm erforderlich, dessen Ziel die adäquate Beschreibung der Möglichkeiten und Grenzen der Delphi-Methode sein sollte. Gerade die bisher praktizierte relative Beliebigkeit bei der Gestaltung dieser Methode und der damit in Verbindung stehende Mangel an professionellen Standards machen die Delphi-Methode bisher anfechtbar (vgl. auch Seeger 1979: 148ff.).

1.2. Zur Theorie der bei Delphi ablaufenden Denkprozesse

Im Mittelpunkt der Diskussion um die Leistungsfähigkeit des Delphi-Ansatzes steht derzeit die Frage, warum man annehmen kann, daß sich aufgrund der bei Delphi-Studien praktizierten verschiedenen Wiederholungen der Befragung die Urteile der Experten von Welle zu Welle verbessern (und nicht nur, z.B. aufgrund von Konformitätsdruck, eine Anpassung an die Mehrheitsmeinung innerhalb einer Gruppe stattfindet).

² Seeger stellt dazu fest: " ... sicher auch die Beliebigkeit der Anwendung haben (neben anderen Faktoren, d. Verf.) es verursacht, daß in der Fachöffentlichkeit recht große Verwirrung über Zweck, Ziel und Wirkung der Methode bereitgemacht hat und bislang einem umfangreichen Verständnis der Methode im Weg gestanden hat (1979: 44)."

Dalkey ging bei seiner Begründung der Delphi-Methode davon aus, daß in „n“ Köpfen mindestens so viel Information enthalten ist, wie in einem (vgl. Dalkey 1969: 411). Die Defizite einer solchen Bestimmung werden inzwischen kritisiert und eine detailliertere Beschreibung der ablaufenden Denkprozesse wird gefordert (vgl. auch Seeger 1979: 151). Dies ist u.E. inzwischen aufgrund von weiteren Arbeiten zur Kognitionspsychologie (vgl. Hippler/ Schwarz/ Noelle-Neumann 1989; Schwarz/ Sudman 1992; Schwarz/ Strack/ Hippler 1990) möglich geworden. Zunächst soll der Prozeß der Urteilsbildung bei Delphi-Befragungen aus dieser Sicht nochmals³ zusammengefaßt werden. Danach wird der Frage nachgegangen, inwieweit in der Theorie mentaler Modelle Hinweise für das Funktionieren der Delphi-Methode enthalten sind.

Es gilt jedoch schon an dieser Stelle festzuhalten, daß eine einheitliche Theorie, die die bei Delphi-Studien ablaufenden Denk- bzw. Informationsverarbeitungsprozesse umfassend erklärt, (derzeit) nicht existiert.

1.2.1. Urteilsbildung aus kognitionspsychologischer Sicht

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

Die den Teilnehmern einer Delphi-Befragung dann nach jeder Welle zurückgemeldeten Informationen sollen bewirken, daß diese den "Suchprozeß" in ihrem Gedächtnis nochmals aufnehmen und nach weiteren für den jeweiligen

³ Zur ausführlicheren Darstellung vgl. Häder/Häder 1994b.

Sachverhalt relevanten Informationen fahnden. Kognitionspsychologisch soll die rückgemeldete Gruppenantwort damit einen gewissen Kontexteffekt bewirken.

Diese kognitionspsychologischen Überlegungen und Experimente zusammenfassend ist festzustellen: Die befragten Experten verfügen über mehr Wissen, als sie in der ersten Befragungswelle für die Beantwortung einer Frage benutzen. Das ursprünglich nichtaktivierte Restwissen wird von ihnen in Verbindung mit den aus der Rückmeldung - z.B. von Maßen der zentralen Lage und der Streuung - gewonnenen Informationen im Rahmen einer Delphi-Befragung genutzt.

1.2.2. Theorie mentaler Modelle

Ein weiterer kognitionspsychologischer Aspekt, der zur Begründung der Leistungsfähigkeit der Delphi-Methode herangezogen werden soll, basiert auf der Theorie mentaler Modelle. Dieser Theorie folgend ist für den Erfolg einer Schätzung bei unsicherem Wissen Situation entscheidend, ob für das Urteilen ein lokales mentales Modell (local MM) oder ein probabilistisches mentales Modell (PMM) benutzt wird. Während bei lokalen MM nur partielles Wissen aktiviert wird, greifen PMM in größerem Umfang auf Erfahrungswissen aus dem Alltag (bzw. gegebenenfalls auf Expertenwissen) zurück.

Aufgrund dieses Ausgangsgedankens kann auf eine vor einiger Zeit an der Delphi-Methode geäußerte Kritik eingegangen werden. So wurde vorgebracht, daß das „Delphi-Verfahren nicht das Modell oder den Hintergrund offenlegt, auf dem die Befragten ihre Meinung aufbauen bzw. ableiten“ (Seeger 1979: 151). Inzwischen liegen jedoch Theorien und Modelle vor, die sich um die Aufdeckung der kognitiven Quellen des Meinungsbildungsprozesses bemühen.

Mentale Modelle dienen in der Kognitionspsychologie zur Beschreibung von Prozessen, die beim Urteilen unter Unsicherheit ablaufen (vgl. Dutke 1994: 33). Auf diese Weise werden zunächst Vorgänge betrachtet und erklärt, bei denen es sich um alltägliche Anforderungen an die menschliche Informationsverarbeitung handelt (Entscheidungen zu treffen, ohne daß dafür vollständiges Wissen vorliegt). Mentale Modelle werden aber auch dazu benutzt, um im Rahmen der Kognitionspsychologie replizierbare, stabile Fehlleistungen zu erklären (vgl. Tversky/ Kahnemann 1974). Hierzu wurden bereits verschiedene experimentelle

Arbeiten vorgelegt (vgl. Nisbett/ Ross 1980, Gigerenzer 1991a, 1991b, Gigerenzer/ Kleinbölting/ Hoffrage 1991 u.a.).

Im Zusammenhang mit dem hier diskutierten Problem - eine kognitionspsychologische Begründung für die Leistungsfähigkeit der Delphi-Methode zu finden - soll besonders auf die Ergebnisse von Studien verwiesen werden, die ermittelten, daß Versuchspersonen bei Experimenten dazu in der Lage sind, Häufigkeitsschätzungen relativ korrekt abzugeben, während ihnen jedoch die Beurteilung von Einzelfällen nicht gelingt. Gigerenzer et al. (1991: 513ff.) schildern das folgende Experiment:

Versuchspersonen bekamen die Aufgabe gestellt, zu sagen, welche von zwei Städten die größere Einwohnerzahl besitze. Dazu waren jeweils zwei Städte (z.B. Solingen und Heidelberg) zu vergleichen („Welche Stadt hat mehr Einwohner?“). Nach der Abgabe ihres Urteils wurden die Zielpersonen danach gefragt, wie sicher sie wären, daß ihre Antwort richtig sei. Diese *Einzelfallbeurteilung* gelingt den Versuchspersonen jedoch nicht: Sie überschätzen im allgemeinen ihre eigene Urteilsfähigkeit. Dieses Phänomen wurde bereits auch in anderen Experimenten ermittelt und mit dem Begriff „overconfidence bias“ bezeichnet (vgl. Nisbett/ Ross 1980: 119f.). Werden jedoch die gleichen Versuchspersonen, nachdem sie 50 solcher Urteile abgegeben haben, danach gefragt, wieviele dieser Urteile richtig gewesen seien, so gelingt ihnen diese *Häufigkeitsschätzung* „erstaunlich akkurat“ (Dutke 1994: 35). Ein solches Phänomen wird von Gigerenzer et al. (1991: 511) als „Confidence-Frequency-Effekt“ bezeichnet. Dieser zunächst überraschende Effekt läßt sich mit Hilfe der Theorie mentaler Modelle erklären:

Demnach wird bei Entscheidungssituationen zunächst nach einem lokalen mentalen Modell gesucht und gehandelt. Dieses ermöglicht den direkten Abruf des erforderlichen Wissens. Lokale MM sind demzufolge auf die Lösung der jeweiligen Aufgabenstellung begrenzt und berücksichtigen kein weiteres Wissen. Außerdem benutzen sie lediglich elementare logische Operationen (z.B.: wenn nicht A, dann B). Damit werden lokale MM beispielsweise für die Lösung von Einzelfallbeurteilungen benutzt und führen zu dem bereits beschriebenen und in Experimenten angetroffenem Overconfidence bias.

Bei Häufigkeitsschätzungen werden jedoch sogenannte „probabilistic mental models“ (PMM) angewendet. Diese unterscheiden sich dadurch von den lokalen MM, daß sie nicht lokal begrenzt sind, sich also nicht nur auf die Entscheidung für eine der vorgegebenen Alternativen beschränken (vgl. auch Dutke 1994: 36).

Außerdem besteht ein weiterer Unterschied darin, daß bei den PMM zusätzliches Wissen in die Lösung der Aufgabenstellung mit einbezogen wird. Schließlich kommen hier - im Unterschied zu den lokalen MM - komplexe Schlußfolgerungsprozesse zum Einsatz. PMM führen so zu den experimentell nachweisbaren besseren Urteilsergebnissen als lokale MM.

Da auch - wie noch im einzelnen darzustellen sein wird - im Rahmen von Delphi-Befragungen von den Experten Aufgaben zu bewältigen sind, für deren Lösung lediglich unvollständiges Wissen vorliegt, ist es naheliegend zu überprüfen, inwieweit die Theorie der PMM zur Erklärung der Ergebnisse bzw. der bei diesem spezifischen Delphi-Ansatz ablaufenden kognitiven Prozesse beitragen kann.

1.2.3. Die allgemein-menschliche Fähigkeit, Meinungen zu schätzen

Einen stärker empirisch begründeten Hinweis auf die menschliche Fähigkeit, nicht nur Meinungen anderer wahrzunehmen, sondern diese sogar "statistisch" zu schätzen, findet sich schließlich auch bei bei Noelle-Neumann in ihrem neu aufgelegten Buch „Öffentliche Meinung. Die Entdeckung der Schweigespirale“. Sie versteht den Menschen zunächst als ein soziales Wesen und schlußfolgert daraus, daß er permanent mit einer "Isolationsfurcht" konfrontiert werde. Diese hat die Autorin mittels empirischer Experimente nachgewiesen (vgl. Noelle-Neumann 1989: 59 ff.; neuere Experimente werden auf S. 303 ff. beschrieben). Zur Vermeidung dieser Isolation beobachten sich die sozialen Wesen untereinander und lernen so, die Meinungen der anderen (insbesondere Mehrheitsmeinungen) zu erkennen. Diese Wahrnehmung bewirke soziale Integration und sei so eine Voraussetzung für das Funktionieren einer Gesellschaft. "Die menschlichen Gesellschaften versichern sich auf diese Weise ihres Zusammenhalts und einer ausreichenden Bereitschaft des einzelnen zum Kompromiß" (ebenda: 167).

Noelle-Neumann geht in ihrer Argumentation noch weiter, sie spricht nicht nur von der Entdeckung einer neuen menschlichen Fähigkeit, der "Wahrnehmung von Meinungsklima" (ebenda: 27) sondern kennzeichnet diese Begabung sogar als "quasistatistisch" (ebenda: 31). Die folgende längere Textstelle enthält eine gute Zusammenfassung der auch für das vorzustellende Delphi-Experiment relevanten Gedanken:

"Besitzen wir überhaupt einen quasi-statistischen Sinn, wie es in der Theorie der öffentlichen Meinung behauptet wird? Können Menschen das Meinungsklima wahrnehmen?"

In allen Ländern, in denen es erprobt wurde, wurde auf Fragen in der Art 'Wie denken die meisten?' 'Sind die meisten dafür oder dagegen?' oder sogar 'Wieviel Prozent der Bevölkerung sind dafür? dagegen?' ohne weiteres geantwortet. Eigentlich müßten die Befragten antworten: 'Wieso fragen Sie mich? Sie sind doch der Meinungsforscher!' Aber das sagen sie nicht. Die Bereitwilligkeit der Schätzungen wird als Indiz angesehen, daß die Menschen diese Stärkeverhältnisse laufend selbst einzuschätzen versuchen, also ganz unabhängig vom Interview.

Aber oft stimmen die Schätzungen nicht. Einstellungen, die in einzelnen Medien vertreten werden, werden oft überschätzt. Es hat sich dafür der Ausdruck 'pluralistic ignorance' eingebürgert. 'Die Bevölkerung täuscht sich über die Bevölkerung.' ... Aber die ganze Bevölkerung, unabhängig vom eigenen Standpunkt, bemerkt, demoskopisch nachweislich, welche Meinungen zu- und welche abnehmen, so wie man merkt, ob es kälter oder wärmer wird. Wie sollte man das erklären wenn nicht durch eine menschliche Fähigkeit, Häufigkeitsverteilungen wahrzunehmen?" (Noelle-Neumann 1989: 316 f.)

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

1.3. Ziel dieses Experiments

Mit dem hier vorzustellenden Test zur Validierung der Delphi-Methode wird *erstens* das Ziel verfolgt, weitere Erfahrungen im Umgang mit dieser Technik zu sammeln und diese entsprechend zu verallgemeinern.⁴

Gerade für Themenstellungen, die sich einer Bearbeitung durch direkte Bevölkerungsbefragungen bzw. Befragungen spezieller Subpopulationen entziehen, da sie

- besonders heikle Fragestellungen beinhalten und/oder
 - bei den Befragten nicht aktuell präsent sind (z.B. da sie in der Vergangenheit liegen)⁵ und/oder
 - zu kompliziert bzw. zu komplex für die Bearbeitung in Form einer Bevölkerungsbefragung sind und/oder
 - sich auf Zielpopulationen richten, die schwer bzw. nicht erreichbar sind⁶,
- für die aber eine Abbildung der Meinungen bzw. Einstellungen aus forschungstheoretischen Erwägungen dennoch unverzichtbar ist, soll die Eignung des Delphi-Ansatzes zur Aktivierung von Expertenwissen untersucht werden.

⁴ Vor dem beschriebenen Hintergrund wurde zunächst ein anderer Test konzipiert: Im Frühjahr 1994 wurden 20 Mitarbeiter des ZUMA darum gebeten, ausgewählte, ihnen nicht bekannte Ergebnisse der Shell-Jugendstudie von 1992 zu schätzen. Um möglichst viele methodisch relevante Aussagen gewinnen zu können, wurden verschiedene Aufgabentypen eingesetzt: die Schätzung vierstufiger Randverteilungen, die Schätzung von Antwortverteilungen bei Alternativfragen sowie die Schätzung von Mittelwerten aus siebenstufigen Antwortvorgaben. Dazu wurden Indikatoren zum Zeitbewußtsein und Wichtigkeitsurteile aus einer Erhebung bei (n=4005) 13- bis 29-jährigen Jugendlichen im Jahr 1991 in Deutschland verwandt. Es fanden insgesamt drei Befragungswellen statt. Der Verlauf dieses ersten Tests ist im Detail an anderer Stelle dokumentiert worden (vgl. Häder/ Häder 1994b).

⁵ Es würde sogar eine wesentliche Bereicherung des sozialwissenschaftlichen Methodenarsenals darstellen, wenn es gelänge - eine ausreichende Validität vorausgesetzt - auch retrospektiv die Ergebnisse von Bevölkerungsbefragungen zu schätzen. Hier mag ein Verweis auf die sich im Rahmen der gegenwärtigen Forschungen zur sozialen Transformation in Ostdeutschland erschließenden Möglichkeiten - z.B. Vergleichsdaten zu erheben bzw. zu schätzen - genügen.

⁶ Dies könnte beispielweise im Rahmen sozialwissenschaftlicher Forschungen zur AIDS-Prävention gelten. Hier wurde - zwar nicht mit Hilfe der Delphi-Methode - „der Zugang über Prostituierte gewählt und weibliche Prostituierte als Expertinnen über ihre Freier befragt“ (Markert 1994: 369). Eine sinnvolle Nutzung des Delphi-Ansatzes könnte hier sicherlich zu einer weiteren Qualifizierung der Ergebnisse beitragen.

Dabei - und dies sei nochmals unterstrichen - will die beschriebene Form der Delphi-Nutzung keinesfalls Befragungen ersetzen. Sie stellt nur bei Problemen wie den genannten eine potentiell mögliche Lösung dar und macht sie damit der empirischen Analyse zugänglich. Rowe et al. (1991: 236) stellen übereinstimmend mit den Autoren in bezug auf die Nutzung von Delphi-Befragungen fest:

„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 (vgl. dazu Coates 1975).“

Zweitens sollen die vorgestellten hypothetischen Überlegungen zur Nutzung von PMM in einem Splitt (vgl. Abschnitt 4) weiter empirisch überprüft werden.

2. Darstellung des Testdesigns

Der Test wurde im November und Dezember 1994 an der Universität Marburg am Institut für Medizinische Soziologie in drei Wellen sowie einer Zusatzbefragung erhoben. Die Befragungen fanden jeweils im Rahmen einer Lehrveranstaltung statt. Dieses Vorgehen ermöglichte eine zeitlich straffe Form der Erhebung. Die Fragebögen sind während eines Seminars unter Anleitung ausgefüllt worden. Die jeweils folgende Welle schloß sich im Abstand einer Woche an, wodurch eine sehr kurze Testzeit erreicht werden konnte. An diesen Test schloß sich eine Befragung (vierte Welle) an, mit deren Hilfe jeweils die Eindrücke der Befragten, die für die Beantwortung eventuell benutzten kognitiven Orientierungshilfen, die für das Ausfüllen benötigte Zeitdauer usw. ermittelt worden sind. Ausfälle sind nicht aufgetreten. Am Test haben insgesamt 32 Studenten teilgenommen.

Das Design wurde am ZUMA erarbeitet. Hier wurden auch die Erhebungsunterlagen für die einzelnen Befragungen angefertigt.

Der Inhalt des Testes bestand darin, einige von einer Subpopulation (westdeutsche Jugendliche im Alter von 18 bis 29 Jahren) erzielte Ergebnisse der Shell Jugendstudie von 1992 zu schätzen.

Im Rahmen des Tests wurde neben Aufgaben, die für alle Teilnehmer identisch waren, folgender Splitt eingesetzt: Ein Teil der Experten ($N_1=16$) hatte die Randverteilungen zweier fünfstufiger Antwortskalen zu schätzen, der andere Teil der Experten ($N_2=16$) dagegen deren Mittelwerte. Damit sollen einige

kognitionspsychologische Fragestellungen (vgl. Abschnitt 1.2.2. dieses Berichts) weiter bearbeitet werden.

3. Ergebnisübersicht

Bei der folgenden Darstellung der Testergebnisse derjenigen Aufgaben, die von der Gesamtheit der Experten zu bearbeiten war, kommen verschiedene Kriterien zum Einsatz, nach denen die Qualität der Ergebnisse des Delphi-Tests beurteilt werden sollen. Diese Kriterien beschreiben jeweils einzelne, aufgrund ihrer Bedeutung unterschiedlich zu gewichtende Aspekte der Testresultate und liefern damit spezifische Hinweise auf die Ergebnisqualität insgesamt.

1. Das schwächste Kriterium besteht darin, ob die Streuung (Range) aller Expertenurteile den wahren Wert überdeckt oder nicht. In der folgenden Auswertung kommt dieser Aspekt aufgrund seiner relativ großen Unschärfe kaum zum Einsatz.

2. Daran schließt sich die Frage an, ob der jeweilige Interdezilbereich der Expertenantworten den entsprechenden wahren Wert überdeckt. Bei diesem schärferen Kriterium werden die extremen Antworten von insgesamt 20% der Experten nicht berücksichtigt. Denkbar ist, auch den Interquartilsbereich der Streuungen als Kriterium zu benutzen. Hierauf wird jedoch aufgrund der zu geringen Zahl beteiligter Experten weitgehend verzichtet.

3. Weiterhin kann die Breite der Interdezilbereiche pro Welle als ein Kriterium für die Antwortqualität benutzt werden. Eine sich von Welle zu Welle verringernde Streuung, die auf eine Vereinheitlichung der Expertenmeinungen hindeutet, gibt so ebenfalls Auskunft über die Qualität der Ergebnisse. Dieses Kriterium baut jedoch auf den vorangegangenen auf, d.h. es setzt voraus, daß die jeweilige Streuung der Expertenschätzungen den wahren Wert überdeckt.

4. Wesentlich für die Beurteilung der Ergebnisqualität einer Delphi-Studie ist die Annäherung des Gruppenergebnisses (wiedergegeben als Mittelwert) an den wahren Wert. Hierbei wird untersucht, ob sich nach den verschiedenen Befragungswellen das Schätzergebnis gegenüber der ersten Welle verbessert hat.

5. Ein weiteres wesentliches Kriterium besteht in der Veränderung (Erhöhung oder Verringerung) der mittleren Schätzfehler der einzelnen Experten für gleiche Aufgabentypen von Welle zu Welle.

Letztlich erlaubt erst die Betrachtung und Bewertung der Gesamtheit dieser verschiedenen Kriterien eine Beurteilung der erbrachten Schätzleistungen für die unterschiedlichen Aufgabentypen insgesamt.

Die nachfolgende Darstellung der Ergebnisse erfolgt analog zu der im Fragebogen gewählten Reihenfolge⁷.

3.1. Übersicht über die Ergebnisse der Schätzung vierstufiger Antwortverteilungen

Tabelle 1 gibt zunächst einen Überblick über die Ergebnisse der Schätzung vierstufiger Antwortverteilungen.

Zunächst muß festgestellt werden, daß in neun von 20 finalen Schätzungen der wahre⁸ Wert außerhalb des Interdezilbereichs der dritten Schätzung liegt.

Als Erklärung liegt hier die geringe Erfahrung der Studenten mit statistischen Verteilungen bzw. beim Umgang mit empirisch-sozialwissenschaftlichen Daten überhaupt nahe (vgl. Abschnitt 3.4.). Es zeigt sich, daß für die Bewältigung dieser Art von Aufgaben ein spezifisches Expertenwissen, d.h. konkret Erfahrungen beim Umgang mit statistischen Verteilungen, erforderlich ist. Dieses war bei den Testteilnehmern offenbar nicht ausreichend ausgebildet. In einem ähnlich angelegten Test waren hingegen empirisch-sozialwissenschaftlich erfahrene Experten wesentlich erfolgreicher bei der Absolvierung derartiger Aufgaben gewesen (vgl. Häder/ Häder 1994b).

Möglicherweise läßt sich das schlechte Schätzergebnis für Frage 5 außerdem auf die Asymmetrie der zu schätzenden Verteilung zurückführen. Diese Verteilung weist den größten Schiefekoeffizienten auf (0.62). Die vier anderen zu schätzenden Verteilungen sind entweder nahezu symmetrisch oder leicht linksschief. Die Stärke der Wölbung scheint hingegen keinen Einfluß auf die Güte der Schätzungen zu haben.

⁷ Muster der Erhebungsunterlagen sind im Anhang dieses Berichtes enthalten.

⁸ Unter „wahrer Wert“ wird im folgenden jeweils der in der Shell Jugendstudie 1992 ermittelte Wert verstanden.

Tabelle 1:
Mittelwerte (arithmetische Mittel - MW - in Prozent) und Streuungen (Interdezilbereiche - IDB - in Prozent) der Schätzungen aller Experten von Welle zu Welle bei den fünf Fragen mit vierstufigem Antwortmodell

	1.Welle		2.Welle		3.Welle		Shell
	MW	IDB	MW	IDB	MW	IDB	MW
1. Frage							
überh. nicht	22	5-50	23	10-40	19	10-40	13
weniger gut	35	22-50	33	25-40	33	20-40	35
gut	31	10-50	31	20-40	32	20-40	36
sehr gut	12	5-20	13	8-20	16	10-20	16
2. Frage							
überh. nicht	21	10-44	20	10-30	19	15-30	9
weniger gut	32	20-40	31	25-40	32	25-40	41
gut	32	15-50	33	25-40	32	25-40	38
sehr gut	15	9-25	16	10-21	17	10-20	12
3. Frage							
überh. nicht	9	2-15	10	10-20	11	5-15	2
weniger gut	20	10-35	21	10-30	23	10-30	21
gut	41	25-59	40	26-50	40	30-50	52
sehr gut	30	10-55	29	20-40	26	15-30	25
4. Frage							
überh. nicht	33	4-60	32	15-50	29	15-44	16
weniger gut	31	10-50	30	15-40	34	20-40	35
gut	20	8-35	23	10-35	24	15-35	37
sehr gut	16	5-50	15	6-20	13	7-20	12
5. Frage							
überh. nicht	22	5-50	22	10-35	21	15-30	32
weniger gut	34	15-50	32	15-50	33	25-40	48
gut	28	10-40	30	20-40	31	20-40	16
sehr gut	16	5-30	16	5-20	16	10-20	4

Tabelle 2 gibt die bei diesem Aufgabentyp aufgetretenen Schätzfehler für jeden einzelnen Testteilnehmer wider.

Tabelle 2:
Summe der Schätzfehler (Summe der χ^2 - Werte für die Fragen mit vierstufigem Antwortmuster) pro Experte und pro Welle für die vierstufigen Antwortschätzungen⁹

ID	1. Welle	2. Welle	3. Welle	Differenz 1.-3.Welle
1	162.8	66.5	156.1	6.7
2	206.5	204.4	135.5	71.0
3	311.9	201.1	155.5	156.4
4	260.9	249.2	174.9	86.0
5	130.2	177.3	191.6	-61.4
6	59.1	170.6	145.3	-86.2
7	570.0	172.9	120.4	449.6
8	449.8	216.3	185.2	264.6
9	127.0	128.2	124.3	2.7
10	378.2	204.2	233.6	144.6
11	466.1	339.0	144.9	321.2
12	781.6	443.9	293.8	487.8
13	294.0	183.4	158.7	135.3
14	774.5	275.9	180.1	594.4
15	125.1	189.6	227.0	-101.9
16	135.4	152.6	77.9	57.5
17	519.6	328.4	191.5	328.1
18	338.2	203.9	69.8	268.4
19	1397.5	316.9	552.4	-154.9
20	350.3	122.2	142.9	207.4
21	245.3	132.1	203.8	41.5
22	520.4	238.3	230.3	290.1
23	338.3	458.5	247.8	90.5
24	912.4	401.4	364.6	547.8
25	432.2	108.7	146.6	286.2
26	100.9	227.6	203.7	-102.8
27	299.6	1355.8	437.7	-138.1
28	197.6	110.7	132.3	65.3
29	421.5	198.1	179.1	242.4
30	399.1	250.5	317.0	82.1
31	186.3	190.0	123.6	62.7
32	216.1	104.5	307.3	-91.2
arithmetisches Mittel	378.4	253.8	204.8	173.6

Sieben negativen Schätzverläufen, bei denen sich in der dritten Welle der Schätzfehler gegenüber der ersten Welle vergrößert hat, stehen 25 positive individuelle Schätzerfolge gegenüber. Dieses Ergebnis ist besonders bemerkenswert. Es zeigt, daß sich - trotz des Fehlens spezieller Kenntnisse für die

⁹ Positive Werte verweisen auf eine Verringerung des Schätzfehlers.

Lösung der Aufgaben - die einzelnen Experten doch den wahren Werten annähern, was auf einen Erkenntnisprozeß während der Phase der Aufgabenbearbeitung schließen läßt. So ist es mehr als zwei Dritteln der Studenten gelungen, ihren Schätzfehler zu verringern. Dies gilt also selbst für Experten, die mit statistischen bzw. sozialwissenschaftlichen Daten wenig vertraut sind.

Beachtet werden sollte außerdem, daß sich die Breite der meisten Interdezilbereiche von der ersten zur dritten Welle deutlich verringert hat. Das deutet darauf hin, daß die Testteilnehmer mit einer extremen Einschätzung ihre Meinung im Laufe der Befragungswellen zur zentralen Lage hin verändert haben. Hinter diesem für Delphi-Studien bekannten Phänomen (vgl. z.B. BMFT 1993, Kerksieck 1972, Brockhoff 1979) verbirgt sich die Frage, ob die beobachtete Konvergenz in Richtung auf entweder den wahren Wert oder aber auf einen bestimmten zentralen Wert (z.B. Median) stattfindet. Hierzu sind weitere Untersuchungen erforderlich.

3.2. Übersicht über die Ergebnisse der Schätzung zweistufiger Antwortverteilungen (Alternativantworten)

Einen Überblick über die Ergebnisse der Schätzung zweistufiger Antwortverteilungen (Alternativantworten) gibt Tabelle 3.

Tabelle 3:
Mittelwerte (arithmetische Mittel - MW -in Prozent) und Streuungen (Interdezilbereiche - IDB - in Prozent) der Schätzungen der Experten von Welle zu Welle bei den vier Fragen mit zweistufigen Antwortvorgaben im Vergleich zu den Ergebnissen der Shell-Studie

	1.Welle		2.Welle		3.Welle		Shell
	MW	IDB	MW	IDB	MW	IDB	MW
6. Frage							
Antwort 1	45	25-65	43	30-60	42	35-50	45
Antwort 2	55	35-75	57	40-70	58	50-65	55
7. Frage							
Antwort 1	44	20-80	52	35-70	55	43-65	41
Antwort 2	56	20-80	48	30-65	45	35-57	59
8. Frage							
Antwort 1	63	45-80	64	55-70	65	60-72	63
Antwort 2	37	20-55	36	30-45	35	28-40	37
9. Frage							
Antwort 1	51	29-75	55	40-75	55	40-65	45
Antwort 2	49	25-71	45	25-60	45	35-60	55

Die tatsächlichen Werte der Shell-Jugendstudie liegen nur in einem Fall (Frage 7) nicht innerhalb des Interdezilbereichs der Schätzungen der dritten Welle. Damit liegt ein weiterer empirischer Beleg für die Möglichkeit vor, prinzipiell mit Hilfe der Delphi-Methode die Ergebnisse von Alternativfragen einer Bevölkerungsbefragung schätzen zu lassen¹⁰, auch wenn sich in diesem Experiment die arithmetischen Mittel der dritten Welle gegenüber denen der ersten Welle geringfügig von den wahren Werten entfernt haben.

Zu klären ist, weshalb Frage 7 von den Experten nicht korrekt geschätzt werden konnte. Möglicherweise spielt auch hier die Asymmetrie der Verteilungen eine Rolle: Frage 7 weist von allen Alternativfragen in den Originaldaten bei der im Test verwendeten Substichprobe die geringste Besetzung der Antwortmöglichkeit 1 auf.

In Tabelle 4 werden wiederum die Schätzerfolge der einzelnen Experten betrachtet. Der Vergleich zwischen der ersten und der dritten Schätzung (Differenz 1.-3. Welle, vgl. Spalte 4) zeigt, daß nur ein Testteilnehmer (ID 23) sich den wahren Ergebnissen *nicht* angenähert hat (negative Differenz in der vierten Spalte). Die Betrachtung der arithmetischen Mittel macht deutlich, daß im Durchschnitt die

¹⁰ In einem Experiment mit ähnlichem Design war es schon Mitarbeitern des ZUMA gelungen, Alternativschätzungen mit hoher Qualität abzugeben.

zweite Welle bessere Ergebnisse erbracht hat als die erste, und die dritte Welle wiederum zu besseren Schätzungen führte als die zweite. Die größte Annäherung an die wahren Werte erfolgte in der zweiten Runde. Dies gilt jedoch nicht zugleich auch für jeden einzelnen Experten.

Insgesamt sind auch diese Ergebnisse ähnlich denen aus einem vorangegangenen Experiment am ZUMA (vgl. Häder/Häder 1994b). Damit kann davon ausgegangen werden, daß es prinzipiell mit der Delphi-Technik gelingt, die Resultate von Alternativfragen aus Befragungen besser zu schätzen, als dies einzelne Experten bei einem einmaligen Versuch vermögen.

Tabelle 4:
Schätzfehler (Durchschnittliche Beträge der Abstände der Schätzungen zu den wahren Werten in den einzelnen Wellen) pro Experte für die vier Alternativfragen, Angaben in Prozentpunkten, sowie Darstellung der Veränderung der Schätzfehler zwischen der ersten und der dritten Welle¹¹

ID	1. Welle	2. Welle	3. Welle	Differenz 1.-3. Welle
1	21	10	9.25	11.75
2	27.75	16.5	12.75	15
3	11	5.5	5.5	5.5
4	9.25	6.75	6	3.25
5	18.5	7.25	4.25	14.25
6	16	19.5	14	2
7	18.25	19.25	7.75	10.5
8	12	8	10.25	1.75
9	13	4.75	5	8
10	10.5	8	8	2.5
11	25.25	26.5	19.5	5.75
12	13.5	9.5	9.5	4
13	14	8	9	5
14	17	9	7.25	9.75
15	14	16.5	11.75	2.25
16	16.75	6.75	7.75	9.5
17	12.75	3.75	3.25	9.5
18	11	13	6.5	4.5
19	21.5	3.25	6.5	15
20	18.25	8.25	12.75	5.5
21	14.75	9.75	13	1.75
22	11.5	14	9	2.5
23	10	12.5	13.25	-3.25
24	27	16.5	17.75	9.25
25	21	6.25	9.25	11.75
26	9	3	8	1
27	19.75	5.25	10.25	9.5
28	11.5	10.5	8	3.5
29	9.75	9	8	1.75
30	17.75	20.5	17.5	.25
31	16.5	4.5	6.75	9.75
32	15.25	7.75	9	6.25
arithmetisches Mittel:	15.78	10.3	9.57	6.21

¹¹ Positive Werte verweisen dabei auf eine Verringerung der Schätzfehlers.

3.3. Übersicht über die Ergebnisse der Schätzung von Mittelwerten

Tabelle 5 enthält die Übersicht über die Ergebnisse der Schätzung von Mittelwerten.

Tabelle 5:
Mittelwerte (arithmetische Mittel -MW-) und Streuungen (Interdezilbereiche -IDB-) von Welle zu Welle für die Gesamtheit der befragten Experten

Frage 10:	1. Welle		2. Welle		3. Welle		Shell MW
	MW	IDB	MW	IDB	MW	IDB	
Harmonie	5.7	4.5-6.4	5.8	5.0-6.2	5.8	5.0-6.2	6.2
Soz.Macht	3.5	2.0-5.0	3.4	2.4-4.7	3.5	2.5-4.2	2.6
Freiheit	6.1	5.0-6.7	6.2	5.8-6.6	6.2	5.8-6.6	6.3
Soz.Ordg.	4.8	3.2-6.0	4.5	3.5-5.4	4.6	4.0-5.5	5.4
Anregung	5.3	4.0-6.2	5.4	4.1-6.1	5.6	5.0-6.0	5.7
Höflichk.	3.8	2.5-5.0	3.7	2.8-4.5	3.5	3.0-4.0	5.0
Reichtum	4.5	3.3-6.0	4.7	4.0-5.5	4.8	4.2-5.3	4.4
Sicherh.	3.9	2.5-5.9	4.0	2.9-5.0	4.2	3.0-5.0	5.0
Kreativ.	4.7	3.1-6.0	4.8	4.0-5.7	4.9	4.0-6.0	5.3
Frieden	5.8	4.7-7.0	6.0	5.0-6.8	6.0	5.5-6.5	6.5
Tradition	3.0	2.0-4.5	2.9	2.0-3.7	2.7	2.0-3.2	3.7
Loslösung	2.8	1.8-4.0	2.7	2.0-3.1	2.6	2.0-3.2	3.1
Familie	5.3	4.0-6.5	5.4	4.2-6.3	5.7	5.0-6.3	6.1
Natur	4.6	3.5-6.0	4.9	3.9-6.0	4.9	4.0-5.7	5.6
Abwechsl.	5.2	3.7-6.2	5.5	4.8-6.0	5.5	5.0-6.1	5.7
Autorität	3.9	2.0-5.0	3.6	2.9-4.9	3.6	3.0-4.0	3.3
Freund.	5.9	5.0-6.8	6.2	5.9-6.5	6.3	6.0-6.5	6.4
Schönh.	3.7	2.0-5.0	3.5	2.5-4.5	3.6	3.0-4.2	5.1

Nur in drei Fällen liegt der wahre Wert außerhalb des Interdezilbereichs der Schätzungen der dritten Welle. Dies betrifft die Wichtigkeit von "Höflichkeit (gute Umgangsformen)", "Achtung vor Tradition (Erhaltung ehrwürdiger Sitten)" und "eine Welt in Schönheit (Schönheit der Natur und Künste)".¹²

Im folgenden soll deshalb geprüft werden, ob die drei Items, deren Schätzung nicht gelungen ist, durch eine besonders starke Streuung und/oder Asymmetrie in den Originaldaten charakterisiert sind. Dies könnte gegebenenfalls als Ursache für die aufgetretenen Probleme bei der Schätzung interpretiert werden.

Die einzelnen Items des Indikators zu den Wichtigkeitsurteilen wurden zur besseren Übersicht in Tabelle 6 nach ihrem Schätzerfolg (durchschnittlicher Betrag des Abstandes zwischen jeweils wahrem Wert und Mittelwert der Schätzung der Mittelwertfragen in der dritten Welle) angeordnet.

¹² Die Mittelwerte dieser Vorgaben konnten auch beim Test am ZUMA nicht adäquat ermittelt werden. Die Schätzungen wichen beim ersten Test hinsichtlich der Werte für Höflichkeit und Tradition sogar ebenfalls in besonders starkem Maße von den wahren bzw. zu schätzenden Werten ab.

Tabelle 6:

Standardabweichung und Schiefe der Originaldaten der Shell Jugendstudie 1992 sowie absoluter Schätzfehler nach der dritten Welle (Indikatoren zu den Wichtigkeitsurteilen)

Item	Standardabweichg.	Schiefe	Fehler
Freiheit	1.158	-2.253	0.1
Anregung	1.392	-1.312	0.1
Freundschaft	1.079	-2.332	0.1
Abwechslung	1.429	-1.307	0.2
Autorität	1.859	0.318	0.3
Harmonie	1.317	-1.969	0.4
Reichtum	1.808	-0.387	0.4
Kreativität	1.661	-0.899	0.4
Familie	1.344	-1.698	0.4
Frieden	1.079	-2.934	0.5
Loslösung	1.875	0.637	0.5
Natur	1.528	-1.152	0.7
Sicherheit	1.959	-0.695	0.8
Soz. Ordnung	1.563	-0.998	0.8
Soz. Macht	1.742	.906	0.9
Tradition	1.926	0.136	1.0
Höflichkeit	1.694	-0.665	1.5
Schönheit	1.735	-0.761	1.5

Diese Übersicht vermittelt zunächst den Eindruck, als würde mit steigender Standardabweichung und mit sinkendem Betrag der Schiefe in den Originaldaten der absolute Schätzfehler größer. Tatsächlich ist der lineare Korrelationskoeffizient zwischen den Variablen Standardabweichung und Fehler positiv ($R=0.52$) und mit $p<0.05$ signifikant. Zwischen dem Betrag der Schiefe und dem Fehler besteht dagegen ein gegenläufiger Zusammenhang ($R= -0.46$), der jedoch nicht signifikant ist. Damit ist allerdings noch nicht hinreichend erklärt, warum gerade die Items Tradition, Höflichkeit und Schönheit in beiden Tests nicht erfolgreich geschätzt werden konnten. Die Standardabweichungen um die Mittelwerte der Originaldaten sind zwar jeweils relativ hoch, nehmen aber keine Werte an, die die Fehlleistungen allein erklären würden.

Als eine mögliche ergänzende Erklärung (ohne sie an dieser Stelle vertiefend darstellen zu können) für die in diesem Experiment wie auch schon in einem Test bei ZUMA übereinstimmend mißlungenen Schätzungen bietet sich an, daß die Originaldaten der Shell-Studie aufgrund des Phänomens der social desirability eine systematische Verzerrung aufweisen, d.h. die befragten Jugendlichen für die Werte "Höflichkeit", "Achtung der Tradition" sowie "eine Welt in Schönheit" eine höhere Bedeutung angegeben haben als sie tatsächlich von ihnen verinnerlicht ist. Würde man einer solchen Argumentation folgen, bedeutete dies, daß die Delphi-Methode sogar zu valideren Ergebnisse führt als die eigentliche Befragung, da dieser Effekt bei Delphi weitgehend ausgeschaltet sein dürfte.

Insgesamt hat dieser Test bei den Mittelwertschätzungen jedoch gute Ergebnisse erbracht. Drei "Problemfällen" stehen 15 gelungene Schätzungen gegenüber.

Auch die Darstellung der Schätzergebnisse der einzelnen Experten (vgl. Tabelle 7) unterstreicht diese insgesamt positive Einschätzung. Nur vier Testteilnehmer (ID 7; 18; 27 und 32) haben sich von der ersten zur dritten Welle von den wahren Werten entfernt.

Im Mittel haben sich die Schätzfehler aber von Welle zu Welle verringert, wobei sich die Schätzleistungen wiederum insbesondere in der zweiten gegenüber der ersten Welle verbessert haben. Dieses Ergebnis korrespondiert mit denen anderer Delphi-Experimente (vgl. z.B. BMFT 1993; Brockhoff 1979).

Tabelle 7:
Mittlere absolute Schätzfehler pro Experte und pro Welle für die Mittelwertschätzungen

ID	Welle 1	Welle 2	Welle 3	Diff. ¹³
1	1.14	0.71	0.64	0.50
2	1.64	1.05	0.83	0.81
3	1.14	1.04	0.62	0.52
4	1.37	0.83	0.81	0.56
5	0.89	0.65	0.81	0.08
6	1.69	1.13	1.31	0.38
7	0.76	1.01	0.89	-0.13
8	1.14	1.00	1.00	0.14
9	0.67	0.44	0.44	0.23
10	1.31	0.86	0.60	0.71
11	1.00	0.87	0.98	0.02
12	0.74	0.51	0.54	0.20
13	0.69	0.72	0.50	0.19
14	1.06	0.68	0.48	0.58
15	0.84	0.97	0.64	0.20
16	0.99	0.82	0.76	0.23
17	0.62	0.70	0.54	0.08
18	0.41	0.58	0.57	-0.16
19	0.93	0.81	0.60	0.33
20	1.32	0.57	0.67	0.65
21	1.02	0.69	0.82	0.20
22	1.01	1.01	0.91	0.10
23	1.51	1.14	1.02	0.49
24	1.71	1.13	0.86	0.85
25	0.60	0.48	0.49	0.11
26	1.19	0.89	1.01	0.18
27	0.73	0.80	0.82	-0.09
28	0.93	0.43	0.52	0.41
29	0.88	0.64	0.64	0.24
30	1.18	0.98	0.89	0.29
31	0.64	0.53	0.47	0.17
32	0.58	0.57	0.70	-0.12
arithmetisches Mittel:	1.01	0.79	0.73	0.28

¹³ Positive Werte verweisen auf eine Verringerung des Schätzfehlers.

In einem ähnlich angelegten empirischen Versuch zur Beurteilung der Delphi-Methode wurde eine Punktschätzung als gut bezeichnet, falls der Mittelwert nach der dritten Welle nicht mehr als 0.2 von dem wahren Wert entfernt lag (vgl. Häder/Häder 1994b). Möglicherweise sind die Experten jedoch nicht in der Lage, den Lageparameter auf eine Nachkommastelle genau zu schätzen. So ist in diesem Test für alle 18 Items zu beobachten, daß die meisten Testteilnehmer ganzzahlige Werte oder .5-Werte bei der Beurteilung bevorzugen. Eine Konsequenz ist, daß die Parameterschätzungen in den meisten Fällen etwas verzerrt sein werden. So bedeutet z.B. die Zahl 6 bei einer Reihe von Experten nicht „exakt 6“, sondern „ungefähr 6“. Ebenso heißt 6.5 bei diesen Personen offenbar „zwischen 6 und 7“. Dieses Phänomen soll anhand einer Häufigkeitsverteilung der 3. Welle für das Item „anregendes Leben“ demonstriert werden, für dessen Punktschätzung immerhin ca. zwei Drittel der Testteilnehmer ganzzahlige bzw. .5-Werte wählten.

Tabelle 8:
Häufigkeitsverteilung der Schätzungen für das Item Anregendes Leben der 3. Welle

Wert	Häufigkeit	gültige Prozent	Sum. Prozent
4.5	1	3.1	3.1
5.0	6	18.8	21.9
5.1	1	3.1	25.0
5.3	1	3.1	28.1
5.5	6	18.8	46.9
5.6	4	12.5	59.4
5.7	3	9.4	68.8
6.0	9	28.1	96.9
6.5	1	3.1	100.0

3.4. Ergebnisse der Zusatzbefragung

Aus der Darstellung der Ergebnisse der Zusatzbefragungen (vierte Welle) lassen sich Motivationsstärke, Verständnis der Aufgabenstellung und verwendete Orientierungshilfen der Expertengruppe näher charakterisieren. Diese Beschreibung kann genutzt werden, um weitere Hinweise auf Gründe für die unterschiedliche Qualität der Schätzleistungen bei den einzelnen Aufgabentypen zu erlangen.

- Zur *Verständlichkeit* des Anliegens der Befragung:

Generell verlangt die relativ ungewöhnliche Aufgabe, nicht über eigene Meinungen zu berichten, sondern über die vermuteten Ergebnisse einer repräsentativen Befragung Auskunft zu geben, nach einer klaren Erläuterung. Daß es nicht in vollem Umfang gelungen ist, das Anliegen des Tests deutlich zu machen und Verständnis dafür zu erzeugen, belegen die folgenden Antworten der Zusatzbefragung:

Auf die Frage, wie verständlich ihnen das Anliegen des Tests war, gaben nur 14 Studenten an, daß es ihnen eher verständlich gewesen sei, 13 Studenten war es dagegen nur teilweise verständlich und 5 Testteilnehmer haben das Anliegen des Tests eher nicht verstanden. Insgesamt ist damit der Anteil derjenigen, die die Testaufgaben vollständig verstanden haben, zu gering, weil dieser Aspekt entscheidenden Einfluß auf das Gelingen des Experimentes überhaupt hat.

Die Studenten wurden weiterhin danach gefragt, inwieweit sie sich bei ihren Schätzungen davon haben leiten lassen, wie sie selbst wohl die einzelnen Fragen beantwortet hätten. Dabei handelte es sich um eine *Kontrollfrage*, da zuvor die Anweisung gegeben worden war, sich bei den Schätzungen *nicht* an der eigenen Meinung zu orientieren. Das Ergebnis zeigt, daß sich immerhin sieben Studenten eher stark (Antwortstufe 4) an der eigenen Meinung orientiert haben, während nur drei Testteilnehmer angaben, bei der Absolvierung der Testaufgaben vollkommen von der eigenen Meinung abstrahiert zu haben. Dies kann als ein ergänzender Hinweis auf Probleme hinsichtlich der Verständlichkeit des Anliegens unseres Experimentes bei den Studenten bewertet werden.

Schließlich dokumentiert die Frage nach der vermuteten *Kompetenz* der für den Test rekrutierten Experten (also der Kommilitonen) ein nur in mittlerem Maße (Mittelwert 2.8 bei 5stufiger Skala) ausgeprägtes Vertrauen in das Fachwissen der anderen Testteilnehmer. Dies würde dafür sprechen, daß kein ausgeprägter Konformitätsdruck aufgrund der Rückmeldungen entstanden sein dürfte, sondern die Studenten sich weitgehend auf die eigenen Schätzungen verlassen haben. Dies bedeutet allerdings zugleich, daß den rückgemeldeten Informationen weniger Autorität zugeschrieben wurde.

- Als *Orientierungen*, die bei den Schätzungen der ersten Welle benutzt wurden, wurden am stärksten die Erfahrungen mit anderen Jugendlichen (MW: 1.9) sowie die Leitung spontan durch das Gefühl (MW: 2.1) berichtet. Eine wesentlich

geringere Rolle spielten dagegen Erfahrungen mit ähnlichen Fragestellungen (MW: 3.5) sowie mit statistischen Verteilungsformen (MW: 3.8). Dies unterstreicht die nur eingeschränkte Kompetenz der Studenten als Experten zur Lösung derartiger Aufgabenstellungen insbesondere im Vergleich zum vorangegangenen Test mit Mitarbeitern des ZUMA.

- Im Rahmen des Tests wurde erstmals zu ermitteln versucht, ob und in wieweit etwa bestehende *Stereotype* über Jugendliche für die Beantwortung der Fragen eine Rolle gespielt haben. Operationalisiert lautete die Fragestellung in der Zusatzbefragung: "Ich habe mich bei meinen Schätzungen leiten lassen, davon, was man so über Jugendliche liest und hört." Es stellte sich heraus, daß damit tatsächlich eine weitere wichtige Orientierung der Studenten bei ihren Antworten ermittelt worden ist. Lediglich spontane Antworten nach dem Gefühl und die Erfahrungen mit anderen Jugendlichen wurden noch häufiger als Orientierung bei den Schätzungen der ersten Welle angegeben als Stereotype.

- Der für das Ausfüllen jeweils benötigten *Zeitfonds*, wie er von der Expertengruppe in Marburg berichtet wurde, verringerte sich jeweils von Welle zu Welle. Von den Studenten wurden folgende Durchschnittswerte angegeben: 1. Welle: 15 Minuten, 2. Welle: 13 Minuten, 3. Welle: 10 Minuten. Da der Test mit den Marburger Studenten unter Klausurbedingungen stattgefunden hatte, war der Versuchsleiter in der Lage, diese von den Experten angegebenen Zeitspannen zu beurteilen. Demnach wurden sie von den Teilnehmern des Tests deutlich unterschätzt. Realistisch sei es dagegen, jeweils die doppelte Zeit als für die Bearbeitung der Schätzaufgaben benötigt anzusetzen.

Im Vergleich zum vorangegangenen Test bei ZUMA, der ein ähnliches Design und ähnliche Schätzaufgaben zum Inhalt hatte, kann anhand dieser Ergebnisse von einer weitaus größeren methodischen Kompetenz und dem wesentlich besseren Verständnis der Aufgabenstellung bei den ZUMA-Experten ausgegangen werden. Dem steht allerdings gegenüber, daß die Studenten selbst Angehörige der Population sind, deren Befragungsergebnisse geschätzt werden sollten. Ihnen stand als Orientierung für ihre Schätzungen deshalb vor allem die Kenntnis von Denkinhalten Gleichaltriger zur Verfügung.

Insgesamt muß jedoch festgestellt werden, daß bei dem hier vorgestellten Test hinsichtlich der Kompetenz nicht von einem optimalen Expertengremium gesprochen werden kann, was sich in der Qualität einiger Schätzungen niedergeschlagen hat. Für die Beobachtung typischer, bei Delphi-Studien ablaufender Prozesse scheint die Testanordnung mit den rekrutierten Experten jedoch hinreichend geeignet.

3.6. Zusammenfassung

Insgesamt, so kann ein weiteres Mal festgestellt werden, gelingt es prinzipiell, mit Hilfe der Delphi-Methode die Ergebnisse von Bevölkerungsbefragungen zu schätzen. Vor allem die folgenden Erkenntnisse aus diesem Test gilt es hervorzuheben:

Obwohl sich die am Test beteiligten Studenten weder durch besondere Erfahrungen mit empirisch-sozialwissenschaftlichen Untersuchungen noch eine tiefgehende Kenntnis statistischer Verteilungen auszeichnen, ist es fast allen gelungen, die Qualität der Schätzungen von Welle zu Welle zu verbessern. Dies läßt sich insbesondere daran zeigen, daß für die Gesamtheit der Testteilnehmer bei allen drei Aufgabentypen von der ersten über die zweite zur dritten Welle eine Annäherung an die „wahren Werte“ stattgefunden hat.

Dies deutet darauf hin, daß die eingangs erwähnten Denkprozesse, bei denen im Laufe der Aufgabenbearbeitung immer mehr Kontextwissen für die Problemlösung einbezogen wird, was sich in der Konstruktion immer hochwertigerer mentaler Modelle niederschlägt, tatsächlich zu einer höheren Qualität der finalen Schätzung führen, als dies bei einer nur einmaligen Urteilsabgabe zu erwarten ist. Dafür spricht z.B., daß unter den vorgegebenen möglichen Orientierungshilfen für die Schätzungen in der zweiten und dritten Welle am häufigsten angegeben wurde, nochmals über den Inhalt der Aufgaben nachgedacht zu haben.

Damit dürfte es erwiesen sein, daß Delphi zu einem differenzierten Erkenntnisprozeß während der Aufgabenbearbeitung (hervorgerufen durch wiederholte mentale Auseinandersetzung mit der inhaltlichen Aufgabenstellung sowie den Rückmeldungen) führt, der positiven Einfluß auf die Qualität der Ergebnisse hat.

Wie gut allerdings die Schätzungen der letzten Welle *absolut* sind, hängt wesentlich von der generellen Eignung der Experten, das heißt ihrer Sachkompetenz schon zu Beginn der Delphi-Studie ab. In dieser Beziehung - und dies war bei der Auswahl von Medizinstudenten des ersten Studienjahres als Testteilnehmer auch nicht anders zu erwarten - war das Expertengremium nicht optimal zusammengesetzt. Das Ziel des Tests - nachzuweisen, daß eine wiederholte Urteilsabgabe bei unsicherem Wissen zu einer höheren Qualität des Urteils führt - konnte jedoch mit den Studenten eindeutig erreicht werden.

Ein weiteres interessantes Testergebnis besteht darin, daß die verschiedenen Aufgabentypen mit unterschiedlichem Erfolg absolviert worden sind. So sind die finalen Punktschätzungen zu den Wichtigkeitsurteilen insgesamt von höherer Qualität als die Verteilungsschätzungen zu Fragen des Zeitbewußtseins. Der Frage, inwiefern hier eine Regelmäßigkeit vermutet werden kann, derzufolge es möglicherweise einfacher ist, Punktschätzungen abzugeben, kann allerdings aufgrund dieser Testanordnung nicht nachgegangen werden. Weitere Anhaltspunkte hierzu soll aber der im folgenden vorzustellende Aufgabensplitt liefern.

4. Punkt- vs. Verteilungsschätzungen: Ein methodischer Test

Ein Ziel dieser Erhebung war es, ein bei einer vorangegangenen Testbefragung aufgetauchtes Problem genauer zu untersuchen. Dazu wurden, wie in Abschnitt 2 erläutert, die Experten in zwei Gruppen (jeweils N=16) eingeteilt. Ein Teil der Testteilnehmer, im folgenden Gruppe 1 genannt, hatte die Randverteilungen zweier fünfstufiger Antwortskalen zu schätzen, der andere Teil der Testteilnehmer ausschließlich deren Mittelwerte. Alle anderen Fragen wurden für beide Gruppen gleich gestellt.¹⁴ Mit dieser Testanordnung sollte untersucht werden, ob es besser gelingt, Punkt- oder Verteilungsschätzungen vorzunehmen.

¹⁴ Der Leser möge dabei beachten, daß die Rückmeldung an die Experten auch nicht für die Gruppen getrennt durchgeführt wurde. Die rückgemeldeten Werte unterschieden sich nur für die beiden Splittfragen.

4.1. Kognitionspsychologischer Rahmen des Tests

Die Theorie mentaler Modelle wurde dazu entwickelt, um kognitive Prozesse zu erklären, die bei unter einer gewissen Unsicherheit getroffenen Entscheidungen ablaufen. Im Abschnitt 1.3.3. wurden dazu bereits Experimente zitiert, die gezeigt haben, daß sich mit Hilfe lokaler mentaler Modelle gewisse Fehlleistungen, z.B. der overconfidence bias, erklären lassen. Probabilistische mentale Modelle haben dagegen bei Häufigkeitsschätzungen zu erstaunlich korrekten Entscheidungen geführt. In diesen Zusammenhang soll zunächst auch der Splitt zur Punkt- bzw. Verteilungsschätzung eingeordnet werden.

Das von Gigerenzer et al. (1991) geschilderte Experiment beinhaltete die Lösung der gleichen Aufgabe (die Schätzung, ob ein zuvor abgegebenes Urteil richtig sei) auf unterschiedliche Weise (als Einzelfallbeurteilung bzw. als Häufigkeitsschätzung für eine Reihe von Urteilen). Ähnlich ist auch die im hier zu schildernden Splitt gestellte Aufgabe angelegt: Für das gleiche Item sollte einmal das Ergebnis als Mittelwert geschätzt werden und im anderen Fall sollte versucht werden, die Anwohnerhäufigkeit jeder einzelnen Antwortstufe anzugeben.

Da es sich bei den im Rahmen der Delphi-Befragung getroffenen Entscheidungen um Schätzungen handelt, die in der Regel ohne vollständiges Wissen getroffen werden müssen, ist es naheliegend, Grundgedanken der Theorie mentaler Modelle prinzipiell auch auf Entscheidungen bei Delphi-Studien zu übertragen.

Einleitend sollen kurz einige dafür relevante Gedanken wiedergegeben werden: Im Rahmen der Theorie der PMM spielen die Bildung von Referenzklassen (vgl. Brunswik 1943: 257), von sogenannten „probability cues“ sowie die eigentliche Aufgabenstellung, die „target variable“ (Gigerenzer et al. 1991: 507ff.), die entscheidende Rolle. Eine Referenzklasse, wie sie in dem von Gigerenzer et al. (1991) geschilderten Experiment (vgl. Abschnitt 1.3.3.) gebildet werden könnte, würde alle deutschen Städte umfassen, zu denen auch Heidelberg und Solingen gehören. Als probability cues können nun solche Annahmen dienen, die - resultierend aus der Aufgabenstellung, die größere der beiden Städte herauszufinden - Hinweise auf die jeweilige Einwohnerzahl erlauben. So kann angenommen werden, daß:

- von größeren Städten eher eine Mannschaft in der Bundesliga spielt,
- größere Städte kürzere Kraftfahrzeugkennziffern besitzen,

- wenn es sich bei einer der beiden Städte um die Hauptstadt eines Bundeslandes handelt, diese dann die größere ist usw.

In der Theorie der PMM wird weiterhin darauf verwiesen, daß Probability cues die Eigenschaft haben, (lediglich) mit einer bestimmten Wahrscheinlichkeit auf die target variable, den interessierenden Sachverhalt, hinzudeuten.¹⁵

Um die Vermutung zu unterstützen, daß zur Lösung der Aufgaben innerhalb dieses Tests PMM benutzt worden sind, müßten sich zunächst solche Referenzklassen benennen lassen. Derartige probability cues könnten z.B. sein:

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

Bei Delphi-Befragungen vermögen die Experten nun die jeweiligen Wahrscheinlichkeiten, mit denen die einzelnen probability cues auf den erfragten

¹⁵ Gigerenzer et al. (1991: 509) stellten beispielsweise fest, daß die Benutzung des Fußballmannschafts- probability cues in der Bundesligaspielsaison 1988/89 mit einer Wahrscheinlichkeit von 91% zu einer richtigen Antwort bei der Frage nach der größeren Einwohnerzahl einer westdeutschen Stadt führen würde.

Sachverhalt hindeuten, (besonders) fachmännisch einzuschätzen.¹⁶ Die zur Lösung der Aufgaben herangezogenen probability cues dürften in beiden Gruppen identisch sein, da formal jeweils nach den gleichen Sachverhalten gefragt wurde (Einstellung zur deutschen Wiedervereinigung sowie allgemeine Lebensziele von Jugendlichen). Wenn nun Unterschiede bei der Aufgabenbewältigung auftreten, so dürften diese auf die Art der gestellten Aufgabe und auf den damit jeweils verbundenen unterschiedlichen Schwierigkeitsgrad („target variable“) zurückzuführen sein. Der Test wird zeigen, ob es möglich ist, für beide Aufgabentypen geeignete PMM zu bilden, oder ob nur eine Aufgabenart lösbar ist. Die daraus für die Gestaltung von Delphi-Befragungen resultierenden Implikationen liegen nahe.

4.2. Zur Homogenität der Splitt-Gruppen

Um die Ergebnisse der beiden Splittfragen miteinander vergleichen zu können, ist es zunächst aus methodischer Sicht notwendig, daß die beiden Gruppen hinsichtlich Motivation und Expertenwissen untereinander homogen sind.

Gegenstand dieses Unterabschnitts ist es daher, die Homogenität der beiden Subpopulationen statistisch zu belegen. Dies erfolgt einerseits anhand der Ergebnisse der verschiedenen, für beide Gruppen in gleicher Weise gestellten Testaufgaben, also mit Ausnahme des Splitts (denn dessen methodische Legitimation soll gerade nachgewiesen werden), andererseits aufgrund der Antworten der Zusatzbefragung in beiden Gruppen.

- Zur Homogenität des Antwortverhaltens bei den Testaufgaben in beiden Splittgruppen

Tabellen 9 und 10 enthalten zunächst einen Überblick über die Ergebnisse der Schätzungen der vierstufigen Antwortverteilungen nach Gruppen getrennt.

¹⁶ Gigerenzer et al. (1991: 510) stellen eine umfangreiche Literatur (vgl. Gigerenzer 1984, Hasher/Goldstein/Toppino 1977, Howell/Burnett 1978, Zacks/Hasher/Sanft 1982) zusammenfassend fest: „A large literature exists that suggests that (a) memory is often (...) excellent in storing frequency information from various environments and (b) the registering of event occurrences for frequency judgments is a fairly automatic cognitive process requiring very little attention or conscious effort“ (Gigerenzer et al. 1991: 510).

Auffällig ist, daß die Interdezilbereiche beider Gruppen für die dritte Welle von Frage 1 die entsprechenden Werte der Shell-Studie überdecken. Auf der anderen Seite überdeckt kein Interdezilbereich (3. Welle) der beiden Gruppen für Frage 5 der Verteilungsschätzungen die Werte der Shell-Studie. Da nur 11 (Gruppe 1) bzw. 10 (Gruppe 2) der 20 Interdezilbereiche die Werte der Shell-Studie überdecken, müssen die Verteilungsschätzungen - auch für die Gruppen getrennt - als insgesamt nicht zufriedenstellend angesehen werden.

Die Unterschiede in den Schätzungen zwischen den Gruppen können vernachlässigt werden: Die maximale Abweichung zwischen den Mittelwerten der Schätzungen der dritten Welle beträgt 7 Prozentpunkte.

Tabelle 9:
Lage- (arithmetisches Mittel - MW) und Streuungsmaß (Interdezilbereich - IDB) der Schätzungen der Experten der Gruppe 1 für die verschiedenen Wellen bei den fünf Fragen mit vierstufigem Antwortmodell im Vergleich zu den Ergebnissen der Shell-Studie (Angaben in Prozent)

	1. Welle		2. Welle		3. Welle		Shell
	MW	IDB	MW	IDB	MW	IDB	MW
1. Frage							
überh. nicht	23	5-50	24	8-40	20	10-40	13
weniger gut	34	22-50	31	25-40	30	15-40	35
gut	30	10-50	32	20-50	35	20-45	36
sehr gut	13	5-20	13	8-20	15	10-20	16
2. Frage							
überh. nicht	21	10-50	20	7-30	18	15-20	9
weniger gut	31	20-40	31	25-37	31	25-40	41
gut	31	10-45	33	24-50	34	27-40	38
sehr gut	17	10-25	16	10-22	17	10-20	12
3. Frage							
überh. nicht	11	1-29	12	5-21	12	5-20	2
weniger gut	19	10-32	22	10-30	25	10-33	21
gut	41	14-60	39	26-50	37	30-40	52
sehr gut	30	10-60	27	18-37	26	15-30	25
4. Frage							
überh. nicht	34	15-60	32	15-46	32	20-44	16
weniger gut	35	12-55	33	20-40	34	20-45	35
gut	21	7-35	23	10-38	22	15-30	37
sehr gut	10	3-17	12	6-20	12	7-20	12
5. Frage							
überh. nicht	23	5-50	21	10-34	22	10-30	32
weniger gut	36	15-60	31	15-48	33	20-40	48
gut	26	10-50	31	20-40	29	20-42	16
sehr gut	15	5-30	17	10-27	16	10-20	4

Tabelle 10:
Lage- (arithmetisches Mittel - MW) und Streuungsmaß (Interdezilbereich - IDB) der Schätzungen der Experten der Gruppe 2 für die verschiedenen Wellen bei den fünf Fragen mit vierstufigem Antwortmodell im Vergleich zu den Ergebnissen der Shell-Studie (Angaben in Prozent)

	1. Welle		2. Welle		3. Welle		Shell MW
	MW	IDB	MW	IDB	MW	IDB	
1. Frage							
überh. nicht	20	5-50	22	10-40	20	10-40	13
weniger gut	35	20-50	35	25-45	35	28-45	35
gut	33	15-50	29	20-40	28	20-40	36
sehr gut	12	5-20	14	5-25	17	10-30	16
2. Frage							
überh. nicht	22	7-44	20	10-30	20	10-30	9
weniger gut	33	10-40	31	25-40	32	20-40	41
gut	32	15-56	34	25-40	30	20-35	38
sehr gut	13	8-20	15	10-20	18	10-25	12
3. Frage							
überh. nicht	7	2-10	8	5-15	10	5-15	2
weniger gut	22	10-36	21	10-30	21	10-30	21
gut	41	30-50	40	25-50	43	40-50	52
sehr gut	30	10-55	31	20-50	26	15-35	25
4. Frage							
überh. nicht	33	0-60	32	10-50	25	10-45	16
weniger gut	27	8-45	28	14-40	34	20-40	35
gut	20	8-30	23	10-35	27	15-40	37
sehr gut	21	5-70	17	5-30	14	5-30	12
5. Frage							
überh. nicht	21	10-50	23	10-40	20	15-30	32
weniger gut	31	15-40	32	15-50	32	25-45	48
gut	30	20-40	30	20-40	32	20-40	16
sehr gut	18	10-35	15	5-20	16	10-20	4

Auch bei den anderen Schätzaufgaben, die beide Gruppen zu absolvieren hatten, zeigt sich ein ähnliches Bild.

Betrachtet man die Schätzerfolge der beiden Gruppen für die vier Fragen mit den Antwortalternativen¹⁷, so unterscheiden sich diese nicht von dem Gesamtschätzergebnis: Der Interdezilbereich von Frage 7 überdeckt in beiden Gruppen den Wert der Shell-Studie nicht, während die Interdezilbereiche der anderen drei Fragen die aus der Shell-Studie erhaltenen Werte überdecken.

Die Frage, ob sich die beiden Gruppen voneinander unterscheiden, soll für die Fragen 6 bis einschließlich 9 durch den Kolmogorov-Smirnov-Zweistichproben-Test untersucht werden. Die Werte der K-S-Statistik sind in Tabelle 11, nach Wellen getrennt, dargestellt.

¹⁷ Auf die tabellarische Darstellung der Einzelergebnisse der beiden Gruppen bei den Alternativschätzungen soll an dieser Stelle verzichtet werden.

Tabelle 11: <i>Werte der Kolmogorov-Smirnov-Statistik nach Wellen getrennt</i>			
	1. Welle	2. Welle	3. Welle
Frage 6	.37500	.25000	.25000
Frage 7	.25000	.31250	.31250
Frage 8	.31250	.12500	.25000
Frage 9	.12500	.31250	.12500

Die kritischen Werte dieses Tests betragen bei den Stichprobenumfängen $N_1=16$ und $N_2=16$: 0.4375 (Signifikanzniveau 5%) und 0.5625 (Signifikanzniveau 0.01%). In keinem der Fälle wird die kritische Grenze überschritten. Man kann daher davon ausgehen, daß die Schätzungen für die Antwortalternativen einer (homogenen) Population entstammen (Signifikanzniveau jeweils 5%).

Genauer betrachtet werden nun die Mittelwertschätzungen der beiden Gruppen. In den Tabellen 12 und 13 sind die entsprechenden Ergebnisse dargestellt.

Tabelle 12: <i>Lage- (arithmetisches Mittel - MW) und Streuungsmaß (Interdezilbereich - IDB) der Schätzungen der Experten der Gruppe 1 für die verschiedenen Wellen für die Mittelwertschätzungen im Vergleich zu den Ergebnissen der Shell-Studie (Angaben in Prozent)</i>							
	1. Welle		2. Welle		3. Welle		Shell MW
	MW	IDB	MW	IDB	MW	IDB	
Harmonie	6.0	5.3-6.5	6.0	5.3-7.0	5.9	5.1-6.2	6.2
Soz. Macht	3.1	1.5-4.8	3.3	2.2-4.7	3.3	2.8-4.2	2.6
Freiheit	6.2	5.5-6.8	6.4	6.0-7.0	6.3	5.8-7.0	6.4
Soz. Ordnung	5.3	3.7-6.3	4.7	3.8-5.5	4.5	3.5-5.0	5.4
Anregung	5.7	4.8-6.5	5.3	4.1-6.0	5.6	5.0-6.0	5.7
Höflichkeit	3.7	2.1-5.0	3.6	2.5-4.5	3.4	3.0-4.0	5.1
Reichtum	4.8	3.3-6.9	4.6	3.5-5.5	4.7	3.5-5.3	4.4
Sicherheit	3.9	2.0-6.0	3.9	2.8-5.5	3.9	3.0-5.0	5.0
Kreativität	5.1	4.0-6.5	4.9	3.8-6.0	5.0	4.2-5.9	5.3
Frieden	6.0	4.7-7.0	6.1	5.3-6.8	6.1	5.6-6.5	6.5
Tradition	2.8	2.0-4.0	2.9	2.0-3.7	2.6	2.2-3.2	3.7
Loslösung	3.0	1.8-4.0	2.8	2.0-3.5	2.7	2.0-3.3	3.1
Familie	5.2	3.6-6.5	5.5	4.0-6.3	5.6	5.0-6.1	6.1
Natur	5.0	3.7-6.0	4.8	3.9-5.8	4.9	4.0-5.7	5.6
Abwechslung	5.4	3.8-7.0	5.3	4.8-6.0	5.7	5.4-6.1	5.7
Autorität	4.0	1.9-6.0	3.5	2.6-4.8	3.4	2.4-4.0	3.3
Freundschaft	6.1	5.0-7.0	6.2	6.0-6.8	6.3	6.0-6.5	6.4
Schönheit	4.3	3.2-6.0	3.8	2.6-5.0	3.8	3.1-4.6	5.1

Tabelle 13:
Lage- (arithmetisches Mittel - MW) und Streuungsmaß (Interdezilbereich - IDB) der Schätzungen der Experten der Gruppe 2 für die verschiedenen Wellen für die Mittelwertschätzungen im Vergleich zu den Ergebnissen der Shell-Studie (Angaben in Prozent)

	1. Welle		2. Welle		3. Welle		Shell MW
	MW	IDB	MW	IDB	MW	IDB	
Harmonie	5.3	4.0-6.0	5.6	5.0-6.0	5.7	4.5-6.5	6.2
Soz. Macht	4.0	2.5-6.0	3.5	2.5-5.0	3.6	2.5-5.0	2.6
Freiheit	5.9	5.0-6.7	6.0	5.0-6.5	6.1	5.0-6.5	6.4
Soz. Ordnung	4.3	3.0-5.8	4.4	3.5-5.0	4.7	4.0-5.5	5.4
Anregung	5.0	3.5-6.0	5.5	4.0-6.4	5.6	5.0-6.0	5.7
Höflichkeit	3.8	3.0-6.0	3.7	3.0-5.0	3.6	3.0-4.5	5.1
Reichtum	4.2	2.5-5.0	4.9	4.0-5.6	4.9	4.5-5.5	4.4
Sicherheit	4.0	2.5-5.6	4.1	3.0-5.0	4.4	3.0-5.7	5.0
Kreativität	4.3	3.0-6.0	4.7	4.0-5.5	4.7	3.5-6.0	5.3
Frieden	5.7	4.5-7.0	6.0	4.9-7.0	6.0	5.0-6.5	6.5
Tradition	3.1	2.0-5.0	2.9	2.0-4.0	2.8	2.0-3.8	3.7
Loslösung	2.7	2.0-3.3	2.5	2.0-3.1	2.6	2.0-3.0	3.1
Familie	5.5	4.8-6.3	5.4	4.2-6.5	5.7	5.4-6.0	6.1
Natur	4.2	3.5-4.7	4.9	3.5-6.0	4.8	4.3-5.5	5.6
Abwechslung	4.9	4.0-6.0	5.6	5.0-6.1	5.4	5.0-5.9	5.7
Autorität	3.9	3.3-4.9	3.8	3.0-5.4	3.8	3.0-4.0	3.3
Freundschaft	5.7	5.3-6.0	6.1	5.5-6.5	6.2	6.0-6.5	6.4
Schönheit	3.1	2.0-4.1	3.4	2.5-4.2	3.4	3.0-3.9	5.1

Betrachtet man die Ergebnisse der Mittelwertschätzungen nach Gruppen getrennt, so stellt man fest, daß diese etwas schlechter sind als die Schätzungen der Gesamtpopulation. So überdecken die Interdezilbereiche von 13 Items die Mittelwerte der Shell-Studie. Dabei unterschätzen beide Gruppen die Items Höflichkeit und Schönheit, Gruppe 1 überschätzt das Item soziale Macht und unterschätzt die Items soziale Ordnung und Frieden. Hingegen unterschätzt Gruppe 2 die Items Loslösung, Familie und Natur. Die deutliche Unterschätzung der Items Höflichkeit und Natur könnte, wie oben schon erwähnt, auf einen systematischen Fehler in der Shell-Studie aufgrund von Antwortverzerrungen in Richtung der sozialen Erwünschtheit hindeuten.

Der maximale absolute Unterschied zwischen den beiden Gruppen beträgt in der ersten Welle 1.2, nach der dritten Welle nur noch 0.5 Skalenpunkte. Dabei haben sich die Mittelwerte der Gruppen um maximal 0.8 (Gruppe 1) bzw. 0.7 Skalenpunkte (Gruppe 2) verändert. Die Mittelwerte der Gruppen haben sich mit den Wellen einander angenähert: Betrug die durchschnittliche absolute Abweichung der 18 Items zwischen den Gruppen in der ersten Welle noch 0.52, so sank diese zur dritten Welle auf 0.22. Die Experten der Gruppe 2 waren

hinsichtlich des resultierenden Gruppenmittels etwas mehr zur Veränderung ihrer Urteile bereit als die Experten der Gruppe 1.

Die Ergebnisse des Mittelwertvergleichs der Schätzungen der 18 Items der Gruppen unter Verwendung des doppelten t-Tests sind in Tabelle 14 dargestellt.

Tabelle 14:
T-Tests auf den Unterschied der Schätzungen der beiden Gruppen (Nullhypothese: Die Gruppenmittelwerte unterscheiden sich nicht) Angabe der t-Werte (t) und p-Werte (p)

	1. Welle		2. Welle		3. Welle	
	t	p	t	p	t	p
Harmonie	2.495	.018	1.645	.111	.986	.332
Soz. Macht	-2.210	.035	-.521	.606	-.735	.468
Freiheit	1.219	.232	3.007	.005	1.297	.204
Soz. Ordnung	2.927	.006	1.163	.254	-.894	.379
Anregung	2.360	.025	-.664	.512	-.119	.906
Höflichkeit	-.319	.752	-.140	.890	-1.170	.251
Reichtum	1.458	.155	-1.349	.187	-1.333	.193
Sicherheit	-.336	.739	-.589	.560	-1.896	.068
Kreativität	2.458	.020	.414	.682	1.212	.235
Frieden	.906	.372	.417	.679	.507	.616
Tradition	-.717	.479	.114	.910	-.864	.395
Loslösung	.969	.340	1.039	.307	.648	.522
Familie	-.589	.560	.502	.619	-.652	.519
Natur	1.861	.073	-.126	.900	.253	.802
Abwechslung	1.226	.230	-1.907	.066	1.596	.121
Autorität	.275	.785	-.946	.352	-1.858	.073
Freundschaft	1.511	.141	.664	.512	1.214	.234
Schönheit	3.116	.004	1.316	.198	1.582	.124

Die Werte in Tabelle 14 verdeutlichen, daß sich die Schätzungen der Mittelwerte im Fortgang der Erhebung aufeinander zu bewegt haben. Der maximale t-Wert der Mittelwertunterschiede der beiden Gruppen für die 18 Items betrug in der ersten Welle noch 3.12, in der dritten Welle sank dieser Wert auf 1.89.

Um die Homogenität der einzelnen Gruppen hinsichtlich der Mittelwertschätzungen aufzuzeigen, sollte allerdings der Zweistichprobentest von Kolmogorov-Smirnov verwendet werden. Die Ergebnisse sind in Tabelle 15 dargestellt.

Der kritische Wert von 0.4375 (Signifikanzniveau 5%) wird in zwei von 18 Fällen erreicht, einmal sogar überschritten. Da hier jedoch ein multiples Testproblem vorliegt, kann die Hypothese der Homogenität der Gruppen beibehalten werden.¹⁸

¹⁸ Verwendet man z.B. den multiplen Test nach Bonferroni (vgl. Pruscha 1989: 120f.), so läßt sich die Nullhypothese der Homogenität beibehalten. Diese würde sogar bei einem Signifikanzniveau von 0.2 bei Anwendung des Bonferroni-Tests beibehalten werden.

Tabelle 15:
*Werte der K-S-Statistik für den Gruppenvergleich der Mittelwerte
nach Wellen getrennt*

	1. Welle	2. Welle	3. Welle
Harmonie	.43750	.31250	.25000
Soz. Macht	.37500	.18750	.18750
Freiheit	.31250	.50000	.25000
Soz. Ordnung	.50000	.25000	.25000
Anregung	.37500	.31250	.25000
Höflichkeit	.25000	.18750	.18750
Reichtum	.25000	.43750	.18750
Sicherheit	.18750	.25000	.50000
Kreativität	.37500	.18750	.31250
Frieden	.37500	.18750	.25000
Tradition	.18750	.12500	.18750
Loslösung	.25000	.25000	.25000
Familie	.18750	.25000	.37500
Natur	.43750	.18750	.18750
Abwechslung	.37500	.31250	.43750
Autorität	.12500	.25000	.31250
Freundschaft	.37500	.25000	.37500
Schönheit	.56250	.25000	.43750

- Zur Homogenität der Splittgruppen hinsichtlich der Antworten in der Zusatzbefragung

Die beiden Splittgruppen, so zeigen die Antworten auf die Zusatzbefragung, unterscheiden sich weder hinsichtlich des Verständnisses des Testanliegens, der Orientierungshilfen bei den Schätzungen der ersten und der folgenden Wellen noch hinsichtlich des Durchschnittsalters und der zum Ausfüllen der Fragebogen verwendeten Zeit signifikant.

Aufgrund dieser Betrachtungen scheint die für den Vergleich der Splittaufgaben notwendige Homogenität der Gruppen gewährleistet zu sein.

4.3. Übersicht zu den Ergebnissen des Splitts

In Tabelle 16 sind zunächst als Übersicht die Schätzergebnisse für die Splittfragen¹⁹ dargestellt. Es ist deutlich erkennbar, daß die Experten der Gruppe 2 nicht in der Lage waren, die Verteilungen korrekt zu schätzen. So werden nur fünf von zehn Mittelwerte der Shell-Studie von den Interdezilbereichen der

¹⁹ Im folgenden werden die Fragen für die Schätzergebnisse des Splitts als Fragen 11 und 12 bezeichnet, obwohl die Verteilungsschätzungen für die Verteilungen mit den fünfstufigen Antwortvorgaben in der Variante II des Fragebogens als Fragen Nr. 6 und Nr. 7 aufgeführt waren.

Schätzungen der dritten Welle überdeckt. Hingegen scheinen die Experten in der Lage zu sein, die Mittelwerte der Verteilungen korrekt zu schätzen. Zwar weichen die Punktschätzungen jeweils um ca. 0.3 vom Wert der Shell-Studie ab. Doch liegen die Werte der Shell-Studie sogar für alle drei Wellen in den jeweiligen *Interquartilsbereichen* aller drei Wellen. Diese sind in Tabelle 17 dargestellt.

Tabelle 16: <i>Mittelwerte und Interdezilbereiche für die Fragen des Splitts</i>							
	1. Welle		2. Welle		3. Welle		Shell
	MW	IDB	MW	IDB	MW	IDB	MW
Frage 11:							
sehr dafür	17	5-40	11	5-20	11	5-20	21
eher dafür	26	20-40	24	15-40	28	20-30	32
unentschieden	30	10-40	32	20-50	32	20-40	24
eher dagegen	16	5-30	22	15-30	20	15-25	18
sehr dagegen	11	5-20	11	5-20	9	5-15	5
Frage 12:							
n. so ehrgeizig	13	5-25	11	5-15	11	10-15	1
	17	10-20	21	15-30	21	10-30	8
	25	10-50	34	30-45	32	25-40	32
	25	8-50	22	15-30	23	15-30	42
sehr ehrgeizig	20	5-40	12	10-15	13	10-15	17
Frage 11:	2.9	1.9-3.8	2.8	2.0-3.4	2.8	2.0-3.5	2.5
Frage 12:	3.5	3.0-4.0	3.5	3.0-3.9	3.4	3.0-3.9	3.7

Tabelle 17: <i>Mittelwerte und Interquartilsbereiche (IQB) für die Fragen des Splitts (ausschließlich Mittelpunktschätzungen)</i>							
	1. Welle		2. Welle		3. Welle		Shell
	MW	IQB	MW	IQB	MW	IQB	MW
Frage 11:	2.9	2.0-3.5	2.8	2.4-3.1	2.8	2.3-3.3	2.5
Frage 12:	3.5	3.1-3.9	3.5	3.0-3.8	3.4	3.1-3.7	3.7

Berechnet man aus den Verteilungsschätzungen der Gruppe 2 für jeden Testteilnehmer und jede Welle arithmetische gewichtete Mittelwerte, so läßt sich die Güte der Punktschätzungen, die man aus den Verteilungsschätzungen erhält, beurteilen. Diese Ergebnisse sind in Tabelle 18 dargestellt:

Tabelle 19:
Ergebnisse des doppelten t-Tests und des Zweistichproben K-S-Tests nach Wellen getrennt

Frage 11					
	K-S	N ₁	N ₂	t-Wert	p-Wert
Welle 1	.1875	16	16	-.92	.367
Welle 2	.3750	16	16	-2.45	.021
Welle 3	.5625	14	16	-3.52	.002
Frage 12					
	K-S	N ₁	N ₂	t-Wert	p-Wert
Welle 1	.3125	16	16	-1.16	.254
Welle 2	.7500	16	16	-4.42	.000
Welle 3	.6696	14	16	-3.95	.001

Dieses Ergebnis wird auch durch Aussagen der Testteilnehmer unterstützt.²⁰

Abschließend werden nun die Schätzergebnisse der einzelnen Experten für die Fragen des Splitts betrachtet.

Tabelle 20 kann man entnehmen, daß sich die Schätzergebnisse für Frage 11 über die Wellen trotz der insgesamt schlechten Schätzung verbessert haben. Das gleiche gilt auch für Frage 12. Man sieht ebenfalls anhand der χ^2 -Werte, daß die Verteilungsschätzung der Frage 12 eher gelungen ist als die der Frage 11. Vier Verschlechterungen von Welle 1 zur Welle 3 bei Frage 11 stehen zehn Verbesserungen gegenüber. Bei Frage 12 ist dieses Verhältnis etwas ungünstiger: Dort beträgt es sechs zu neun.

²⁰ So wurden zwei Testteilnehmern, einem aus Gruppe 1, dem anderen aus Gruppe 2, die jeweils andere Variante zur Beurteilung vorgelegt. Beide Experten sagten übereinstimmend, daß die Aufgabe, die Randverteilungen einer fünfstufigen Skala zu schätzen, zu schwer sei. Hingegen beurteilten sie die Aufgabe der Schätzung einer vierstufigen Verteilung als lösbar.

Tabelle 18:
Mittelwerte und Quantile für die Splittfragen; Individuelle Mittelwerte berechnet als gewogene arithmetische Mittel der Verteilungsschätzungen

Frage 11		Streuungsmaße				
	MW	10	25	50	75	90
Welle 1	3.6	2.9	3.2	3.6	3.9	4.5
Welle 2	3.8	3.4	3.6	3.7	4.0	4.3
Welle 3	3.8	3.6	3.6	3.7	4.0	4.1
Shell	3.7					
Frage 12						
	MW	10	25	50	75	90
Welle 1	3.1	2.0	2.9	3.2	3.7	3.9
Welle 2	3.4	3.2	3.3	3.4	3.6	3.7
Welle 3	3.4	3.0	3.4	3.4	3.5	3.7
Shell	2.5					

Die Mittelwertschätzung für Frage 11, die aus den fünfstufigen Antwortvorgaben berechnet wurde, kann als gute Schätzung angesehen werden. Der wahre Wert der Shell-Studie liegt in jeder Welle innerhalb des Interdezilbereiches und sogar innerhalb des (restriktiveren) Interquartilbereiches.

Weniger gelungen sind hingegen die Ergebnisse für Frage 12. Hier überdeckt der Interdezilbereich den Wert der Shell-Studie in der zweiten und dritten Welle nicht. Der Wert der Shell-Studie liegt auch nicht im Interquartilsbereich der 1. Welle. Die finale Schätzung weicht um 1.1 Skalenpunkte vom wahren Wert ab.

Im folgenden soll untersucht werden, in welchem Stadium der Delphi-Erhebung, d.h. ab welcher Welle, die Schätzungen der beiden Splittgruppen sich signifikant unterscheiden.

Tabelle 19 zeigt dazu die Ergebnisse der t-Tests und der Kolmogorov-Smirnov-Tests für die entsprechenden Fragen. Der kritische Wert für die K-S-Statistik liegt trotz der unterschiedlichen Stichprobenumfänge bei 0.4375 (Signifikanzniveau 5%). Aufgrund dieser Ergebnisse muß davon ausgegangen werden, daß die „echte“ Mittelwertschätzung und die Mittelwertschätzung, die man aus der Verteilungsschätzung erhält, sich schon nach der zweiten Welle voneinander unterscheiden (Signifikanzniveau 5%).

Tabelle 20:

Individuelle χ^2 -Werte für Frage 11 Gruppe 2 (5 Antwortkategorien). Die Werte der Shell-Studie werden als erwartete Werte angenommen

ID	Welle 1	Welle 2	Welle 3	Differenz
1	46.2	74.0	----	---
2	35.4	45.1	27.8	7.6
3	11.3	11.7	34.5	-23.2
4	41.4	16.5	12.6	28.8
5	26.2	27.9	24.8	1.4
6	39.2	14.2	3.6	35.6
7	33.1	23.3	17.6	15.5
8	8.8	22.9	12.6	-3.8
9	45.1	34.7	----	---
10	11.3	57.0	34.5	-23.2
11	57.7	51.4	3.6	54.1
12	30.9	18.2	21.2	9.7
13	57.0	29.0	6.1	50.9
14	12.9	12.6	16.4	-3.5
15	143.2	48.5	22.4	120.8
16	74.2	26.5	23.2	51.0

Summe	673.9	513.6	260.7	

Tabelle 21:

Individuelle χ^2 -Werte für Frage 12 Gruppe 2 (5 Antwortkategorien). Die Werte der Shell-Studie werden als erwartete Werte angenommen

ID	Welle 1	Welle 2	Welle 3	Differenz
1	105.4	105.4	-----	---
2	64.1	117.5	115.4	-51.3
3	33.8	106.2	196.9	-163.1
4	395.6	125.0	115.4	280.2
5	92.6	106.2	90.9	1.7
6	43.5	111.2	242.9	-199.4
7	227.4	161.8	89.8	137,6
8	64.1	214.2	228.7	-164.6
9	85.6	40.6	118.6	-33.0
10	612.7	246.7	127.0	485.7
11	419.3	159.2	276.6	152.7
12	1707.3	149.7	99.9	1607.4
13	395.6	109.0	106.2	289,4
14	9.7	105.4	105.4	-95.7
15	400.6	381.8	246.7	153.9
16	38.4	64.2	90.9	-52.5

Summe	4695.5	2304.1	2251.4	

Tabelle 22:
Individuelle χ^2 -Werte für Frage 11 Gruppe 1 (Mittelwertschätzung). Die Werte der Shell-Studie werden als erwartete Werte angenommen

ID	Welle 1	Welle 2	Welle 3	Differenz
1	0.1	0.1	0.3	-0.2
2	0.7	0.1	0.1	0.6
3	0.7	0.3	0.3	0.4
4	0.4	0.1	0.4	0.0
5	0.3	0.2	0.0	0.3
6	0.1	0.0	0.0	0.1
7	0.2	0.7	0.2	0.0
8	0.1	0.1	0.1	0.0
9	0.3	0.2	0.1	0.2
10	0.1	0.3	0.1	0.0
11	0.1	0.0	0.1	0.0
12	0.7	0.0	0.0	0.7
13	0.1	0.1	0.1	0.0
14	0.3	0.0	0.5	-0.2
15	0.1	0.1	0.1	0.0
16	0.1	0.0	0.0	0.1
Summe	4.3	2.4	2.4	

Tabelle 23:
Individuelle χ^2 -Werte für Frage 12 Gruppe 1 (Mittelwertschätzung). Die Werte der Shell-Studie werden als erwartete Werte angenommen

ID	Welle 1	Welle 2	Welle 3	Differenz
1	0.1	0.1	0.1	0.0
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	0.4	0.1	0.1	0.3
5	0.0	0.0	0.0	0.0
6	0.1	0.1	0.1	0.0
7	0.0	0.0	0.0	0.0
8	0.0	0.1	0.1	-0.1
9	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0
11	0.0	0.1	0.1	-0.1
12	0.0	0.0	0.3	-0.3
13	0.1	0.0	0.0	0.1
14	0.1	0.0	0.0	0.1
15	0.1	0.1	0.1	0.0
16	0.1	0.0	0.0	0.1
Summe	1.1	0.8	1.0	

Die individuellen Ergebnisse für die Mittelwertschätzungen sind deutlich besser: Bei Frage 11 haben sich nur zwei Experten von der ersten zur dritten Welle verschlechtert, bei Frage 12 hat sich der χ^2 -Wert von drei Experten in der dritten gegenüber der ersten Welle erhöht.

4.4. Zusammenfassung

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

Das Ergebnis unseres Tests erlaubt weiterhin eine Präzisierung der in Abschnitt 1.2.3. getroffenen Aussagen zur allgemein-menschlichen Fähigkeit, Meinungen zu schätzen (vgl. Noelle-Neumann 1989): So hat sich nur bei einem Aufgabentyp - bei der Schätzung von Mittelwerten - gezeigt, daß die befragten Studenten dazu in der Lage sind, Mehrheitsmeinungen richtig wiederzugeben. Differenziertere Auskünfte über relativ feine Abstufungen dieses Urteils gelingen ihnen jedoch nicht. Die kompliziertere Schätzung von einzelnen Skalenpunkten übersteigt zumindest die Kompetenz der an diesem Test beteiligten Studenten. So wird tatsächlich nur ein *Meinungsklima* wahrgenommen und nicht etwa detaillierte Meinungsdifferenzierungen.

Für die Gestaltung von Delphi Befragungen gilt es - die Ergebnisse dieses Tests verallgemeinernd - zu vermerken, daß differenziertere, über eine Wiedergabe des Meinungsklimas (auf dem Niveau von Mittelwerten) hinausgehende Urteile zumindest von Experten ohne statistische Grundkenntnisse nicht erwartet werden können.

In unseren theoretischen Ausgangsüberlegungen hatten wir die bei Delphi ablaufenden kognitiven Prozesse grundsätzlich als Urteilen unter Unsicherheit beschrieben. Die Schätzungen, die im hier vorgestellten Test von den Untersuchungspersonen abzugeben waren, sind typische Beispiele für derartige Denkprozesse, die tatsächlich auch mit (mehr oder weniger Erfolg) bewältigt worden sind. Die von Gigerenzer et al. beschriebene Bildung mentaler Modelle läßt sich damit plausibel auf die Erklärung kognitiver Prozesse, die bei der Bearbeitung einer Delphi-Studie zur Schätzung von Ergebnissen einer Bevölkerungsbefragung zu leisten sind, anwenden. Es hat sich jedoch auch gezeigt, daß weitere intervenierende Bedingungen zu beachten sind. So tritt das erwartete Ergebnis nicht unter allen Umständen ein. Die von den Testteilnehmern im Rahmen ihrer

mentalen Modelle aktivierten probability cues reichen lediglich, um über Mehrheitsmeinungen zuverlässig Auskunft zu geben. Die für die Schätzung von Verteilungen erforderlichen cues wie z.B. statistische Kenntnisse, das Wissen über die bei Bevölkerungsbefragungen zu erwartenden empirischen Untersuchungsergebnisse usw. fehlen jedoch.

Schließlich besteht ein wesentliches Ergebnis dieser Überlegungen darin, daß mit der Theorie mentaler Modelle eine taugliche methodologische Grundlage für die Aufdeckung, Erklärung und Vorhersage von bei Delphi-Befragungen ablaufenden Prozessen gefunden worden ist.

5. Ausblick

Es kann aufgrund der hier vorgelegten Testergebnisse zunächst ein weiteres Mal festgestellt werden, daß sich die Delphi-Methode dazu eignet, die Ergebnisse von Bevölkerungsbefragungen zu schätzen. Dieses Urteil basiert auf einem Vergleich der Ergebnisse eines drei Wellen umfassenden Delphi-Ansatzes

- erstens mit einmaligen Schätzungen der Experten. So zeigt ein Vergleich der Mittelwerte der Schätzungen der ersten Welle für alle Experten mit denen der dritten Welle bei letzteren deutlich bessere Ergebnisse;
- und zweitens isolierten, individuellen Urteilen.

Auch der Vergleich der individuellen Schätzfehler zeigt, daß sich diese in der dritten Welle gegenüber denen der ersten Welle verringert haben. Diese Einschätzung gilt für alle eingesetzten Arten an Schätzaufgaben.

Die Delphi-Methode hat sich damit - unabhängig vom Grad der Kompetenz der befragten Experten - bewährt, um latent vorhandenes Wissen zu aktivieren bzw. zu reproduzieren und dadurch zu einem Erkenntnisprozeß zu führen. So gelang selbst bei Aufgaben (vierstufige Verteilungsschätzungen), die die methodische und inhaltliche Kompetenz der Testteilnehmer offenbar überstiegen, eine Verringerung des Schätzfehlers von der ersten zur dritten Welle.

Weiterhin konnte in einem speziellen methodischen Experiment gezeigt werden, daß es unter der gegebenen Testanordnung besser gelungen ist, Punktschätzungen vorzunehmen als Verteilungsschätzungen. Inwiefern dieses Ergebnis für Delphi-

Studien generell verallgemeinert werden kann, muß weiteren Studien vorbehalten bleiben.

Darüber hinaus existiert eine Fülle von methodischen Problemen im Zusammenhang mit der Evaluation der Delphi-Methode, die es in weiteren Tests zu bearbeiten gilt. Diese werden sich z.B. auf Prinzipien der Expertenrekrutierung und optimale Strategien der Ergebnismeldung beziehen. Gleichzeitig gilt es, die theoretischen Grundlagen der Methode weiter auszuarbeiten.

Literatur

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

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

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

BMFT, 1993: Deutscher Delphi-Bericht zur Entwicklung von Wissenschaft und Technik, im Auftrag des Bundesministeriums für Forschung und Technologie (BMFT), Bonn.

Bodenhausen, G. V., Wyer, R. S., 1987: Social cognition and social reality: Information acquisition and use in the laboratory and the real world, S.6-41 in: H.-J. Hippler/ N. Schwarz/ S. Sudman (Hrsg.) Social information processing and survey methodology, New York: Springer.

Brockhoff, K., 1979: Delphi-Prognosen im Computer-Dialog. Experimentelle Erprobung und Auswertung kurzfristiger Prognosen. Tübingen: Paul Siebeck.

Brunswik, E., 1943: Organismis achievement and environmental probability, in: Psychological Review, 50: 255-272.

Coates, J. F., 1975, In Defense of Delphi: A Review of Delphi Assessment, Expert Opinion, Forecasting and Group Process by H. Sackman, Technological Forecasting and Social Change 7: 193-194.

Dalkey, N. C., 1969: The Delphi Method: An Experimental Study of Group Opinion. RAND RM 5888-PR, June.

Dutke, St., 1994: Mentale Modelle: Konstrukte des Wissens und Verstehens. Kognitionspsychologische Grundlagen für die Software-Ergonomie. Stuttgart Göttingen: Verlag für Angewandte Psychologie

Geschka, H., 1977: Delphi. In: Bruckmann, G. (Hrsg.), Langfristige Prognosen. Möglichkeiten und Methoden der Langfristprognostik komplexer Systeme. Würzburg/Wien.

Gigerenzer, G., 1984: External validity of laboratory experiments: The frequency-validity relationship, in: American Journal of Psychology, 97: 185-195.

Gigerenzer, G., 1991a: From tool to theories. A heuristic of discovery in cognitive psychology. In: Psychological Review, 98: 254-267.

Gigerenzer, G., 1991b: How to make cognitive illusions disappear: Beyond „heuristics and biases“, in: W. Stroebe/ M. Hewstone (Hrsg.): European review of social psychology, London: Wiley.

Gigerenzer, G./ Hoffrage, U./ Kleinbölting, H., 1991: Probabilistic mental models: A Brunswikian theory of confidence, in: Psychological Review, 98: 506-528.

Gordon, T. J./Helmer, O., 1964: Report on a Long Range Forecasting Study. Rand Paper P-2982. Santa Monica, Cal: Rand Corporation.

Häder, M./Häder, S., 1994a: Die Grundlagen der Delphi-Methode. Ein Literaturbericht, ZUMA-Arbeitsbericht Nr. 94/02.

Häder, M./Häder, S., 1994b: Ergebnisse einer experimentellen Studie zur Delphi-Methode, ZUMA-Arbeitsbericht Nr. 94/05.

Häder, M./Häder, S., 1995: Ein kognitionspsychologischer Zugang zur theoretischen Fundierung der Delphi-Methode, erscheint in: ZUMA NACHRICHTEN 37.

Hasher, L./Goldstein, D./Toppino, T., 1977: Frequency and the conference of referential validity, in: *Journal of Verbal Learning and Verbal Behavior*, 16: 107-112.

Helmer, O./Rescher, N., 1960: *On the Epistemology of the Inexact Sciences*. Project Rand Report R-353.

Hippler, H.-J./Schwarz, N./Noelle-Neumann, E., 1989: *Response Order Effects in Dichotomous Questions: The Impact of Administration Mode*. ZUMA-Arbeitsbericht 89/17. Mannheim.

Howell, W. C./Burnett, S., 1978: Uncertainty measurement: A cognitive taxonomy, in: *Organizational Behavior and Human Performance*, 22: 45-68.

Jugend '92, Maschinenlesbares Codebuch, ZA-Nr. 2323, Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln.

Kerksieck, H.-J., 1972: *Methoden der technologischen Vorausschau im Dienste der Forschungsplanung industrieller Unternehmen unter besonderer Berücksichtigung der Delphi-Methode*. Dissertation. Mannheim.

Linstone, H. A./Turoff, M. (Hrsg.), 1975: *The Delphi Method*. London u.a.: Addison-Wesley.

Markert, St., 1994: Risikoverhalten von Freiern. S. 369-375 in: W. Heckmann/M. A. Koch (Hrsg.), *Sexualverhalten in Zeiten von AIDS*. Berlin: edition sigma.

Nisbett, R. E./Ross, L., 1980: *Human Inference: Strategies and shortcomings of social judgement*. Englewood Cliffs: Prentice-Hall.

Noelle-Neumann, E., 1989: *Öffentliche Meinung. Die Entdeckung der Schweigespirale*, Frankfurt, M. Berlin: Ullstein.

Pruscha, H., 1989: *Angewandte Methoden der Mathematischen Statistik*. Stuttgart: Teubner.

Reuband, K.H., 1990: Interviews, die keine sind. "Erfolge" und "Mißerfolge" beim Fälschen von Interviews. Kölner Zeitschrift für Soziologie und Sozialpsychologie 42: 706-733.

Rowe, G./Wright, G./Bolger, F., 1991: Delphi, A Reevaluation of Research and Theory, Technological Forecasting and Social Change 39: 235-251.

Sackman, H., 1975: Delphi Critique. Expert Opinion, Forecasting, and Group Process. Lexington/Toronto/London: Lexington Books.

Saliger, E./Kunz, C., 1981: Zum Nachweis der Effizienz der Delphi-Methode. In: Zeitschrift für Betriebswirtschaft 51: 470-480.

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 Forschungsgebietes. ZUMA-Arbeitsbericht 90/07. Mannheim.

Seeger, Th., 1979: Die Delphi-Methode. Expertenbefragungen zwischen Prognose und Gruppenmeinungsbildungsprozessen; überprüft am Beispiel von Delphi-Befragungen im Gegenstandsbereich Information und Dokumentation. Diss., Freiburg: HochschulVerlag.

Tversky, A./Kahnemann, D., 1974: Judgement under uncertainty: Heuristics and biases, in: Science, 185: 1124-1131.

Woundenberg, F., 1991: An Evaluation of Delphi, Technological Forecasting and Social Change 40: 131-150.

Zacks,R.T./ Hasher, L./ Sanft, H., 1982: Automatic encoding of event frequency: Further findings, in: Journal of Experimental Psychology: Learning, Memory, and Cognition, 8: 106-116.

Anlagenverzeichnis

1. Fragebogen für die Erhebung der ersten Welle (Variante 1)	49
2. Fragebogen für die Erhebung der zweiten Welle (Variante 1).....	55
3. Fragebogen für die Erhebung der dritten Welle (Variante 1)	62
4. Fragebogen für die Erhebung der vierten Welle (Zusatzbefragung)	69

Anlage 1: Fragebogen für die Erhebung der ersten Welle.

November 1994

Im Rahmen unserer methodischen Tests zur Delphi-Technik wollen wir uns an Sie wenden. Wir möchten mit Ihnen den Versuch wagen, in mehreren Wellen Antworten einer Befragung von Jugendlichen zu schätzen. Es versteht sich von selbst, daß das Ganze nur dann sinnvoll ist, wenn niemand schummelt und sich die Originaldaten besorgt.

Im Jahr 1992 wurden in einer repräsentativen Befragung ca. 2000 Jugendliche in Westdeutschland im Alter von 18 bis 29 Jahren nach ihrer Meinung zu verschiedenen Sachverhalten befragt. Einige der Fragen sind im folgenden aufgeführt.

Wir bitten Sie nun zu schätzen, wie die Antworten bei einzelnen Fragen in dieser Untersuchung ausgefallen sind. Uns interessieren sowohl Antwortverteilungen als auch Mittelwerte. Geben Sie also bitte jeweils den gesuchten Wert an.

Berücksichtigen Sie dabei unbedingt: Es interessiert hier nicht Ihre persönliche Ansicht zu den Sachverhalten!

Gehen Sie bitte davon aus, daß keine Antwortverweigerungen aufgetreten sind.

Wie es bei Delphi-Studien üblich ist, möchten wir Ihnen gezielte Rückinformationen über die Antworten des Expertengremiums in den einzelnen Wellen geben. Um den Verlauf der Meinungsbildung detailliert verfolgen zu können, benötigen wir Ihren Namen auf dem Fragebogen, d.h. die Daten werden auch personenbezogen gespeichert. Es ist selbstverständlich, daß wir die Angaben vertraulich behandeln und daß die Teilnahme freiwillig ist.

Vielen Dank für Ihre Mitarbeit!

Dr. Michael Häder
Dr. Sabine Häder
Dr. Andreas Ziegler

1. Weile (Variante I)

Verteilungsschätzungen:

1. Frage

Die meisten Menschen gehen ja ganz unterschiedlich mit der Zeit in ihrem Leben um und planen auch unterschiedlich. Wie gut beschreiben die Sätze auf diesen Kärtchen Deine Meinung?

Heute ist heute und morgen ist morgen.

überhaupt nicht%
weniger gut%
gut%
sehr gut%

2. Frage (gleicher Fragetext)

Was ich nächste Woche machen werde, überlege ich mir dann, wenn es soweit ist.

überhaupt nicht%
weniger gut%
gut%
sehr gut%

3. Frage (gleicher Fragetext)

Ich tue am liebsten spontan das, wozu ich gerade Lust habe.

überhaupt nicht%
weniger gut%
gut%
sehr gut%

4. Frage (gleicher Fragetext)

Ich höre gerne Geschichten aus guten alten Zeiten.

überhaupt nicht%
weniger gut%
gut%
sehr gut%

5. Frage (gleicher Fragetext)

Im Leben ist alles Zufall.

überhaupt nicht%
weniger gut%
gut%
sehr gut%

Hinweis: Die folgenden Fragen sind als Alternativen gedacht. Die Antworten müssen deshalb jeweils zusammen 100% ergeben, z.B. A1 + A2 = 100%; B1 + B2 = 100% und so weiter.

6. Frage

Auf diesen Karten stehen jeweils unterschiedliche Meinungen, wie man mit seinem Leben umgehen kann. Sage mir bitte, welcher Meinung Du eher zustimmst.

A1 Ich betrachte mein Leben als eine Aufgabe, für die ich da bin und für die ich alle Kräfte einsetze. Ich möchte in meinem Leben etwas leisten, auch wenn das oft schwer und mühsam ist.

....%

A2 Ich möchte mein Leben genießen und mich nicht mehr abmühen als nötig. Man lebt schließlich nur einmal, die Hauptsache ist doch, daß man etwas von seinem Leben hat.

....%

7. Frage (gleicher Fragetext)

B1 Ich finde es wichtig, mein Leben so einzurichten, daß ich ein ganz anderer Mensch bin, der anders ist als alle anderen Menschen in meiner Umgebung.

....%

B2 Ich finde es wichtig, mein Leben so einzurichten, daß betont wird, was ich mit anderen Menschen um mich herum gemeinsam habe und worin wir uns ähnlich sind.

....%

8. Frage (gleicher Fragetext)

C1 Ich finde es am wichtigsten, daß ich im Leben selbständig bin und wirklich selbständig meine eigenen Interessen und Ziele verfolge.

....%

C2 Ich finde es am wichtigsten, daß ich im Leben und meinen Entscheidungen die Interessen und Ziele anderer Menschen berücksichtige.

....%

9. Frage (gleicher Fragetext)

D1 Ich richte mein Leben so ein, daß die Dinge, die ich tue, mir sogleich und direkt etwas bringen und daß ich unmittelbar einen Nutzen sehe oder Spaß daran habe.

....%

D2 Ich richte mein Leben so ein, daß die Dinge, die ich tue, sich langfristig auszahlen, daß ich später im Leben einmal etwas davon habe und auf lange Frist die Früchte meines Tuns ernten kann.

....%

Mittelwertschätzungen

10. Frage

Auf dieser Liste stehen einige Dinge, die wichtig als Werte für das eigene Leben sein können, was man anstrebt und wie man leben möchte. Sage Du mir bitte bei jedem, wie wichtig es Dir ist für Dein Leben.

1 = wäre nicht wichtig und 7 = wäre äußerst wichtig

01. Innere Harmonie (in Frieden mit mir selbst)

Mittelwert: ...

02. Soziale Macht (Kontrolle über andere, Dominanz)

Mittelwert: ...

03. Freiheit (Freiheit des Handelns und des Denkens)

Mittelwert: ...

04. Soziale Ordnung (Stabilität der Gesellschaft)

Mittelwert: ...

05. Ein anregendes Leben (anregende Erfahrungen)

Mittelwert: ...

06. Höflichkeit (gute Umgangsformen)

Mittelwert: ...

07. Reichtum (materieller Besitz, Geld)

Mittelwert: ...

08. Nationale Sicherheit (Schutz meiner Nation gegen Feinde)

Mittelwert: ...

09. Kreativität (Originalität, Phantasie)

Mittelwert: ...

10. Eine Welt in Frieden (frei von Krieg und Konflikt)

Mittelwert: ...

11. Achtung vor Tradition (Erhaltung ehrwürdiger Sitten)

Mittelwert: ...

12. Loslösung (von weltlichen Belangen)

Mittelwert: ...

13. Familiäre Sicherheit (Sicherheit für die geliebten Personen)

Mittelwert: ...

14. Einheit mit der Natur (Einpassung in die Natur)

Mittelwert: ...

15. Ein abwechslungsreiches Leben (erfüllt mit Herausforderungen, Neuem und Veränderungen)

Mittelwert: ...

16. Autorität (ein Recht zu führen und zu bestimmen)

Mittelwert: ...

17. Wahre Freundschaft (enge unterstützende Freunde)

Mittelwert: ...

18. Eine Welt der Schönheit (Schönheit der Natur und Künste)

Mittelwert: ...

Hinweis: Die beiden nächsten Fragen haben eine 5-stufige Antwortskala.

11. Frage

Wie stehst Du zur Vereinigung von ehemaliger DDR und alter Bundesrepublik von heute aus gesehen?

1 sehr dafür und 5 sehr dagegen

Mittelwert: ...

12. Frage

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

Ich bin bei den Zielen, die ich mir für mein Leben setze ...

1 nicht so anspruchsvoll oder ehrgeizig

5 sehr anspruchsvoll und ehrgeizig

Mittelwert: ...

Wie sicher waren Sie sich bei der Beantwortung der Fragen? Wenn Sie sich völlig unsicher waren, kreuzen Sie bitte das Kästchen ganz links an, wenn Sie sich völlig sicher waren, kreuzen Sie das Kästchen ganz rechts an. Mit den Kästchen zwischen den beiden Endpunkten können Sie abstimmen.

völlig unsicher	<input type="checkbox"/>	völlig sicher						
--------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	------------------

Nochmals vielen Dank für Ihre Mitarbeit, wir melden uns bald wieder.

Anlage 2: Fragebogen für die Erhebung der zweiten Welle (Variante 1).

November 1994

Vielen Dank für Ihre Antworten! Unser Delphi-Test geht nun in die nächste Runde. Wir teilen Ihnen dazu die Ergebnisse der ersten Welle mit und bitten um ein zweites Urteil.

Bitte denken Sie wieder an folgendes:

- Schätzen Sie bitte - wie schon beim letzten Mal - die Ergebnisse einer Befragung, bei der ca. 2000 Jugendliche in Westdeutschland im Alter von 18 bis 29 Jahren um ihrer Meinung zu verschiedenen Sachverhalten gebeten wurden.
- Gehen Sie bitte wieder davon aus, daß keine Antwortverweigerungen aufgetreten sind.
- Nicht vergessen: Es interessiert hier nicht Ihre persönliche Ansicht zu den Sachverhalten!
- Wir geben Ihnen als Rückinformation an: die Mittelwerte (MW) und als Streuungsmaß die Interdezilbereiche (IDB), d.h. die Streubreite der mittleren 80 Prozent aller Urteile. Diese Werte ergeben sich aus den Urteilen von insgesamt 32 Testteilnehmern.

Vielen Dank für Ihre Mitarbeit!

Dr. Michael Häder
Dr. Sabine Häder
Dr. Andreas Ziegler

2. Welle (Variante I)

Verteilungsschätzungen:

1. Frage

Die meisten Menschen gehen ja ganz unterschiedlich mit der Zeit in ihrem Leben um und planen auch unterschiedlich. Wie gut beschreiben die Sätze auf diesen Kärtchen Deine Meinung?

	MW	IDB%
Heute ist heute und morgen ist morgen.			
überhaupt nicht	22%	5 - 50%%
weniger gut	35%	22 - 50%%
gut	31%	10 - 50%%
sehr gut	12%	5 - 20%%

2. Frage (gleicher Fragetext)

Was ich nächste Woche machen werde, überlege ich mir dann, wenn es soweit ist.

überhaupt nicht	21%	10 - 40%%
weniger gut	32%	20 - 40%%
gut	32%	15 - 50%%
sehr gut	15%	10 - 25%%

3. Frage (gleicher Fragetext)

Ich tue am liebsten spontan das, wozu ich gerade Lust habe.

überhaupt nicht	9%	2 - 15%%
weniger gut	20%	10 - 35%%
gut	41%	25 - 55%%
sehr gut	30%	10 - 54%%

4. Frage (gleicher Fragetext)

Ich höre gerne Geschichten aus guten alten Zeiten.

überhaupt nicht	33%	4 - 60%%
weniger gut	31%	10 - 50%%
gut	20%	8 - 35%%
sehr gut	16%	5 - 50%%

5. Frage (gleicher Fragetext)

Im Leben ist alles Zufall.

überhaupt nicht	22%	5 - 50%%
weniger gut	34%	15 - 50%%
gut	28%	10 - 40%%
sehr gut	16%	5 - 30%%

Hinweis: Die folgenden Fragen sind als Alternativen gedacht. Die Antworten müssen deshalb jeweils zusammen 100% ergeben, z.B. A1 + A2 = 100%; B1 + B2 = 100% und so weiter.

6. Frage

Auf diesen Karten stehen jeweils unterschiedliche Meinungen, wie man mit seinem Leben umgehen kann. Sage mir bitte, welcher Meinung Du eher zustimmst.

A1 Ich betrachte mein Leben als eine Aufgabe, für die ich da bin und für die ich alle Kräfte einsetze. Ich möchte in meinem Leben etwas leisten, auch wenn das oft schwer und mühsam ist.

MW: 45% IDB: 25 - 65%%

A2 Ich möchte mein Leben genießen und mich nicht mehr abmühen als nötig. Man lebt schließlich nur einmal, die Hauptsache ist doch, daß man etwas von seinem Leben hat.

MW: 55% IDB: 30 - 75%%

7. Frage (gleicher Fragetext)

B1 Ich finde es wichtig, mein Leben so einzurichten, daß ich ein ganz anderer Mensch bin, der anders ist als alle anderen Menschen in meiner Umgebung.

MW: 44% IDB: 20 - 80%%

B2 Ich finde es wichtig, mein Leben so einzurichten, daß betont wird, was ich mit anderen Menschen um mich herum gemeinsam habe und worin wir uns ähnlich sind.

MW: 56% IDB: 20 - 80%%

8. Frage (gleicher Fragetext)

C1 Ich finde es am wichtigsten, daß ich im Leben selbständig bin und wirklich selbständig meine eigenen Interessen und Ziele verfolge.

MW: 63% IDB: 45 - 80%%

C2 Ich finde es am wichtigsten, daß ich im Leben und meinen Entscheidungen die Interessen und Ziele anderer Menschen berücksichtige.

MW: 37% IDB: 20 - 55%%

9. Frage (gleicher Fragetext)

D1 Ich richte mein Leben so ein, daß die Dinge, die ich tue, mir sogleich und direkt etwas bringen und daß ich unmittelbar einen Nutzen sehe oder Spaß daran habe.

MW: 51% IDB: 29 - 75%%

D2 Ich richte mein Leben so ein, daß die Dinge, die ich tue, sich langfristig auszahlen, daß ich später im Leben einmal etwas davon habe und auf lange Frist die Früchte meines Tuns ernten kann.

MW: 49% IDB: 25 - 71%%

Mittelwertschätzungen

10. Frage

Auf dieser Liste stehen einige Dinge, die wichtig als Werte für das eigene Leben sein können, was man anstrebt und wie man leben möchte. Sage Du mir bitte bei jedem, wie wichtig es Dir ist für Dein Leben.

1 = wäre nicht wichtig und 7 = wäre äußerst wichtig

01. Innere Harmonie (in Frieden mit mir selbst)

MW: 5.7 IDB: 4.5 - 6.4 geschätzter Mittelwert: ...

02. Soziale Macht (Kontrolle über andere, Dominanz)

MW: 3.5 IDB: 2.0 - 5.0 geschätzter Mittelwert: ...

03. Freiheit (Freiheit des Handelns und des Denkens)

MW: 6.1 IDB: 5.0 - 6.7 geschätzter Mittelwert: ...

04. Soziale Ordnung (Stabilität der Gesellschaft)

MW: 4.8 IDB: 3.2 - 6.0 geschätzter Mittelwert: ...

05. Ein anregendes Leben (anregende Erfahrungen)

MW: 5.3 IDB: 4.0 - 6.2 geschätzter Mittelwert: ...

06. Höflichkeit (gute Umgangsformen)
 MW: 3.8 IDB: 2.5 - 5.0 geschätzter Mittelwert: ...
07. Reichtum (materieller Besitz, Geld)
 MW: 4.5 DB: 3.3 - 6.0 geschätzter Mittelwert: ...
08. Nationale Sicherheit (Schutz meiner Nation gegen Feinde)
 MW: 3.9 IDB: 2.5 - 5.9 geschätzter Mittelwert: ...
09. Kreativität (Originalität, Phantasie)
 MW: 4.7 IDB: 3.1 - 6.0 geschätzter Mittelwert: ...
10. Eine Welt in Frieden (frei von Krieg und Konflikt)
 MW: 5.8 IDB: 4.7 - 7.0 geschätzter Mittelwert: ...
11. Achtung vor Tradition (Erhaltung ehrwürdiger Sitten)
 MW: 3.0 IDB: 2.0 - 4.5 geschätzter Mittelwert: ...
12. Loslösung (von weltlichen Belangen)
 MW: 2.8 IDB: 1.8 - 4.0 geschätzter Mittelwert: ...
13. Familiäre Sicherheit (Sicherheit für die geliebten Personen)
 MW: 5.3 IDB: 4.0 - 6.5 geschätzter Mittelwert: ...
14. Einheit mit der Natur (Einpassung in die Natur)
 MW: 4.6 IDB: 3.5 - 6.0 geschätzter Mittelwert: ...

15. Ein abwechslungsreiches Leben (erfüllt mit Herausforderungen, Neuem und Veränderungen)

MW: 5.2 IDB: 3.7 - 6.2 geschätzter Mittelwert: ...

16. Autorität (ein Recht zu führen und zu bestimmen)

MW: 3.9 IDB: 2.0 - 5.0 geschätzter Mittelwert: ...

17. Wahre Freundschaft (enge unterstützende Freunde)

MW: 5.9 IDB: 5.0 - 6.8 geschätzter Mittelwert: ...

18. Eine Welt der Schönheit (Schönheit der Natur und Künste)

MW: 3.7 IDB: 2.0 - 5.0 geschätzter Mittelwert: ...

Hinweis: Die beiden nächsten Fragen haben eine 5-stufige Antwortskala.

11. Frage

Wie stehst Du zur Vereinigung von ehemaliger DDR und alter Bundesrepublik von heute aus gesehen?

1 sehr dafür und 5 sehr dagegen

MW: 2.9 IDB: 2.0 - 3.8 geschätzter Mittelwert: ...

12. Frage

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

Ich bin bei den Zielen, die ich mir für mein Leben setze ...

1 nicht so anspruchsvoll oder ehrgeizig

5 sehr anspruchsvoll und ehrgeizig

MW: 3.5 IDB: 3.0 - 4.0 geschätzter Mittelwert: ...

Und nun bitten wir Sie wieder um Ihre Einschätzung:

Wie sicher waren Sie sich bei der Beantwortung der Fragen? Wenn Sie sich völlig unsicher waren, kreuzen Sie bitte das Kästchen ganz links an, wenn Sie sich völlig sicher waren, kreuzen Sie das Kästchen ganz rechts an. Mit den Kästchen zwischen den beiden Endpunkten können Sie abstufen.

völlig unsicher	<input type="checkbox"/>	völlig sicher						
--------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	------------------

Nochmals vielen Dank für Ihre Mitarbeit, wir melden uns bald wieder.

Anlage 3: Fragebogen für die Erhebung der dritten Welle.

Dezember 1994

Vielen Dank für Ihre Antworten! Unser Delphi-Test geht nun in die letzte Runde. Wir teilen Ihnen dazu wieder die Ergebnisse der vorigen Welle mit und bitten um ein drittes Urteil.

Bitte denken Sie auch dieses Mal wieder an folgendes:

- Geben Sie bitte unbedingt Ihre "Versuchspersonen-Nummer" auf diesem Fragebogen und nachher auch auf dem Zusatzbogen an.
- Schätzen Sie bitte - wie schon beim letzten Mal - die Ergebnisse einer Befragung, bei der ca. 2000 Jugendliche in Westdeutschland im Alter von 18 bis 29 Jahren um ihrer Meinung zu verschiedenen Sachverhalten gebeten wurden.
- Gehen Sie bitte wieder davon aus, daß keine Antwortverweigerungen aufgetreten sind.
- Nicht vergessen: Es interessiert hier nicht Ihre persönliche Ansicht zu den Sachverhalten!
- Wir geben Ihnen als Rückinformation an: die Mittelwerte (MW) und als Streuungsmaß die Interdezilbereiche (IDB), d.h. die Streubreite der mittleren 80 Prozent aller Urteile. Diese Werte ergeben sich aus den Urteilen von insgesamt 32 Testteilnehmern.

Vielen Dank für Ihre Mitarbeit!

Dr. Michael Häder
Dr. Sabine Häder
Dr. Andreas Ziegler

3. Welle (Variante I)

Verteilungsschätzungen:

1. Frage

Die meisten Menschen gehen ja ganz unterschiedlich mit der Zeit in ihrem Leben um und planen auch unterschiedlich. Wie gut beschreiben die Sätze auf diesen Kärtchen Deine Meinung?

	MW	IDB%
Heute ist heute und morgen ist morgen.			
überhaupt nicht	23%	10 - 40%%
weniger gut	33%	25 - 40%%
gut	31%	20 - 40%%
sehr gut	13%	8 - 20%%

2. Frage (gleicher Fragetext)

Was ich nächste Woche machen werde, überlege ich mir dann, wenn es soweit ist.

überhaupt nicht	20%	10 - 30%%
weniger gut	31%	25 - 40%%
gut	33%	25 - 40%%
sehr gut	16%	10 - 21%%

3. Frage (gleicher Fragetext)

Ich tue am liebsten spontan das, wozu ich gerade Lust habe.

überhaupt nicht	10%	5 - 20%%
weniger gut	21%	10 - 30%%
gut	40%	26 - 50%%
sehr gut	29%	20 - 40%%

4. Frage (gleicher Fragetext)

Ich höre gerne Geschichten aus guten alten Zeiten.

überhaupt nicht	32%	15 - 50%%
weniger gut	30%	15 - 40%%
gut	23%	10 - 35%%
sehr gut	15%	5 - 20%%

5. Frage (gleicher Fragetext)

Im Leben ist alles Zufall.

überhaupt nicht	22%	10 - 35%%
weniger gut	32%	15 - 50%%
gut	30%	20 - 40%%
sehr gut	16%	5 - 20%%

Hinweis: Die folgenden Fragen sind als Alternativen gedacht. Die Antworten müssen deshalb jeweils zusammen 100% ergeben, z.B. A1 + A2 = 100%; B1 + B2 = 100% und so weiter.

6. Frage

Auf diesen Karten stehen jeweils unterschiedliche Meinungen, wie man mit seinem Leben umgehen kann. Sage mir bitte, welcher Meinung Du eher zustimmst.

A1 Ich betrachte mein Leben als eine Aufgabe, für die ich da bin und für die ich alle Kräfte einsetze. Ich möchte in meinem Leben etwas leisten, auch wenn das oft schwer und mühsam ist.

MW: 43% IDB: 30 - 60%%

A2 Ich möchte mein Leben genießen und mich nicht mehr abmühen als nötig. Man lebt schließlich nur einmal, die Hauptsache ist doch, daß man etwas von seinem Leben hat.

MW: 57% IDB: 40 - 70%%

7. Frage (gleicher Fragetext)

B1 Ich finde es wichtig, mein Leben so einzurichten, daß ich ein ganz anderer Mensch bin, der anders ist als alle anderen Menschen in meiner Umgebung.

MW: 52% IDB: 35 - 70%%

B2 Ich finde es wichtig, mein Leben so einzurichten, daß betont wird, was ich mit anderen Menschen um mich herum gemeinsam habe und worin wir uns ähnlich sind.

MW: 48% IDB: 30 - 65%%

8. Frage (gleicher Fragetext)

C1 Ich finde es am wichtigsten, daß ich im Leben selbständig bin und wirklich selbständig meine eigenen Interessen und Ziele verfolge.

MW: 64% IDB: 55 - 70%%

C2 Ich finde es am wichtigsten, daß ich im Leben und meinen Entscheidungen die Interessen und Ziele anderer Menschen berücksichtige.

MW: 36% IDB: 30 - 45%%

9. Frage (gleicher Fragetext)

D1 Ich richte mein Leben so ein, daß die Dinge, die ich tue, mir sogleich und direkt etwas bringen und daß ich unmittelbar einen Nutzen sehe oder Spaß daran habe.

MW: 55% IDB: 40 - 75%%

D2 Ich richte mein Leben so ein, daß die Dinge, die ich tue, sich langfristig auszahlen, daß ich später im Leben einmal etwas davon habe und auf lange Frist die Früchte meines Tuns ernten kann.

MW: 45% IDB: 25 - 60%%

Mittelwertschätzungen

10. Frage

Auf dieser Liste stehen einige Dinge, die wichtig als Werte für das eigene Leben sein können, was man anstrebt und wie man leben möchte. Sage Du mir bitte bei jedem, wie wichtig es Dir ist für Dein Leben.

1 = wäre nicht wichtig und 7 = wäre äußerst wichtig

01. Innere Harmonie (in Frieden mit mir selbst)

MW: 5.8 IDB: 5.0 - 6.2

geschätzter Mittelwert: ...

02. Soziale Macht (Kontrolle über andere, Dominanz)

MW: 3.4 IDB: 2.4 - 4.7

geschätzter Mittelwert: ...

03. Freiheit (Freiheit des Handelns und des Denkens)

MW: 6.2 IDB: 5.8 - 6.6

geschätzter Mittelwert: ...

04. Soziale Ordnung (Stabilität der Gesellschaft)

MW: 4.5 IDB: 3.5 - 5.4

geschätzter Mittelwert: ...

05. Ein anregendes Leben (anregende Erfahrungen)

MW: 5.4 IDB: 4.1 - 6.1

geschätzter Mittelwert: ...

06. Höflichkeit (gute Umgangsformen)	MW: 3.7 IDB: 2.8 - 4.5	geschätzter Mittelwert: ...
07. Reichtum (materieller Besitz, Geld)	MW: 4.7 DB: 4.0 - 5.5	geschätzter Mittelwert: ...
08. Nationale Sicherheit (Schutz meiner Nation gegen Feinde)	MW: 4.0 IDB: 2.9 - 5.0	geschätzter Mittelwert: ...
09. Kreativität (Originalität, Phantasie)	MW: 4.8 IDB: 4.0 - 5.7	geschätzter Mittelwert: ...
10. Eine Welt in Frieden (frei von Krieg und Konflikt)	MW: 6.0 IDB: 5.0 - 6.8	geschätzter Mittelwert: ...
11. Achtung vor Tradition (Erhaltung ehrwürdiger Sitten)	MW: 2.9 IDB: 2.0 - 3.7	geschätzter Mittelwert: ...
12. Loslösung (von weltlichen Belangen)	MW: 2.7 IDB: 2.0 - 3.1	geschätzter Mittelwert: ...
13. Familiäre Sicherheit (Sicherheit für die geliebten Personen)	MW: 5.4 IDB: 4.2 - 6.3	geschätzter Mittelwert: ...
14. Einheit mit der Natur (Einpassung in die Natur)	MW: 4.9 IDB: 3.9 - 6.0	geschätzter Mittelwert: ...

15. Ein abwechslungsreiches Leben (erfüllt mit Herausforderungen, Neuem und Veränderungen)

MW: 5.5 IDB: 4.8 - 6.0 geschätzter Mittelwert: ...

16. Autorität (ein Recht zu führen und zu bestimmen)

MW: 3.6 IDB: 2.9 - 4.9 geschätzter Mittelwert: ...

17. Wahre Freundschaft (enge unterstützende Freunde)

MW: 6.2 IDB: 5.9 - 6.5 geschätzter Mittelwert: ...

18. Eine Welt der Schönheit (Schönheit der Natur und Künste)

MW: 3.6 IDB: 2.5 - 4.5 geschätzter Mittelwert: ...

Hinweis: Die beiden nächsten Fragen haben eine 5-stufige Antwortskala.

11. Frage

Wie stehst Du zur Vereinigung von ehemaliger DDR und alter Bundesrepublik von heute aus gesehen?

1 sehr dafür und 5 sehr dagegen

MW: 2.8 IDB: 2.1 - 3.4 geschätzter Mittelwert: ...

12. Frage

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

Ich bin bei den Zielen, die ich mir für mein Leben setze ...

1 nicht so anspruchsvoll oder ehrgeizig

5 sehr anspruchsvoll und ehrgeizig

MW: 3.5 IDB: 3.0 - 3.8 geschätzter Mittelwert: ...

Und nun bitten wir Sie wieder um Ihre Einschätzung:

Wie sicher waren Sie sich bei der Beantwortung der Fragen? Wenn Sie sich völlig unsicher waren, kreuzen Sie bitte das Kästchen ganz links an, wenn Sie sich völlig sicher waren, kreuzen Sie das Kästchen ganz rechts an. Mit den Kästchen zwischen den beiden Endpunkten können Sie abstufen.

völlig unsicher	<input type="checkbox"/>	völlig sicher						
--------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	------------------

Nochmals vielen Dank für Ihre Mitarbeit. Wir werden Sie über die Ergebnisse unseres Tests informieren.

Anlage 4: Fragebogen für die Erhebung der vierten Welle (Zusatzbefragung).

Nach dem Abschluß der letzten Welle wollen wir noch einige zusätzliche Informationen für die Auswertung sammeln. Bitte beantworten Sie deshalb nun noch die Fragen im folgenden Bogen, in dem es um Ihre persönlichen Eindrücke geht.

Sie können Ihre Antworten jeweils abstufen. Bitte kreuzen Sie bei jeder Frage das zutreffende Kästchen an!

1. Wie verständlich war für Sie das Anliegen unseres Tests?

völlig verständlich	<input type="checkbox"/>	überhaupt nicht verständlich				
------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	------------------------------------

2. Wie sehr haben Sie sich bisher mit Fragen des Zeitbewußtseins - wie sie in unserem Test vorkamen - beschäftigt?

sehr beschäftigt	<input type="checkbox"/>	gar nicht beschäftigt				
---------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	-----------------------------

3. Wie sehr haben Sie sich bisher mit Wichtigkeitsurteilen beschäftigt?

sehr beschäftigt	<input type="checkbox"/>	gar nicht beschäftigt				
---------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	-----------------------------

4. Wie stark haben Sie sich bei Ihrer ersten Schätzung von folgendem leiten lassen?

Ich habe mich leiten lassen...

a) von meinen Erfahrungen mit ähnlichen Fragestellungen:

sehr stark	<input type="checkbox"/>	gar nicht				
---------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------

b) von meinen Erfahrungen mit statistischen Verteilungsformen:

sehr stark gar nicht

c) spontan von meinem Gefühl:

sehr stark gar nicht

d) von meinen Erfahrungen mit anderen Jugendlichen:

sehr stark gar nicht

e) davon, was man so über Jugendliche liest und hört:

sehr stark gar nicht

f) davon, wie ich die Fragen wohl selbst beantwortet hätte:

sehr stark gar nicht

g) von etwas anderem, nämlich:

5. Wovon haben Sie sich bei Ihren Schätzungen in den folgenden Wellen leiten lassen?

a) Ich habe (nochmals) über den Inhalt nachgedacht.

trifft völlig zu trifft gar nicht zu

b) Ich habe über die rückgemeldeten statistischen Werte nachgedacht.

trifft völlig zu	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	trifft gar nicht zu
------------------------	---	---------------------------

c) Ich war bemüht, mein ursprüngliches Urteil beizubehalten.

trifft völlig zu	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	trifft gar nicht zu
------------------------	---	---------------------------

6. Für wie kompetent halten Sie Studenten mit soziologischer Ausbildung für die Schätzung solcher Befragungsergebnisse?

sehr kompetent	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	gar nicht kompetent
-------------------	---	------------------------

7. Hat Ihnen die Mitarbeit an unserem Test Spaß gemacht oder keinen Spaß gemacht?

sehr viel Spaß	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	gar keinen Spaß
----------------------	---	-----------------------

Tragen Sie bitte nun noch die Antworten auf folgende Fragen ein!

8. Wie lange haben Sie etwa für das Ausfüllen der Fragebögen benötigt?

In der ersten Welle etwa:

In der zweiten Welle etwa:

In der dritten Welle etwa:

.... Minuten

.... Minuten

.... Minuten

9. Wie alt sind Sie?

..... Jahre

10. Gibt es noch etwas, das wir über diesen Test von Ihnen wissen sollten? Wir interessieren uns für alles, was Ihnen beim Ausfüllen aufgefallen ist!

Zunächst nochmals vielen Dank für Ihre Mitarbeit. Erste Ergebnisse dieses Tests werden wir Ihnen bald vorstellen können.

ZUMA-Arbeitsberichte

- 80/15 Gerhard Arminger, Willibald Nagl, Karl F. Schuessler
Methoden der Analyse zeitbezogener Daten. Vortragsskripten der ZUMA
Arbeitstagung vom 25.9. bis 5.10.79
- 81/07 Erika Brückner, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer, Peter Schmidt
Methodenbericht zum "ALLBUS 1980"
- 81/19 Manfred Küchler, Thomas P. Wilson, Don H. Zimmerman
Integration von qualitativen und quantitativen Forschungsansätzen
- 82/03 Gerhard Arminger, Horst Busse, Manfred Küchler
Verallgemeinerte Lineare Modelle in der empirischen Sozialforschung
- 82/08 Glenn R. Carroll
Dynamic analysis of discrete dependent variables: A didactic essay
- 82/09 Manfred Küchler
Zur Messung der Stabilität von Wählerpotentialen
- 82/10 Manfred Küchler
Zur Konstanz der Recallfrage
- 82/12 Rolf Porst
"ALLBUS 1982" - Systematische Variablenübersicht und erste Ansätze zu
einer Kritik des Fragenprogramms
- 82/13 Peter Ph. Mohler
SAR - Simple AND Retrieval mit dem Siemens-EDT-
Textmanipulationsprogramm
- 82/14 Cornelia Krauth
Vergleichsstudien zum "ALLBUS 1980"
- 82/21 Werner Hagstotz, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer
Methodenbericht zum "ALLBUS 1982"
- 83/09 Bernd Wegener
Two approaches to the analysis of judgments of prestige: Interindividual
differences and the general scale
- 83/11 Rolf Porst
Synopsis der ALLBUS-Variablen. Die Systematik des ALLBUS-
Fragenprogramms und ihre inhaltliche Ausgestaltung im ALLBUS 1980 und
ALLBUS 1982
- 84/01 Manfred Küchler, Peter Ph. Mohler
Qualshop (ZUMA-Arbeitstagung zum "Datenmanagement bei qualitativen
Erhebungsverfahren") - Sammlung von Arbeitspapieren und -berichten,
Teil I + II
- 84/02 Bernd Wegener
Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-Prestige-
Skala
- 84/03 Peter Prüfer, Margrit Rexroth
Erfahrungen mit einer Technik zur Bewertung von Interviewerverhalten

- 84/04 Frank Faulbaum
Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982
- 84/05 Jürgen Hoffmeyer-Zlotnik
Wohnquartiersbeschreibung. Ein Instrument zur Bestimmung des sozialen Status von Zielhaushalten
- 84/07 Gabriele Hippler, Hans-Jürgen Hippler
Reducing Refusal Rates in the Case of Threatening Questions: The "Door-in-the-Face" Technique
- 85/01 Hartmut Esser
Befragtenverhalten als "rationales Handeln" - Zur Erklärung von Antwortverzerrungen in Interviews
- 85/03 Rolf Porst, Peter Prüfer, Michael Wiedenbeck, Klaus Zeifang
Methodenbericht zum "ALLBUS 1984"
- 86/01 Dagmar Krebs
Zur Konstruktion von Einstellungsskalen im interkulturellen Vergleich
- 86/02 Hartmut Esser
Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung
- 86/03 Bernd Wegener
Prestige and Status as Function of Unit Size
- 86/04 Frank Faulbaum
Very Soft Modeling: The Logical Specification and Analysis of Complex Process Explanations with Arbitrary Degrees of Underidentification and Variables of Arbitrary Aggregation and Measurement Levels
- 86/05 Peter Prüfer, Margrit Rexroth (Übersetzung Dorothy Duncan)
On the Use of the Interaction Coding Technique
- 86/06 Hans-Peter Kirschner
Zur Kessler-Greenberg-Zerlegung der Varianz der Meßdifferenz zwischen zwei Meßzeitpunkten einer Panel-Befragung
- 86/07 Georg Erdmann
Ansätze zur Abbildung sozialer Systeme mittels nicht-linearer dynamischer Modelle
- 86/09 Heiner Ritter
Einige Ergebnisse von Vergleichstests zwischen den PC- und Mainframe-Versionen von SPSS und SAS
- 86/11 Günter Rothe
Bootstrap in generalisierten linearen Modellen
- 87/01 Klaus Zeifang
Die Test-Retest-Studie zum ALLBUS 1984 - Tabellenband
- 87/02 Klaus Zeifang
Die Test-Retest-Studie zum ALLBUS 1984 - Abschlußbericht
- 87/04 Barbara Erbslöh, Michael Wiedenbeck
Methodenbericht zum "ALLBUS 1986"
- 87/05 Norbert Schwarz, Julia Bienias
What Mediates the Impact of Response Alternatives on Behavioral Reports?

- 87/06 Norbert Schwarz, Fritz Strack, Gesine Müller, Brigitte Chassein
The Range of Response Alternatives may determine the Meaning of the Question: Further Evidence on Informative Functions of Response Alternatives
- 87/07 Fritz Strack, Leonard L. Martin, Norbert Schwarz
The Context Paradox in Attitude Surveys: Assimilation or Contrast?
- 87/08 Gudmund R. Iversen
Introduction to Contextual Analysis
- 87/09 Seymour Sudman, Norbert Schwarz
Contributions of Cognitive Psychology to Data Collection in Marketing Research
- 87/10 Norbert Schwarz, Fritz Strack, Denis Hilton, Gabi Naderer
Base-Rates, Representativeness, and the Logic of Conversation
- 87/11 George F. Bishop, Hans-Jürgen Hippler, Norbert Schwarz, Fritz Strack
A Comparison of Response Effects in Self-Administered and Telephone Surveys
- 87/12 Norbert Schwarz
Stimmung als Information. Zum Einfluß von Stimmungen und Emotionen auf evaluative Urteile
- 88/01 Antje Nebel, Fritz Strack, Norbert Schwarz
Tests als Treatment: Wie die psychologische Messung ihren Gegenstand verändert
- 88/02 Gerd Bohner, Herbert Bless, Norbert Schwarz, Fritz Strack
What Triggers Causal Attributions? The Impact of Valence and Subjective Probability
- 88/03 Norbert Schwarz, Fritz Strack
The Survey Interview and the Logic of Conversation: Implications for Questionnaire Construction
- 88/04 Hans-Jürgen Hippler, Norbert Schwarz
"No Opinion"-Filters: A Cognitive Perspective
- 88/05 Norbert Schwarz, Fritz Strack
Evaluating One's Life: A Judgment of Subjective Well-Being
- 88/06 Norbert Schwarz, Herbert Bless, Gerd Bohner, Uwe Harlacher, Margit Kellenbenz
Response Scales as Frames of Reference: The Impact of Frequency Range on Diagnostic Judgments
- 88/07 Michael Braun
ALLBUS-Bibliographie (7. Fassung, Stand: 30.6.88)
- 88/08 Günter Rothe
Ein Ansatz zur Konstruktion inferenzstatistisch verwertbarer Indices
- 88/09 Ute Hauck, Reiner Trometer
Methodenbericht International Social Survey Program - ISSP 1987
- 88/10 Norbert Schwarz
Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction
- 88/11 Norbert Schwarz, B. Scheuring (sub.)
Judgments of relationship satisfaction: Inter- and intraindividual comparison strategies as a function of questionnaire structure
- 88/12 Rolf Porst, Michael Schneid
Ausfälle und Verweigerungen bei Panelbefragungen - Ein Beispiel -

- 88/13 Cornelia Züll
SPSS-X. Anmerkungen zur Siemens BS2000 Version
- 88/14 Michael Schneid
Datenerhebung am PC - Vergleich der Interviewprogramme "internerv+" und "THIS"
- 88/15 Norbert Schwarz, Bettina Scheuring
Die Vergleichsrichtung bestimmt das Ergebnis von Vergleichsprozessen: Ist - Idealdiskrepanzen in der Partnerwahrnehmung
- 88/16 Norbert Schwarz, Bettina Scheuring
Die Vergleichsrichtung bestimmt das Ergebnis von Vergleichsprozessen: Ist - Idealdiskrepanzen in der Beziehungsbeurteilung
- 89/01 Norbert Schwarz, George F. Bishop, Hans-J. Hippler, Fritz Strack
Psychological Sources of Response Effects in Self-Administered And Telephone Surveys
- 89/02 Michael Braun, Reiner Trometer, Michael Wiedenbeck
Methodenbericht. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften - ALLBUS 1988
- 89/03 Norbert Schwarz
Feelings as Information: Informational and Motivational Functions of Affective States
- 89/04 Günter Rothe
Jackknife and Bootstrap: Resampling-Verfahren zur Genauigkeitsschätzung von Parameterschätzungen
- 89/05 Herbert Bless, Gerd Bohner, Norbert Schwarz und Fritz Strack
Happy and Mindless? Moods and the Processing of Persuasive Communications
- 89/06 Gerd Bohner, Norbert Schwarz und Stefan E. Hormuth
Die Stimmungs-Skala: Eine deutsche Version des "Mood Survey" von Underwood und Froming
- 89/07 Ulrich Mueller
Evolutionary Fundamentals of Social Inequality, Dominance and Cooperation
- 89/08 Robert Huckfeldt
Noncompliance and the Limits of Coercion: The Problematic Enforcement of Unpopular Laws
- 89/09 Peter Ph. Mohler, Katja Frehsen und Ute Hauck
CUI - Computerunterstützte Inhaltsanalyse. Grundzüge und Auswahlbibliographie zu neueren Anwendungen
- 89/10 Cornelia Züll, Peter Ph. Mohler
Der General Inquirer III - Ein Dinosaurier für die historische Forschung
- 89/11 Fritz Strack, Norbert Schwarz, Brigitte Chassein, Dieter Kern, Dirk Wagner
The Salience of Comparison Standards and the Activation of Social Norms: Consequences for Judgments of Happiness and their Communication
- 89/12 Jutta Kreiselmaier, Rolf Porst
Methodische Probleme bei der Durchführung telefonischer Befragungen: Stichprobenziehung und Ermittlung von Zielpersonen, Ausschöpfung und Nonresponse, Qualität der Daten
- 89/13 Rainer Mathes
Modulsystem und Netzwerktechnik. Neuere inhaltsanalytische Verfahren zur Analyse von Kommunikationsinhalten

- 89/14 Jutta Kreiselmeier, Peter Prüfer, Margrit Rexroth
Der Interviewer im Pretest. Evaluation der Interviewerleistung und Entwurf eines neuen Pretestkonzepts. April 1989
- 89/15 Henrik Tham
Crime as a Social Indicator
- 89/16 Ulrich Mueller
Expanding the Theoretical and Methodological Framework of Social Dilemma Research
- 89/17 Hans-J. Hippler, Norbert Schwarz, Elisabeth Noelle-Neumann
Response Order Effects in Dichotomous Questions: The Impact of Administration Mode
- 89/18 Norbert Schwarz, Hans-J. Hippler, Elisabeth Noelle-Neumann, Thomas Münkler
Response Order Effects in Long Lists: Primacy, Recency, and Asymmetric Contrast Effects
- 89/19 Wolfgang Meyer
Umweltberichterstattung in der Bundesrepublik Deutschland
- 89/20 Michael Braun, Reiner Trometer
ALLBUS Bibliographie (8. Fassung, Stand: 30.6.1989)
- 89/21 Günter Rothe
Gewichtungen zur Anpassung an Statusvariablen. Eine Untersuchung am ALLBUS 1986
- 89/22 Norbert Schwarz, Thomas Münkler, Hans-J. Hippler
What determines a "Perspective"? Contrast Effects as a Function of the Dimension Tapped by Preceding Questions
- 89/23 Norbert Schwarz, Andreas Bayer
Variationen der Fragenreihenfolge als Instrument der Kausalitätsprüfung: Eine Untersuchung zur Neutralisationstheorie devianten Verhaltens
- 90/01 Norbert Schwarz, Andreas Bayer
Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis
- 90/02 Norbert Schwarz, Fritz Strack, Hans-J. Hippler, George Bishop
The Impact of Administration Mode on Response Effects in Survey Measurement
- 90/03 Norbert Schwarz, Herbert Bless, Gerd Bohner
Mood and Persuasion: Affective States Influence the Processing of Persuasive Communications
- 90/04 Michael Braun, Reiner Trometer
ALLBUS-Bibliographie 90
- 90/05 Norbert Schwarz, Fritz Strack
Context Effects in Attitude Surveys: Applying Cognitive Theory to Social Research
- 90/06 Norbert Schwarz, Herbert Bless, Fritz Strack, Gisela Klumpp, Annette Simons
Ease of Retrieval as Information: Another Look at the Availability Heuristic
- 90/07 Norbert Schwarz, Fritz Strack, Hans-J. Hippler
Kognitionspsychologie und Umfrageforschung: Themen und Befunde eines interdisziplinären Forschungsgebietes
- 90/08 Norbert Schwarz, Hans-J. Hippler
Response Alternatives: The Impact of their Choice and Presentation Order

- 90/09 Achim Koch
Externe Vergleichsdaten zum ALLBUS 1984, 1986, 1988
- 90/10 Norbert Schwarz, Bärbel Knäuper, Hans-J. Hippler, Elisabeth Noelle-Neumann, Leslie Clark
Rating Scales: Numeric Values May Change the Meaning of Scale Labels
- 91/01 Denis J. Hilton
Conversational Inference and Rational Judgment
- 91/02 Denis J. Hilton
A Conversational Model of Causal Explanation
- 91/03 Mood Effects on Interpersonal Preferences: Evidence for Motivated Processing Strategies
- 91/04 Joseph P. Forgas
Affective Influences on Interpersonal Perception
- 91/05 Norbert Schwarz, Herbert Bless
Constructing Reality and Its Alternatives: An Inclusion/Exclusion Model of Assimilation and Contrast Effects in Social Judgment
- 91/06 Herbert Bless, Roland F. Fellhauer, Gerd Bohner, Norbert Schwarz
Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben
- 91/07 Norbert Schwarz, Bärbel Knäuper, E. Tory Higgins
Der Einfluß von Rangordnungsaufgaben auf nachfolgende Denkprozesse: Zur Aktivierung prozeduraler Sets
- 91/08 Bettina Scheuring, Norbert Schwarz
Selbstberichtete Verhaltens- und Symptommhäufigkeiten: Was Befragte aus Antwortvorgaben des Fragebogens lernen
- 91/09 Norbert Schwarz, Herbert Bless
Scandals and the Public's Trust in Politicians: Assimilation and Contrast Effects
- 91/10 Rolf Porst
Ausfälle und Verweigerungen bei einer telefonischen Befragung
- 91/11 Uwe Blien, Heike Wirth, Michael Müller
Identification risk for microdata stemming from official statistics
- 91/12 Petra Beckmann
Methodological Report ISSP 1989
- 91/13 Martina Wasmer, Achim Koch, Michael Wiedenbeck
Methodenbericht zur "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 1990
- 91/14 Uwe Blien, Oded Löwenbein
Einkommensanalysen auf der Grundlage amtlicher Daten und Umfragedaten: Ergebnisse zur betrieblichen Seniorität und Arbeitslosigkeit
- 91/15 Petra Beckmann, Peter Mohler, Rolf Uher
ISSP Basic Information on the ISSP Data Collection 1985 -1994
- 91/16 Norbert Schwarz
In welcher Reihenfolge fragen? Kontexteffekte in standardisierten Befragungen
- 91/17 Ellen D. Riggle, Victor C. Ottati, Robert S. Wyer, Jr., James Kuklinski, Norbert Schwarz
Bases of Political Judgments: The Role of Stereotypic and Non-stereotypic Information

- 91/18 Dagmar Krebs
Was ist sozial erwünscht? Der Grad sozialer Erwünschtheit von Einstellungsisems
- 91/19 Michael Braun, Reiner Trometer
ALLBUS-Bibliographie
- 91/20 Michael Schneid
Einsatz computergestützter Befragungssysteme in der Bundesrepublik Deutschland
- 91/21 Rolf Porst, Michael Schneid
Software-Anforderungen an computergestützte Befragungssysteme
- 91/22 Ulrich Mueller
The Reproductive Success of the Elites in Germany, Great Britain, Japan and the USA during the 19th and 20th Century
- 92/01 P.H. Hartmann, B. Schimpl-Neimanns
Zur Repräsentativität sozio-demographischer Merkmale des ALLBUS - multivariate Analysen zum Mittelschichtbias der Umfrageforschung
- 92/02 Gerd Bohner, Kimberly Crow, Hans-Peter Erb, Norbert Schwarz
Affect and Persuasion: Mood Effects on the Processing of Message Content and Context Cues and on Subsequent Behavior
- 92/03 Herbert Bless, Gerd Bohner, Traudel Hild, Norbert Schwarz
Asking Difficult Questions: Task Complexity Increases the Impact of Response Alternatives
- 92/04 Wolfgang Bandilla, Siegfried Gabler, Michael Wiedenbeck
Methodenbericht zum DFG-Projekt ALLBUS Baseline-Studie
- 92/05 Frank Faulbaum
Von der Variablenanalyse zur Evaluation von Handlungs- und Prozeßzusammenhängen
- 92/06 Ingwer Borg
Überlegungen und Untersuchungen zur Messung der subjektiven Unsicherheit der Arbeitsstelle
- 92/07 Ingwer Borg, Michael Braun
Arbeitsethik und Arbeitsinvolvement als Moderatoren der psychologischen Auswirkungen von Arbeitsunsicherheit
- 92/08 Eleanor Singer, Hans-Jürgen Hippler, Norbert Schwarz
Confidentiality Assurances in Surveys: Reassurance or Threat?
- 92/09 Herbert Bless, Diane M. Mackie, Norbert Schwarz
Mood Effects on Attitude Judgments: The Independent Effects of Mood before and after Message Elaboration
- 92/10 Ulrich Mueller, Carola Schmid
Ehehäufigkeit und Fruchtbarkeit weiblicher Mitglieder der deutschen Elite
- 92/11 Herbert Bless, Fritz Strack, Norbert Schwarz
The Informative Functions of Research Procedures: Bias and the Logic of Conversation
- 92/12 Norbert Schwarz, Herbert Bless, Michaela Wänke
Subjective Assessment and Evaluation of Change: Lessons from Social Cognition Research
- 92/13 Norbert Schwarz, Hans-J. Hippler
Buffer Items: When Do They Buffer and When Don't They?

- 92/14 Hans-J. Hippler, Norbert Schwarz
The Impact of Administration Modes on Response Effects in Surveys
- 92/15 Michaela Wänke, Norbert Schwarz
Comparative Judgments: How the Direction of Comparison Determines the Answer
- 92/16 Michael Braun, Reiner Trometer
ALLBUS-Bibliographie (11. Fassung, Stand: 30.6.1992)
- 92/17 Anke Nau, Rolf Porst
Projektbericht ZUMA-Methodenpanel Teil 1: Konzeption, Vorbereitung und Durchführung
- 92/18 Michael Schneid
Handbuch ZUMA-Feldsteuerungsprogramm
- 92/19 Paul Lüttinger, Bernhard Schimpl-Neimanns
Amtliche Bildungsstatistik und empirische Sozialforschung
- 92/20 Rolf Porst, Michael Schneid
Fragebogenschreiben mit Microsoft WORD 5.0: Druckformate, Textbausteine und Makros zur effizienten Gestaltung von Fragebogen
- 92/21 Michael Braun, Jacqueline Scott, Duane F. Alwin
Economic Necessity or Self-Actualization? Attitudes toward Women's Labor-force Participation in the East and West
- 92/22 Duane F. Alwin, Michael Braun, Jacqueline Scott
The Separation of Work and the Family: Attitudes Towards Women's Labour-Force Participation in Germany, Great Britain, and the United States
- 92/23 Michaela Wänke, Norbert Schwarz, Herbert Bless
The Availability Heuristic Revisited: Experienced Ease of Retrieval in Mundane Frequency Estimates
- 93/01 Michael Braun, Carmen Eilinghoff, Siegfried Gabler, Michael Wiedenbeck
Methodenbericht zur Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1992
- 93/02 Michael Häder, Sabine Novossadeck
Einstellungsdynamik in Ostdeutschland. Auswertung von Wiederholungsuntersuchungen zur individuellen Bewältigung des sozialen Wandels
- 93/03 Norbert Schwarz
Judgment in a Social Context: Biases, Shortcomings, and the Logic of Conversation
- 93/04 Norbert Schwarz, Michaela Wänke, Herbert Bless
Subjective Assessments and Evaluations of Change: Some Lessons from Social Cognition Research
- 93/05 Norbert Schwarz
Context Effects in Attitude Measurement
- 93/06 Reiner Trometer
Die Operationalisierung des Klassenschemas nach Goldthorpe im ALLBUS
- 93/07 Carola Schmid
Der Zugang zu den Daten der Demographie
- 93/08 Michael Schneid
Fragebogenschreiben mit Word für Windows

- 93/09 Achim Koch
Die Nutzung demographischer Informationen in den Veröffentlichungen mit ALLBUS-Daten
- 93/10 Helmut Schröder
Über den Zusammenhang zwischen Aktivitäten und Zufriedenheit: "Eine kommunale Seniorenbefragung"
- 93/11 Michael Braun, Reiner Trometer
ALLBUS-Bibliographie, 12. Fassung, Stand: 30.9.93
- 93/12
- 93/13 Steven E. Finkel, Peter R. Schrott
Campaign Effects on Voter Choice in the German Election of 1990
- 93/14 Jürgen Hoffmeyer-Zlotnik, Dagmar Krebs
Subjektive Statuszuweisung; Objektive Schichtmessung
- 93/15 Dagmar Krebs
Richtungseffekte von Itemformulierungen
- 93/16 Dagmar Krebs
Social Desirability: The collective conscience? Judging the degree of social desirability in attitude items
- 93/17 Bernhard Krüger, Heiner Ritter, Cornelia Züll
SPSS Einsatz auf unterschiedlichen Plattformen in einem Netzwerk: Daten und Ergebnisaustausch
- 94/01 Jürgen H.P. Hoffmeyer-Zlotnik, Michael Wiedenbeck
Überlegungen zu Sampling, Qualitätsprüfung und Auswertung von Daten aus Teilpopulationen
- 94/02 Michael Häder, Sabine Häder
Die Grundlagen der Delphi-Methode: - Ein Literaturbericht -
- 94/03 Sabine Häder
Auswahlverfahren bei Telefonumfragen
- 94/04 Peter Prüfer, Margrit Rexroth
Ein Verfahren zur Erfassung von Erhebungsproblemen bei Interviews der Hauptstudie
- 94/05 Michael Häder, Sabine Häder
Ergebnisse einer Experimentellen-Studie zur Delphi-Methode
- 94/06 Bernhard Schimpl-Neimanns, Heike Wirth
Bestandsaufnahme und Nutzungsmöglichkeiten amtlicher Mikrodaten der DDR für Sekundäranalysen zur Bildungs- und Einkommensungleichheit
- 94/07 Norbert Schwarz, Hans-J. Hippler
Subsequent Questions May Influence Answers to Preceding Questions in Mail Surveys
- 94/08 Norbert Schwarz, Hans-J. Hippler
The Numeric Values of Rating Scales: A Comparison of their Impact in Mail Surveys and Telephone Interviews
- 94/09 Norbert Schwarz
Cognition, Communication, and Survey Measurement: Some Implications for Contingent Valuation Surveys
- 94/10 Michael Braun, Reiner Trometer
ALLBUS Bibliographie (13. Fassung, Stand 30.8.94)

- 94/11 Achim Koch, Siegfried Gabler, Michael Braun
Konzeption und Durchführung der allgemeinen
Bevölkerungsumfrage der Sozialwissenschaften
(ALLBUS) 1994
- 95/01 Michael Schneid
Fragebogenschreiben mit Microsoft Word für Windows
Version 6.0
- 95/02 Michael Schneid
Disk-By-Mail
Eine Alternative zur schriftlichen Befragung?
- 95/03 Michael Schneid
Einsatz computergestützter Befragungssysteme in
Südamerika, Nah-Ost, Asien, Afrika und Australien.
Eine Fax-Umfrage
- 95/04 Rolf Porst, Christa von Briel
Wären Sie vielleicht bereit, sich gegebenenfalls
noch einmal befragen zu lassen? Oder:
Die Gründe für die Teilnahme an Panel-Befragungen

