

X Die Gewichtung

Aufgrund von Fehlprognosen haben die Meinungsforschungs-Institute in den letzten Jahren zu einem ebenso riskanten wie dubiosen Mittel gegriffen; der „Kunst“ der Gewichtung. Es werden nicht die tatsächlichen Umfrageergebnisse veröffentlicht, sondern diese werden rein rechnerisch erhöht oder erniedrigt, und nur dieses modifizierte Resultat wird publiziert.²⁹

M. Kaase führte dazu im „Rheinischen Merkur“ Nr.37/1987 aus:

„Wegen Verzerrungsmöglichkeiten hat die kommerzielle Wahlforschung das Verfahren der „politischen Gewichtung“ erfunden. Bei diesem höchst umstrittenen und theoretisch völlig unfundierten Ansatz macht man sich paradoxerweise gerade die Abweichung der Befragtenangaben zu ihrer zurückliegenden Wahlentscheidung (der sogenannten Recall-Frage) vom ja zweifelsfrei bekannten Ergebnis einer zurückliegenden Vergleichswahl zunutze, um die aktuelle „Prognose“, die an sich auf den Angaben der Befragten zu ihrer Wahlabsicht beruhen müßte, zu korrigieren. Solche politisch gewichteten „Prognosen“ haben die Wirkung, diese mehr in die Nähe der traditionell zu erwartenden Wahlergebnisse zu rücken und verfälschen damit gerade die eigentlich interessante Nachricht : Wie nämlich zu einem bestimmten Zeitpunkt die Stimmung der Bevölkerung bezüglich der politischen Parteien aussieht.“

Ein typisches Beispiel für die Umverteilung von Prozenten, wie sie vor der Bundestagswahl praktiziert wurde, war in der Wirtschaftswoche Nr. 1/2 vom 2. Januar 1987 dargestellt:

„Behaupten etwa zu viele Befragte, sie hätten beim letztenmal sozialdemokratisch gewählt, so wird das aktuelle Umfrageergebnis für die SPD entsprechend heruntergewichtet. Liegt der Erinnerungswert für die FDP zu niedrig, werden die Liberalen entsprechend hochgewichtet: Deren aktueller Recall-Zuschlag liegt bei rund 100 Prozent.“

Es wird oft behauptet, die erfragten Werte für die SPD und die Grünen müßten erniedrigt und diejenigen für die FDP erhöht werden. Als Gründe dafür werden angeführt, daß ein Teil der SPD- und Grüne-Wähler „erfahrungsgemäß“ bis zum Wahlsonntag „ihre Meinung ändern“ oder daß diese auf Grund von systematischen Verzerrungen in der Repräsentativumfrage überproportional vertreten seien. Von der FDP wird behauptet, sie würde „erfahrungsgemäß“ am Wahltag Zweitstimmen von CDU/CSU-Wählern erhalten und sie sei in Repräsentativumfragen systematisch untervertreten.

Falls es sich um empirisch gesicherte Erkenntnisse und nicht um plausible Stammtisch-Argumente handelt, können solche Modifikationen sinnvoll und nützlich sein, soweit sie der Korrektur systematischer Fehler dienen. Aber es ist absurd, auslosungsbedingte Abweichungen, die einen großen Spielraum unvermeidlich machen (vergleiche Abschnitte IV bis IX), durch Gewichtung kompensieren zu wollen. Dies wird aber gerade in der Praxis routinemäßig gemacht.

Die Größenordnung der Umgewichtung ist atemraubend. Nach Noelle-Neumann werden die

²⁹ Die Methode der Gewichtung von Umfrageergebnissen ist an sich fast so alt wie die "moderne" Demoskopie (vergleiche : The Pre-Election Polls of 1948, Bulletin Nr. 60 des Social Research Council Inc., New York 1949). Neu ist die Art und das Ausmaß der Gewichtung.

tatsächlich ermittelten Ergebnisse bis zu **10% oder 11%** abgeändert.^{30,31} Wie vorhin erwähnt, betrug im Januar 1987 der „aktuelle“ Recall-Zuschlag für die FDP rund 100%, d.h. das Umfrageergebnis wurde verdoppelt. In den Graphiken auf den Seiten X/5 bis X/7 ist die von der Forschungsgruppe Wahlen im Politbarometer des ZDF vorgenommene Umgewichtung in den Jahren 1986/87 detailliert dargestellt.

Die Meinungsforschungs-Institute schaffen sich durch die Gewichtung einen Freiraum, von dessen Existenz und Ausmaß die Öffentlichkeit keine Ahnung hat.

Ein seriöses Institut sollte auf das Gewichtungsdilemma und die Interview-Fehler hinweisen und wenigstens den auslosungsbedingten Spielraum der Umfragergebnisse angeben. Es sollte die Rohdaten offenlegen und nicht nur Zahlen, die es daraus durch eine geheimgehaltene Kosmetik gewinnt.

Das geschieht jedoch nicht. Die einzige Ausnahme ist meines Wissens die Forschungsgruppe Wahlen - d.h. das Politbarometer im Zweiten Deutschen Fernsehen - wo die Existenz und das Ausmaß der Abänderung von Umfrageergebnissen noch in Umrissen erkennbar ist. Die Forschungsgruppe Wahlen hatte sich stets gegen die Gewichtung von Umfragergebnissen gewehrt. Anfang 1986 wurde sie durch politischen Druck gezwungen, diese dubiosen Praktiken ebenfalls einzuführen und zu kaschieren. Die Details dieser Aktion und die Begründung für diese angebliche „Erweiterung“ der Berichterstattung sind in der Box auf Seite X/8 zu lesen.

Für die folgenden Ausführungen halte man sich die Tabellen auf den Seiten X/5 bis X/7 vor Augen, in welchen die Umgewichtungen graphisch dargestellt sind. Die Einführung der Gewichtung beim Politbarometer im März 1986 führte zu einer Erhöhung der Prozentzahlen für die Union bzw. die FDP um durchschnittlich etwa 1,5% bzw. 3% im Jahre 1986. Die Umfrageergebnisse für die SPD bzw. die Grünen wurden um durchschnittlich etwa 3% bzw. 2% nach unten korrigiert. Bei diesen Zahlen handelt es sich um Durchschnittswerte, die Bandbreite der Umgewichtung ist beträchtlich. Die Gewichtung zeigte bei den großen Parteien einen deutlichen zeitlichen Trend. Anfang des Jahres wurde die SPD massiv beschnitten, während die Union ein kräftiges „face lifting“ erhielt: im März -9% für die SPD und +6% für die Union. Im Sommer - Juni bis September - reduzierte sich der Aderlaß der SPD auf 2% bis 4%, während die CDU/CSU zunehmend „selbständiger“ wurde. Im Oktober stand die Union auf eigenen Füßen, und die SPD mußte nur noch die bescheidene Abgabe von einem Prozent entrichten. Im Oktober/November erfolgte dann eine eigentliche Kehrtwendung bei der Gewichtung : Man erinnere sich an die damalige Situation. Die November-Umfrageergebnisse des Politbarometers, welche Anfang Dezember die Wahlkampfarena erreichten, ergaben 49% für die Union, 37% für die SPD, 4% für die FDP und 10% für die Grünen. Damit wäre die FDP wie üblich - d.h. wie bei den meisten Politbarometerumfragen 1986 - an der 5%-Hürde gescheitert und die Union hätte folglich mit 49% die absolute Mehrheit errungen. Bekanntlich erhitzte diese Möglichkeit die Gemüter erheblich. Die CDU/CSU lief Sturm gegen das „Gerede von der absoluten Mehrheit“, gleichzeitig aber führte sie eine vehemente Zweitstimmenkampagne gegen die FDP mit der

³⁰ Man vergleiche dazu die Seiten I/4 und I/5.

³¹ Wirtschaftswoche Nr. 1/2 vom 2. Januar 1987:

"Solch gewaltige Veränderungen sind selbst für die Nestorin des Gewerbes stets wieder ein Ritt über den Bodensee. Elisabeth Noelle-Neumann bekennt: "Ich bin jedesmal wieder von großer Sorge erfüllt." Mit welchen Methoden der Justierung sie die groben Recallgewichtungen noch verfeinert, weiß dabei selbst in Allensbach nur ein kleiner Zirkel von Eingeweihten. Auch die anderen Institute verbergen die Details ihrer Computerprogramme, in die neben den Recallgewichten noch Faktoren wie Veränderungen der Bevölkerungsstruktur oder des Stimmungsklimas und der politischen Prioritäten eingehen."

Begründung, die FDP benötige keine Leihstimmen der Union, denn sie sei sicher über der 5%-Hürde.

In dieser Situation änderte das Politbarometer seine bisherige Gewichtung-Praxis bei den großen Parteien. Zum ersten Mal wurden Prozente von der CDU/CSU zur SPD hin verschoben; der Anteil der Union wurde von 49% auf 47% reduziert, während die SPD eine Mutspritze von 37% auf 38% erhielt. Der Anteil der FDP wurde von 6% im Vormonat auf 7% erhöht, obwohl das zugrundeliegende Umfrageergebnis unverändert bei 4% lag. Durch diese Änderung der bisherigen Gewichtung-Praxis lieferte das Politbarometer - was zweifellos unbeabsichtigt war - den „wissenschaftlichen Beweis“ für die Wahlkampfbehauptungen der Union.

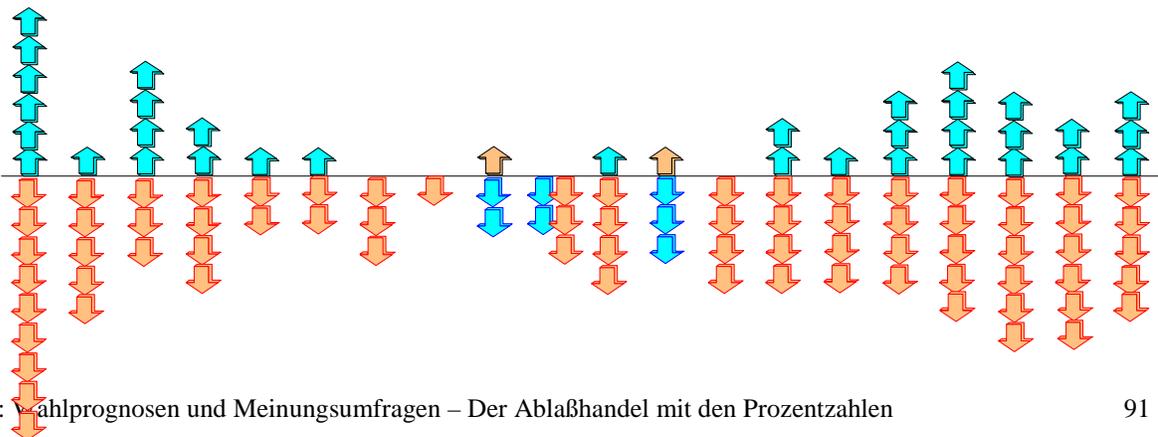
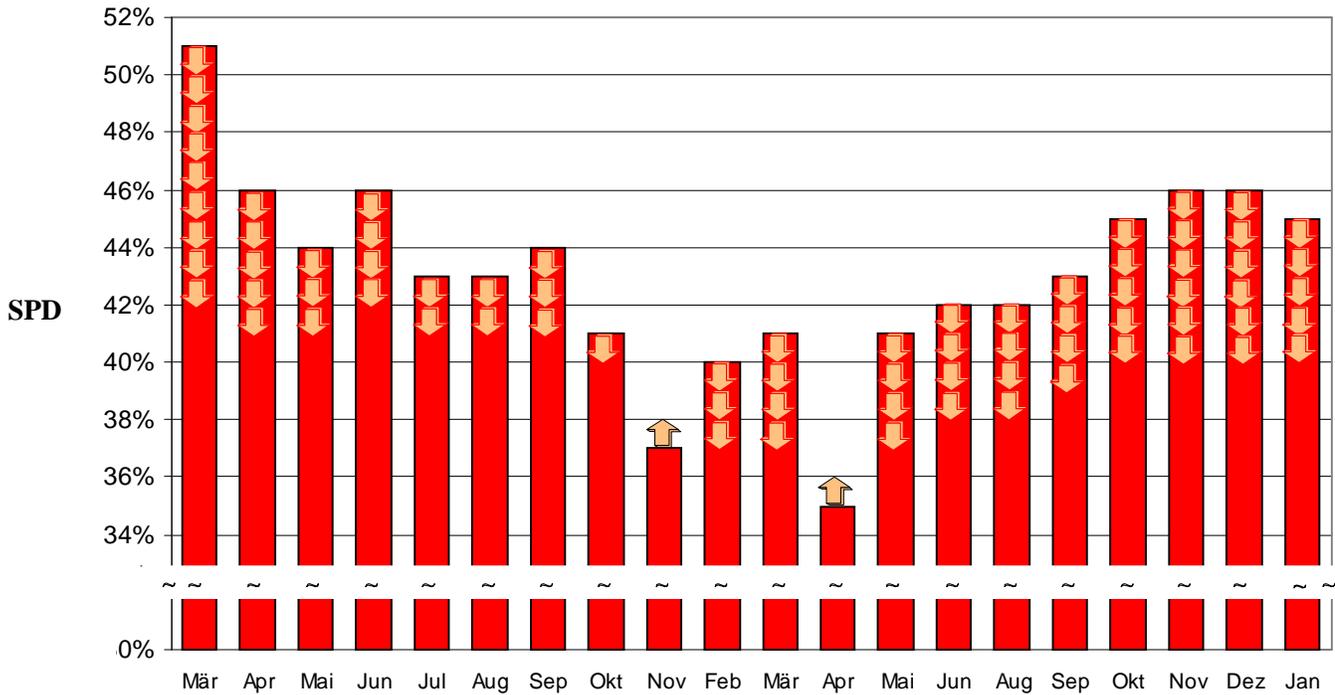
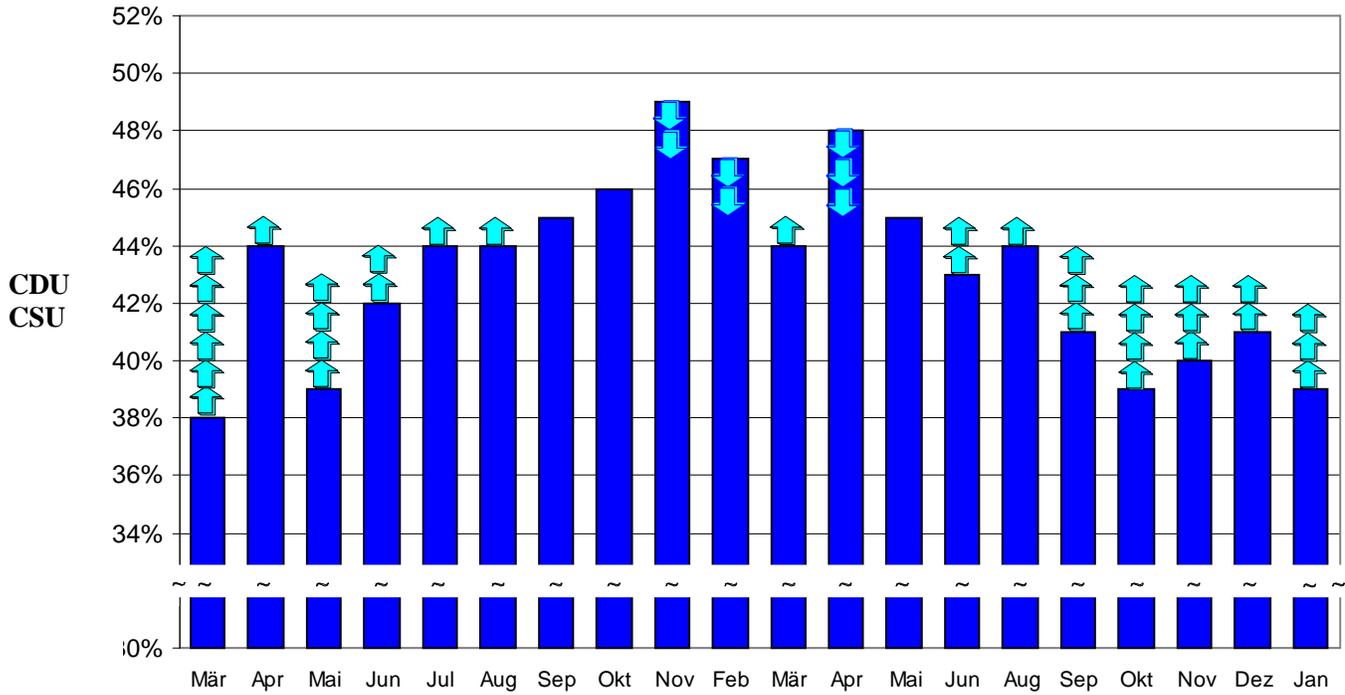
Das Verschweigen der Gewichtungen kann handfeste politische Auswirkungen haben. Ist es zum Beispiel zu verantworten, der FDP Monat für Monat 6% bis 7% auf den Bildschirm zu projizieren, obwohl ihr bei den laufenden Umfragen ermittelter Zweitstimmenanteil deutlich unter 5% liegt?

Mit diesen Zahlen wird doch dem Wähler suggeriert, die FDP sei sicher über der 5%-Hürde. Wer garantiert aber, daß die Wähler am Wahlsonntag für die „aus Erfahrung“ hinzugerechneten 2% tatsächlich aufkommen? Etwa die Gewichtungs-Zauberlehrlinge?

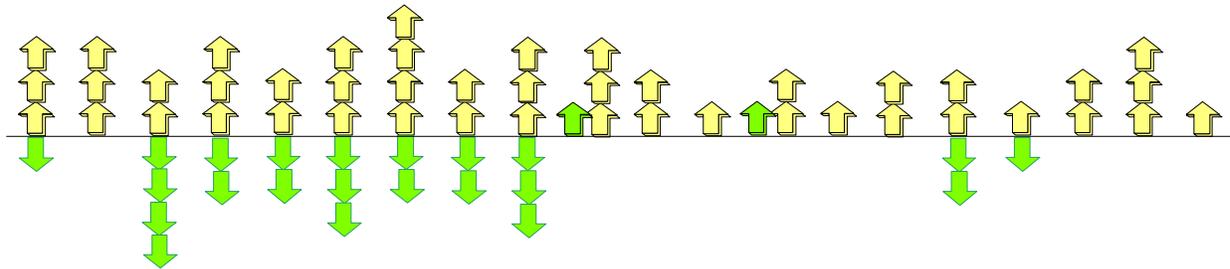
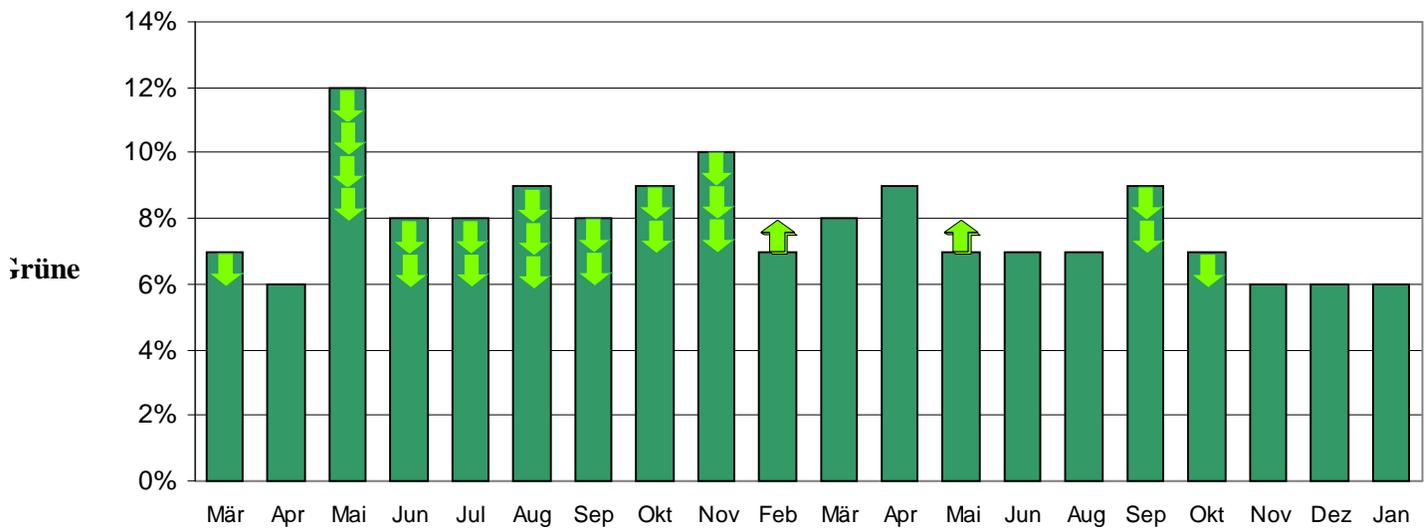
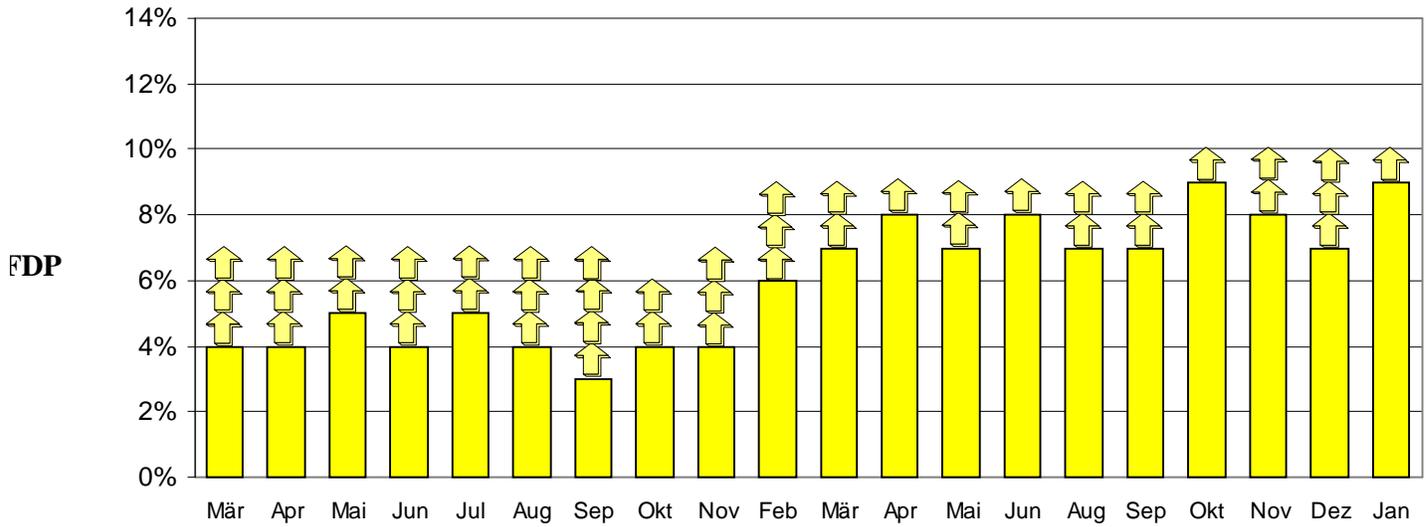
Es mag sein, daß die Stärke der FDP und der CDU/CSU mit Umfrageergebnissen in der Vergangenheit nicht selten unterschätzt worden sind. Ein Wahlforscher hat das Recht, wahrscheinlich sogar die Pflicht, darauf hinzuweisen.

Aber hat er das Recht, der Öffentlichkeit die Unzulänglichkeiten seiner Meßmethoden mit gefälschten Zahlen zu verheimlichen?

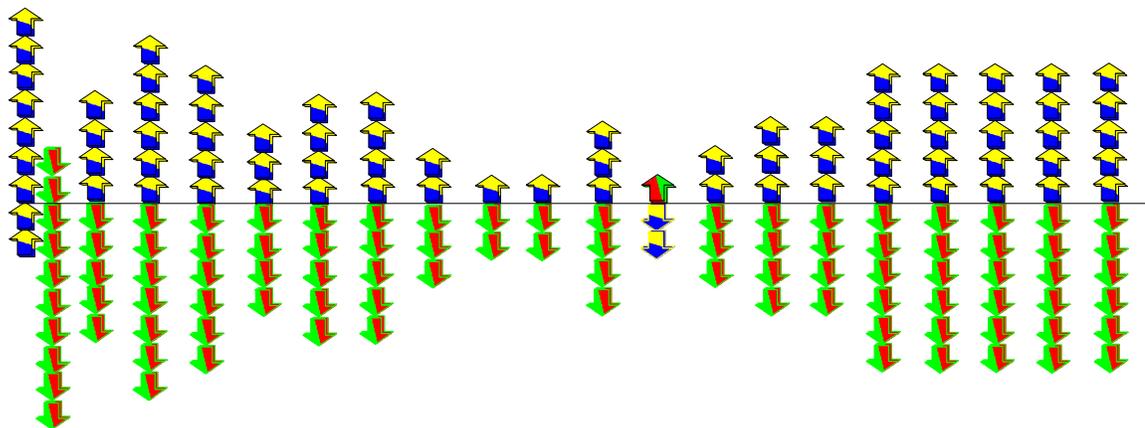
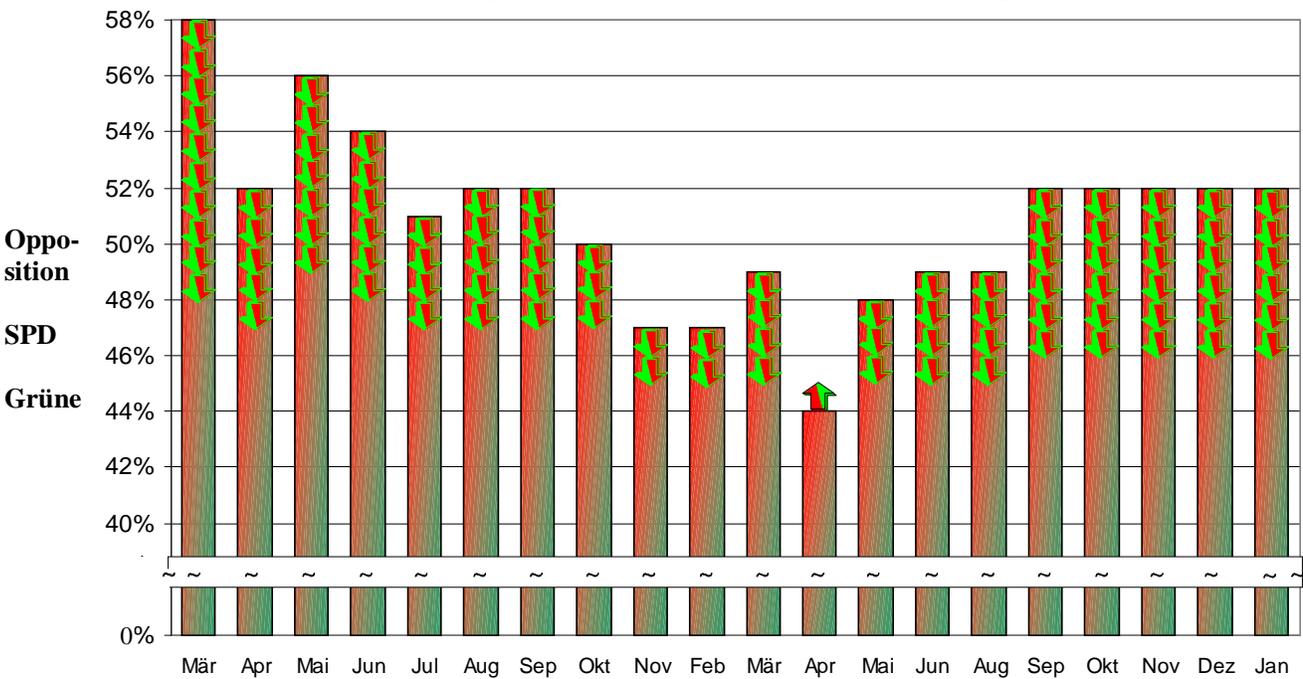
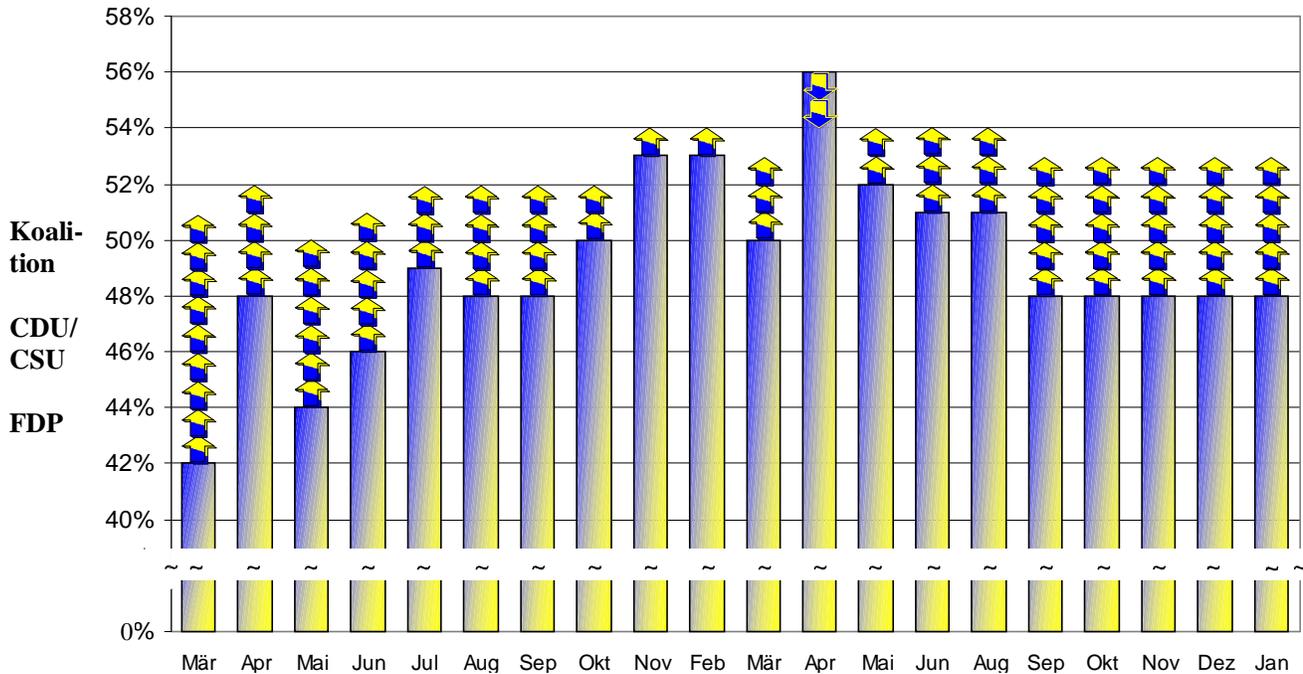
Politbarometer März 1986 bis Januar 1988



Politbarometer März 1986 bis Januar 1988



Politbarometer März 1986 bis Januar 1988



Wie das Politbarometer zur Gewichtung verdonnert wurde

Erwin K. Scheuch, Rheinischer Merkur Nr. 37/1987 (Merkur Extra):

„Wie wichtig Parteien die Veröffentlichung solcher Wahlvoraussagen nehmen, wird an einer Kontroverse über das sogenannte Politbarometer des ZDF deutlich. Diese Sendung berichtet nicht nur über die Meinung der Wähler, sondern pflegt als prominenteste Mitteilung das Ergebnis einer Frage nach der Parteisympathie vorzustellen. Daneben stellen Meinungsforschungsinstitute, die auch im Fernsehen zu Wort kamen und kommen, die sogenannte Sonntagsfrage, die wissen will, was er am nächsten Sonntag täte, gäbe es dann etwas zu wählen. Die Ergebnisse der Sonntagsfrage und der Sympathiefrage zeigten immer eine erhebliche Abweichung. Das kann nicht verwundern, weil eben verschiedene Dinge erfragt wurden. Aber nicht nur das Fragen war verschieden, sondern auch das Behandeln der Antworten, bevor sie publiziert wurden. Kein Umfrageinstitut veröffentlicht nämlich solche Antworten, ohne sie zu verändern, zu gewichten. Damit sollen Fehler ausgeglichen werden, die sich aus Antwortverweigerungen und anderen Störfaktoren ergeben können. Beim Politbarometer wurde das anders gewichtet als bei den übrigen Umfrageinstituten mit ihrer Sonntagsfrage, was an sich nur Fachleute zu interessieren braucht. Ich erlaubte mir in einem Aufsatz darauf zu verweisen, daß bei der Sympathiefrage die CDU 1986 mehrere Prozentpunkte –meist um drei bis vier – unter den Werten der anderen Institute lag, die SPD aber wesentlich stärker erschien als bei den Berichten über die Sonntagsfrage. „Zuschauer mißverstehen das Barometer aber als Aussage über die Wahlchancen“, überschrieb ich meinen Beitrag.

Das las die CDU als eine Bestätigung ihrer Vermutung, das Politbarometer betreibe mit seinen Zahlen Stimmungsmache zu Gunsten der SPD, und setzte entsprechend beim ZDF durch, daß in Zukunft die Forschungsgruppe Wahlen neben den Ergebnissen für die Sympathiefrage auch die Ergebnisse für die Sonntagsfrage vorzustellen habe. Deren bis dahin im ZDF veröffentlichten Ergebnisse sahen bei der Forschungsgruppe Wahlen nämlich auch nicht anders aus als bei den anderen Forschungsinstituten. Wie wichtig aber auch die SPD die Suggestivwirkung guter Ergebnisse für sich selber nahm, zeigt sich an den wütenden Angriffen im „Vorwärts“ auf diese Erweiterung der Berichterstattung.“

Wie das Politbarometer zur Gewichtung verdonnert wurde (Fortsetzung)

Kommentar: Beim Lesen des obigen Artikels von E.K. Scheuch entsteht der Eindruck, die Sympathiefrage und die Sonntagsfrage seien zwei völlig verschiedene Dinge, welche beide auf Umfrageergebnissen beruhen, und das Politbarometer habe vor März 1986 nur die Resultate der Sympathiefrage auf den Bildschirm gebracht, nicht aber diejenigen der Sonntagsfrage.

Das Politbarometer stellt die Sympathiefrage und die Sonntagsfrage. Nach den mir vorliegenden Unterlagen über 21 Monate von Januar 1986 bis November 1987 sowie September/Okttober 1985 – es gelang mir nicht, die Unterlagen für die Monate November und Dezember 1985 zu erhalten – stimmen die Antworten auf die beiden Fragen erstaunlich gut überein. Die Abweichungen liegen für die vier Parteien meist unter einem Prozent. Für jede Partei sind die Abweichungen sowohl positiv als auch negativ und sie heben sich im Durchschnitt praktisch auf.

Wenn das Politbarometer seine Berichterstattung wie gefordert erweitert hätte, dann hätte es meist zweimal die gleichen bzw. nur unwesentlich verschiedene Zahlen auf den Bildschirm bringen müssen. Dies war nicht der Sinn der Aktion. Das Politbarometer sollte die tatsächlichen Ergebnisse umgewichten (d.h. verfälschen), wie dies in der Prognoseindustrie üblich ist. **Der Vorschlag, die Meßergebnisse der Union um 3% bis 4% zu erhöhen, wurde durch die Ereignisse 1986/87 ad absurdum geführt.** Dies hätte vom September 1986 bis Mai 1987 meist

zu einer absoluten Mehrheit für die CDU/CSU geführt. Von November 1986 bis April 1987 wurden die Umfrageergebnisse für die CDU/CSU um durchschnittlich 2% heruntergewichtet, was das Schreckgespenst einer absoluten Mehrheit vom Fenster räumte.

Wie aus den Graphiken auf den vorangehenden Seiten hervorgeht, ergab die Gewichtung 1986/87 keinen erkennbaren SINN außer dem der Glättung **zwecks Kompensation auslosungsbedingter Abweichungen** und der Vortäuschung eines kontinuierlichen zeitlichen Trends. Die Gewichtung stellte sicher, daß die CDU/CSU nie „zu wenige“, aber auch nie „zu viele“ Prozente bekam. Zusammen mit der FDP wurde eine stetige Mehrheit von 50 bis 54% sichergestellt, egal wie die tatsächlichen Umfrageergebnisse lauteten.

Das Politbarometer zu zwingen, seine Umfrageergebnisse so zu fälschen, wie es in der Prognoseindustrie üblich ist, betrachte ich nicht als eine „sinnvolle Erweiterung“ der Berichterstattung. Eine solche würde allenfalls darin bestehen, daß das Politbarometer die Chance seiner Prognose – d.h. den auslosungsbedingten Spielraum – bekannt gibt, wie es auf Seite VII/9 illustriert wird. Dann würde sich das Problem der Gewichtung von selbst lösen ...

Bei Abänderungen der Rohergebnisse in den dargelegten Dimensionen stellt sich die Frage, weshalb überhaupt noch Meinungsumfragen für Wahlprognosen durchgeführt werden. Weshalb legt das Meinungsforschungsinstitut xy nicht einfach die Hand für einen Unionsanteil von x% ins Feuer? Seit den fünfziger Jahren schwankte die CDU/CSU bei den Bundestagswahlen bekanntlich zwischen etwa 45% und 50%, und dazwischen liegen nicht allzu viele Prozentzahlen, die als Prognosewert in Frage kommen Wie in Abschnitt VII gezeigt wurde, ist der auslosungsbedingte Spielraum größer.

Daß es auch „richtige“ Prognosen gibt, liegt zu einem guten Teil an der politischen Stabilität der Bundesrepublik und an den statistisch unvermeidlichen „Zufallstreffern“. Wenn die politische Landschaft in Bewegung geriete, würden den Wahlforschern die Felle sehr bald für jedermann sichtbar davonschwimmen. Einen Vorgeschmack für dieses Szenario lieferten etwa die Landtagswahlen in Bremen im September 1987: Die Form der im Zweiten Deutschen Fernsehen präsentierten Prognose,

CDU unter 30%

drückt so grotesk wie nur möglich das Eingeständnis aus, daß in einer *neuen* Situation nur nichtssagende Prognosen „*sicher*“ sind. Man traut den eigenen Daten nicht mehr, weil sie den bisherigen Rahmen sprengen. Das Wahlergebnis - 23,4% für die CDU, ein Verlust von 10% gegenüber 1983 - bestätigte die dumpfen Ahnungen der Forschungsgruppe Wahlen.

Bei großen Veränderungen - z.B. $\pm 10\%$ bis $\pm 20\%$ - werden die Karten neu verteilt. Dann müßten völlig unbekannte Größen gemessen werden, für deren Abschätzung es keine Präzedenzfälle gibt. Da reicht es nicht mehr aus, vom letzten Wahlergebnis Pi mal Daumen ein paar Prozente abzuziehen oder zu addieren. Nun muß der Wahlforscher eine Aussage machen, von der er weiß, daß er sich damit wahrscheinlich in die Nesseln setzen wird³².

In einer solchen Situation werden wir Zeuge einer bisher in der Zoologie nicht beobachteten Metamorphose: Unsere Politwetterfrösche werden aufgeschreckt aus ihrem seichten Teich hüpfen und als ratlos schnatternde Enten über dem Sumpf der Demoskopie flattern. Jeder fühlt sich allein - und von seinen Daten im Stich gelassen. Die Statistik hilft ihm nicht, sich aus der Schlinge zu

³² Bei einer normalen Umfrage im kommerziellen Alltagsgeschäft spürt der Meinungsforscher nicht die Faust der bevorstehenden Wahl im Nacken. Da ist er frei, seine Daten sprechen zu lassen, was immer sie auch sagen mögen....

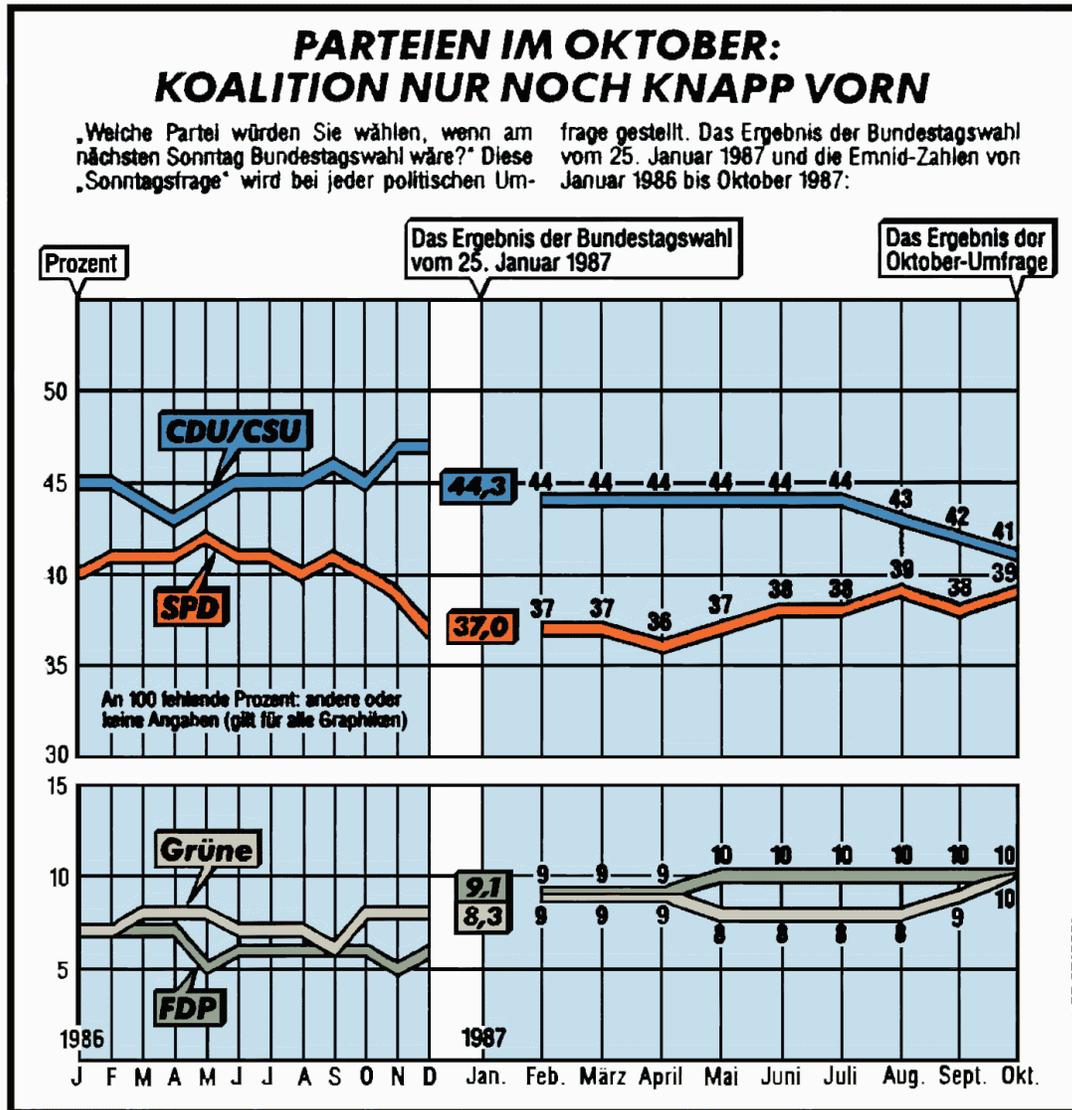
ziehen, in die er sich durch unhaltbare Versprechen verstrickt hat. Denn es treten - wie in den Abschnitten IV bis IX gezeigt wurde - als Folge der Auslosung des repräsentativen Querschnittes Abweichungen in einer Größenordnung auf, die der Meinungsforscher vor der Öffentlichkeit nicht eingestehen kann und will. So bleibt ihm als einziger Orientierungspunkt das Spektrum der „Prognosen“ der Konkurrenz. Die treibende Kraft ist die Angst, möglicherweise als einziger völlig daneben zu liegen. Zum Glück gibt es da die Gewichtung, mit deren Hilfe es möglich wird, die eigene Prognose im Mittelfeld der „Gerüchtebörse“, d.h. der „Konkurrenz“-Prognosen, anzusiedeln³³. Wenn die Sache schiefgeht, dann war man wenigstens nicht allein, was für das umsatzmäßig dominante und daher langfristig lebenswichtige kommerzielle Alltagsgeschäft von Bedeutung ist.

Eine Konsequenz dieser Angleichung (d.h. Koppelung) von Umfrageergebnissen - die manchmal auch bei normalen Änderungen des politischen Kräfteverhältnisses zu beobachten ist - besteht darin, daß die Fehler in der gleichen Richtung auftreten. Beispielsweise schätzen „alle“ Institute die CDU/CSU zu hoch oder „alle“ zu tief ein. Wenn der Zufall zu seinem Recht kommen würde, dann wären Fehler dieser Art extrem unwahrscheinlich.

Auch beim zeitlichen Verlauf der Parteistärken wird der Zufall vergewaltigt. Wenn z.B. im „Spiegel“ Nr. 45/1987 eine Graphik von Emnid abgedruckt ist, in welcher der CDU/CSU von Februar bis Juli 1987 konstant 44% zugeschrieben wird - und ebenso der FDP konstant 10% von Mai bis Oktober 1987 - dann handelt es sich um getürkte (d.h. geglättete bzw. gewichtete) Zahlen. Der Zufall mag ja seine Launen haben, aber so etwas schafft er nicht! (siehe folgende Seite)

³³ Eine neue Anwendung der Schweigespirale für Frau Noelle-Neumann?

Spieglein, Spieglein an der Wand,...
 ist Augstein auch der Sauberste im ganzen Land?
 Spiegel Märkte



Am Hofe König Rudolfs wird die Demoskopie seit langem als Kunst gepflegt: In monatlichen fünf- bis zehseitigen Spieglein-Berichten werden die Untertanen über die Lage im Reiche gründlich informiert. Objektiv! Da kodifiziert nicht einer der unzähligen Schreiber die vorherrschende Meinung a Hofe – nein, all die schönen Geschichtlein werden mit Zahlen belegt! Die Moral fließt förmlich aus den Zahlen, und kein verantwortungsbewußter Zahlengläubiger kann sich ihrem Sog entziehen. Von auslosungsbedingtem Spielraum und von Chance kann man von offizieller Seite in König Rudolfs Reich nichts erfahren. Nicht etwa, weil dann viele der schönen Demagogie-Geschichtlein wie Autos mit geplatzen Reifen davonholpern würden – nein, *Ehrenwort*, davon hat man am Hofe noch nie etwas gehört. Im Jahre *Barschels* und zu Ehren *Waldheims* haben wir deshalb ganz legal eine gutmütige Wanze beim Hof-Demagogen EMNID von König Rudolf installiert. Hören wir, was sie Pfeiffert:

Dies bedeutet in statistischem Sinne mit 90 % Wahrscheinlichkeit ($t = 1,64$), dass das tatsächliche Wahlergebnis für die einzelnen Parteien innerhalb folgender Fehlertoleranzgrenzen liegt:

CDU/CSU	= 42,3 - 49,2 %
SPD	= 32,9 - 41,1 %
F.D.P.	= 5,8 - 10,2 %
Grüne	= 5,8 - 10,2 %

EMNID - Institut GmbH & Co. KG

Tacke

F. Tacke

Urkundenrolle Nr. 5 / Jahrgang 1987

B e s c h e i n i g u n g

Hiermit wird der Emnid-Institut GmbH & Co.KG, Bielefeld, bescheinigt, daß die vorstehende Wahlprognose 1987 (Bundestagswahl am 25. Januar 1987) vom 24. Januar 1987 zu notariellem Protokoll - UR Nr. 1/87 - am 24. Januar 1987 in meine Verwahrung gegeben worden ist. Den versiegelten Umschlag habe ich heute, am 26. Januar 1987 geöffnet.

Bielefeld, 26. Januar 1987



H. Vitor
W. V. V.

XI Schlußfolgerungen

Die statistischen Grundlagen von Wahlprognosen und Meinungsumfragen basieren u.a. auf dem sogenannten schwachen Gesetz der großen Zahlen, vor allem aber auf dem unersättlichen Bedarf der zahlgläubigen Demoskopie-Endverbraucher.

Als Meßinstrument dient den Meinungsforschungs-Instituten eine *Elle*, nämlich die „*schwach repräsentativen*“ Tausender-Querschnitte, die mit einem nicht gerade stubenreinen Zufallsverfahren erstellt werden. Sie suggerieren aber durch die Präsentation ihrer Resultate, sie könnten damit *Millimeter-Bruchteile* messen, als hätten sie eine *Schieblehre*.

Nach den Gesetzen der Statistik ist jedes Meßresultat auch ein potentielles Miß-Resultat. Weil

- 1) die Ausfälle und Verzerrungen unkontrollierbar sind,
- 2) die Elle gegenüber den Schwächen der Interviewer und des Fragebogens immun ist,
- 3) bei der Zufallsauswahl aus Kostengründen erhebliche Abstriche gemacht werden,
- 4) kein Demoskop das Ausmaß des auslosungsbedingten Spielraums eingestehen will und kann,

wird aus Furcht vor dem Mißresultat der untaugliche Versuch gemacht, das Meßresultat mit dem *Gewichtungs-Vorschlaghammer* ins Lot zu rücken. Wen wundert es dann, wenn *diese „Hau den Lukas“*- Prognosen nicht immer auf dieser Welt landen?

Wenn ein Wahlforscher oder ein Meinungsforscher eine Prozentzahl nennt - mit oder ohne Nachkommastelle - oder eine solche dem Auftraggeber zum freien öffentlichen Gebrauch überläßt, dann ist dies unlauterer Wettbewerb, wahrscheinlich sogar Betrug. Er weiß doch ganz genau, daß er mit seinen Methoden grundsätzlich nicht in der Lage ist - auch unter idealen Voraussetzungen nicht -, eine gesicherte Prozentzahl zu ermitteln. Er hat Bereiche für die auslosungsbedingten Abweichungen seiner Ergebnisse und die zugehörige Chance anzugeben und auf die anderen Fehlerquellen hinzuweisen. Das Meßverfahren ist ausreichend zu beschreiben, damit eine Beurteilung der Resultate möglich wird.

Die von der Gewichtung und dem Mißbrauch des Begriffes „repräsentativer Querschnitt“ ausgehende Gefahr für die Öffentlichkeit liegt nicht so sehr in der Manipulation, denn die bevorstehenden Wahlen und die Rückwirkung von Fehlprognosen auf das umsatzmäßig dominante kommerzielle Alltagsgeschäft wirken ausreichend als Korrektiv, oft sogar schon als Über-Korrektiv.

Die eigentliche Gefahr besteht in der nicht beabsichtigten Irreführung durch Falsch-Informationen als Folge einer mißglückten Gewichtung³⁴ und des Verschweigens der unvermeidlichen Fehler, die durch das Interview und die Auslosung des repräsentativen Querschnittes verursacht werden. Man gibt vor, etwas wissenschaftlich gemessen zu haben, während man in Wirklichkeit nur spekuliert hat.

³⁴ So meinte der Geschäftsführer Jung des Meinungsforschungsinstitutes Basis Research GmbH gegenüber der Wirtschaftswoche (Nr.1/2 vom 2.Januar 1987), daß die Gewichtung in letzter Zeit versagt habe: "Die Rohzahlen waren oft genauer als die gewichteten Zahlen."

Besondere Bedeutung kommt der Tatsache zu, daß durch die Gewichtung der Umfrageergebnisse die unabhängige Berichterstattung der verschiedenen Meinungsforschungsinstitute zerstört wird. Welches Institut wagt es schon, auf die Dauer Umfrageergebnisse herauszugeben, die sich wesentlich vom Gros der anderen Institute unterscheiden? Die *Rohergebnisse* sind und bleiben zwar „*unabhängige Zeugen*“, aber durch den Konformitätsdruck werden diese Zeugen mit der Gewichtung oft mundtot gemacht. Die Ausschaltung der unabhängigen Berichterstattung hinterläßt keine Spuren, weil das Ausmaß der Abänderung verschwiegen wird und damit die Symptome des Angleichungsprozesses unterdrückt werden. Meines Erachtens gehen die Kollektiv-Fehlprognosen zum Teil auf diesen „*Meinungsklimadruck*“ unter den Instituten zurück. *Die Rohdaten möglichst vieler Institute über einen längeren Zeitraum stellen die bestmögliche Information dar.*

Die für die Öffentlichkeit bestimmten Wahlprognosen werden meist mit besonderem Aufwand betrieben, weil sie durch die Wahl überprüft werden und der Ruf des Instituts dabei auf dem Spiel steht. Um beim gegenwärtigen Kräfteverhältnis in der Bundesrepublik politisch interessante Aussagen betreffend der Stärken der vier relevanten Parteien machen zu können, ist eine Genauigkeit erforderlich, die die Möglichkeiten der Demoskopie weit übersteigt. Diese Genauigkeit läßt sich, wie in diesem Artikel dargelegt wurde, auch mit erhöhten Aufwand nicht erreichen, denn der erforderliche Stichprobenumfang ist praktisch nicht realisierbar und der Interviewfehler wird dadurch nicht aus der Welt geschafft. Wegen des auslosungsbedingten Spielraums vermögen Meinungsumfragen kaum mehr an gesicherter Information zu liefern, als was man aufgrund der relativen politischen Stabilität in der Bundesrepublik ohnehin schon weiß. Graphische Darstellungen des zeitlichen Verlaufs von Parteistärken, Popularitätswerten usw. sollte man im allgemeinen als Zeichentricksfilme oder „*Comic strips*“ betrachten.

Der Unterschied zwischen den Meinungsforschungsinstituten in der Wahlforschung besteht hauptsächlich darin, daß die einen mit mehr Glück und einem etwas besseren Riecher als die anderen spekulieren, wobei nicht zu bestreiten ist, daß Frau Noelle-Neumann der Konkurrenz manchmal um eine Nasenbreite voraus ist. Aber Spekulation bleibt es allemal.

Entscheidend für Erfolg bzw. Mißerfolg von Prognosen sind die politische Stabilität, der gute Riecher bzw. etwas Glück bei der Gewichtung und die Launen des Zufalls. Mit Wissenschaft hat das Ganze nichts zu tun, die „*Wissenschaft*“ dient lediglich als Aushängeschild.

Bei kommerziellen Umfragen schlägt keine Stunde der Wahrheit, gibt es keine Wahl. Wehrlos ist ein Auftraggeber allerdings nicht. Er kann als Teil seines Auftrages verlangen, daß Wiederholungen der Befragung simuliert werden, wie dies auf S. III/16-19 dargestellt wurde. Bei den handelsüblichen Meinungsumfragen mit einigen hundert bis 2000 Interviews werden dann viele der oft zu Hunderten abgelieferten, angeblich so harten Zahlen wie Butter unter der Simulations-Sonne schmelzen. Dann werden die durch die Auslosung des „repräsentativen Querschnittes“ verursachten Abweichungen in einer Weise zu Tage treten, welche die in den Abschnitten VI bis IX angeführten Spielräume als vergleichsweise „*mustergültig*“ erscheinen lassen. **Je mehr Fragen in einer „Repräsentativumfrage“ gestellt werden und je umfangreicher die Antwortmöglichkeiten sind, desto mehr weicht der „repräsentative Querschnitt“ von einem Miniaturbild ab.** Nicht nur weil die Länge des Fragebogens bei den Interviewern und Befragten Erschöpfungssymptome auslöst, sondern weil die statistische Qualität ausgehöhlt wird. Aufgrund vorläufiger Resultate ist mit einer Vergrößerung der Konfidenzintervalle (= lotteriebdingte Spielräume) um bis zu einem Faktor zwei und mehr zu

rechnen. Dies wird demnächst ausführlich behandelt werden.

Als Auftraggeber würde ich kommerzielle Umfragen grundsätzlich bei zwei verschiedenen Instituten in Auftrag geben - um *grobe* Fehler eher diagnostizieren zu können - und dabei sicherstellen, daß die Institute, was die Meßmethoden betrifft, *unabhängig* voneinander arbeiten. Unterschiede von ein bis zehn Prozent sind - je nach Anzahl und Umfang der gestellten Fragen und Größe des repräsentativen Querschnittes (1000-2000) - nicht als *signifikant*, sondern als *methodenbedingt* zu betrachten.

XII NACHWORT

Die Stunde der Wahrheit - die Wahl - begeht der Wahlforscher feierlich, nachdenklich und offen, etwas abgespannt durch die Strapazen des Wahlkampfes, aber innerlich gestärkt durch den erhöhten Umsatz. Das Resultat wird er gelehrt und aus sicherer Distanz „erklärend“ kommentieren. Hat er - wie es in der letzten Zeit häufiger passierte - mit Dichtung und Gewichtung Pech gehabt, dann wird er an der Gewichtungsschraube weiterdrehen. Ist er mit der zuletzt verwendeten Gewichtung ausnahmsweise nicht auf die Nase gefallen, so wird er dies der neuen „Methode“ zuschreiben. Weshalb sollte er sich die Freude mit dem Gedanken vergällen, daß der Erfolg wohl eher darauf beruht, daß der Zufalls- und Interviewfehler „versehentlich“ weggewichtet wurden? Auch ein blindes Huhn findet mal ein Korn...

So oder so: Die Launen des Zufalls werden dafür Sorge tragen, daß die Gewichtungszauberlehrlinge nicht zur Ruhe kommen. „Die Rohzahlen waren oft genauer als die gewichteten Zahlen“, klagte der Geschäftsführer von Basis-Research der Wirtschaftswoche Nr.1/2 Januar 1987.

Die Wahl wird man überstehen, der Konkurrenz ging es auch nicht viel besser. Ein Gang nach Canossa? Schon eher ein Einkaufsbummel ins nahegelegene Florenz. Und dann eine Erholungsreise an den Golf von Korinth, verbunden mit einem Kulturtrip zu den Tempeln in Delphi.

Auf dem Flug von Mailand nach Athen schlägt man noch schnell im Reiseführer das Wissenswerte zum „Orakel von Delphi“ nach:

„Delphi (7.Jh.v.Chr.) liegt am Südhang des Parnaß (2459m), welcher im Altertum als Sitz des Gottes Apollo galt. Das Orakel Apollos war das berühmteste Orakel des Altertums. Seine große politische Bedeutung fällt in die archaische Zeit. Damals wirkte das Orakel auf die Verfassung der Städte, die Kolonisation und viele politische Unternehmen ein. Noch Platon erkannte Delphis Weisungen eine hohe Bedeutung für die staatliche Ordnung zu.“

Die Ausführungen über den großen politischen Einfluß des Orakels bringen unsere Wahlforscher so richtig auf den Geschmack. Daß selbst Platon die Demoskopie und ihre Weisungen für das Wohl der Gemeinschaft und die Aufrechterhaltung der politischen Ordnung als Notwendigkeit betrachtete, macht unseren Wahlforschern wieder neuen Mut. Welch gute alte Zeit! Anstatt sich wie bisher jede Woche zwecks Entspannung den Glauben an die eigenen Prognosen beim Nervenklempner auf der Couch wieder ausreden zu lassen, könnte man es doch einmal mit dem Orakel versuchen, das seine Kunst offensichtlich viel besser beherrscht.

Im Reiseführer erfährt man, wie eine Konsultation beim Orakel abläuft:

Das Medium von Delphi war eine ältere Frau: die Pythia
Nach einer Opfergabe nahmen die Pythia und ihre ratsuchenden Klienten ein gemeinsames Bad in dem kastalischen Springbrunnen, wo ihr die Probleme vorgelegt wurden. Nach diesem Ritual ging sie in den Tempel und begann mit der Meditation. In ihrem veränderten Bewußtseinszustand fing die Pythia zu sprechen an. Ihre Wahrsprüche wurden den Ratsuchenden nicht direkt mitgeteilt, sondern sie wurden von Priestern interpretiert und in kunstvollen Versen niedergeschrieben, deren Auslegung nicht immer einfach war.

Auf der Fahrt nach Delphi beraten unsere Wallfahrer, was sie der Orakelpriesterin Pythia als Opfergabe bieten könnten. Sie einigen sich auf ein Wertpapier und legen ihre letzte Prognose auf den Altar. Vom Schatzmeister des Tempels nach Nominal- und Kurswert befragt, meinen sie, solche Papiere würden selten zum Nennwert genommen und die Kotierung an der Gerüchtebörse sei erst kürzlich erfolgt. Entscheidend sei aber der innere Wert, und der Selbstkostenpreis habe fast 100.000,- DM betragen. Der Schatzmeister erwidert sachlich aber höflich, der Altar sei keine Endlagerungsstätte und fünf Mark in bar wären ihm lieber. Nach dieser kalten Dusche verspüren unsere Wahlkampfmanager keine Lust mehr auf ein Bad im Nymphen-Springbrunnen. Auf ihre Bitte hin verlegt die Pythia die Reinigungsprozedur in die Sauna. Dort tragen sie ihr mit bewegten Worten die Klage über die Krankheit des repräsentativen Querschnittes vor:

seine **Wankelmütigkeit** !

Den Wahrspruch der Pythia am nächste Morgen

- Achte das Einmaleins der Statistik -

haben unsere Wahlforscher bis heute nicht enträtselt. Als Datenhändler hat man nicht viel Zeit zur Muße, auch werden sie den Verdacht nicht los, daß die Pythia sie im Dampfbad mit der Konkurrenz aus „Wetten daß...“ verwechselt hat. Wie dem auch sei, solange unsere Polit-Wetterfrösche für ihre gewichteten Lottozahlen dank der guten Tarnung reißenden Absatz finden und jedermann - von Politikern bis hin zur Regenbogenpresse - diesem Zahlen-Horoskop hörig ist, verhallt der Wahrspruch der Pythia ungehört. Man kann von unseren Datenhändlern - Kreuzritter für freies Unternehmertum - auch nicht erwarten, daß sie sich freiwillig in den eigenen Speck schneiden. Das wäre gänzlich neu in einer Narren-freien Marktwirtschaft.

In Meinungsforschungskreisen ist die „Pythia vom Bodensee“ eine wohlbekanntes und ein wohlbeachtetes Phänomen.

Bei dieser Institution handelt es sich aber nicht um eine Reinkarnation der Pythia in unserem Jahrhundert - allenfalls um eine von der Frankfurter Allgemeinen (FAZ) bestellte Thronräuberin - sondern um den Kosenamen von Frau Prof. Dr. Dr. h. c. Noelle-Neumann im eingeweihten Blätterwald.³⁵

Daß zwischen den beiden Damen Welten liegen, oder, wie wir in der Schweiz zu sagen pflegen, sie ihr Heu nicht auf der gleichen Bühne haben, geht auch aus dem „Wahr“-spruch der „Pythia vom Bodensee“

Die Schweigespirale, Öffentliche Meinung - unsere soziale Haut hervor, der mit dem Einmaleins der Statistik auf Kriegsfuß steht, soweit es die Anwendung auf die Gegenwartspolitik betrifft.

In den Auswirkungen auf die Staatspolitik ist jedoch eine gewisse Parallelität zwischen den beiden Damen unverkennbar. Ich kann mir allerdings nicht vorstellen, daß Platon den Prozentzahlen der „Pythia vom Bodensee“ so leichtfüßig gefolgt wäre wie Helmut Kohl³⁶.

³⁵ Den Hinweis auf die "Pythia vom Bodensee" verdanke ich Thomas Hochkirchen. Die Bezeichnung "Pythia vom Bodensee" ist seit langem im Umlauf und geht offenbar auf die FAZ zurück.

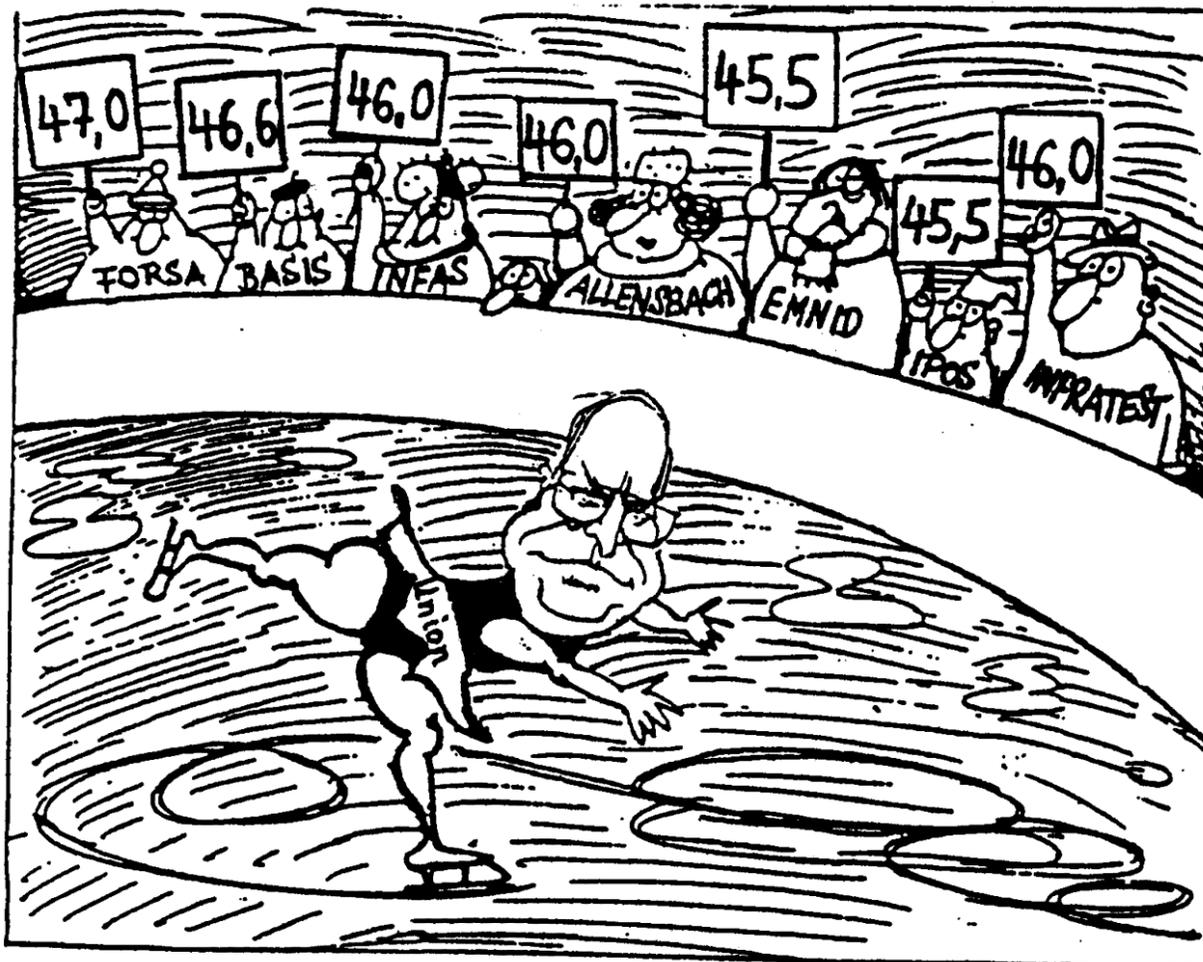
³⁶ und einige seiner Vorgänger im Amt mit der Ausnahme von H. Schmidt: "Ich hasse Demoskopie" erklärte dieser

abrupt in einem Gespräch mit einem Journalisten, in welchem er Dinge aufzählte, die ihm Freude bereiten.

Kohls Prognosen- und Zahlengläubigkeit wird illustriert durch seine erste Stellungnahme zu den Bundestagswahlen vom 25. Januar 1987. Nachdem sich bei den Hochrechnungen am Wahlabend ein Resultat von knapp 44% für die Union abzeichnete - also ein Verlust von etwa 5% - , zeigte er sich im ZDF enttäuscht:

"Wir in der Union haben unser Ziel nicht erreicht. Ich selbst habe gerechnet in einer Größenordnung von 46%, wir sind deutlich unter dieser Zahl geblieben. Das ist betrüblich.
Wir haben Verluste erlitten, Verluste, mit denen niemand in diesem Umfang gerechnet hat."

Aus dieser Aussage geht hervor, daß er bzw. seine demoskopische Ratgeberin die Vorstellung haben, der Unterschied von 46% und 44% sei so eklatant, daß er einem Demoskop unmöglich entgehen könne. Als Erklärung für den Unterschied führte er die gegenüber 1983 um 5% geringere Wahlbeteiligung an. Auf den folgenden Seiten ist ein Statistik-Kurs in Bildform für das Bundeskanzleramt abgedruckt.

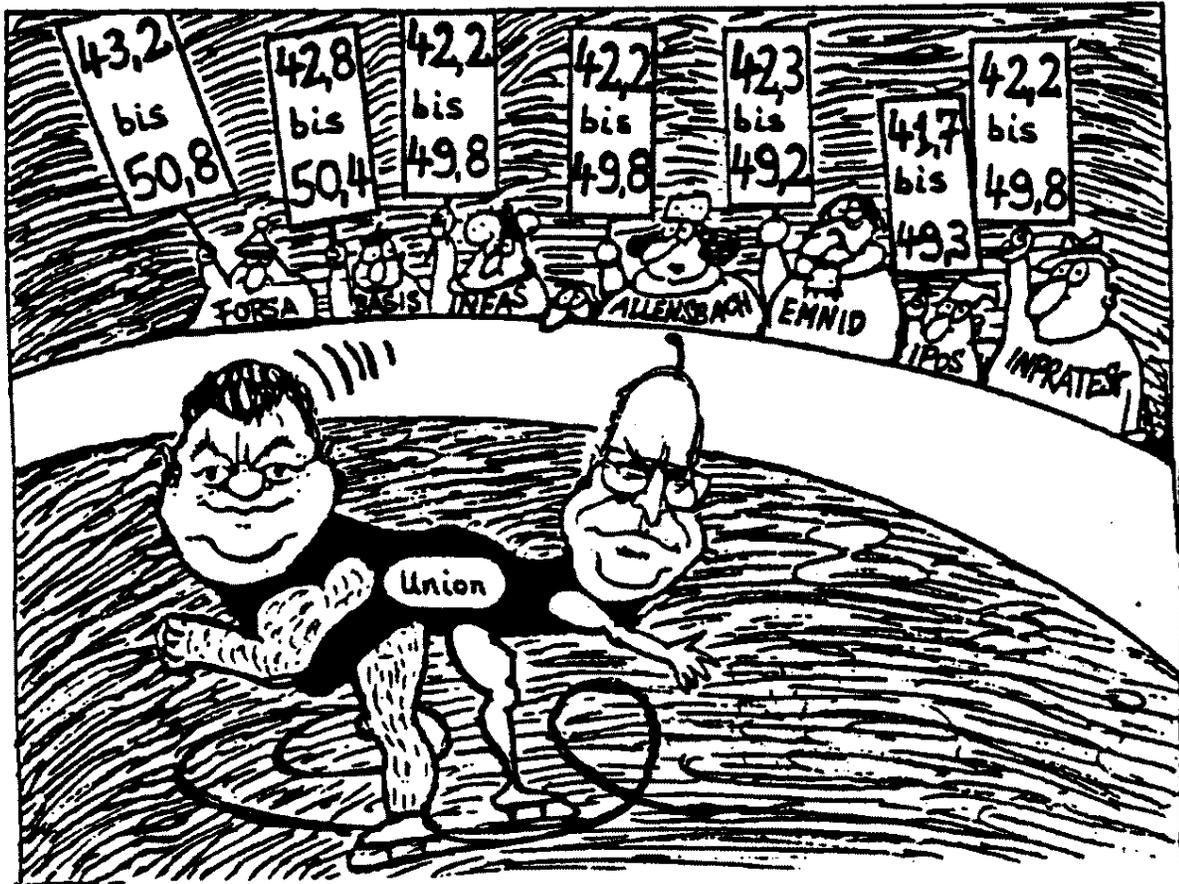


frei nach Mannheimer Morgen

Diagnose : Die Mitglieder der Jury haben geschummelt. Sie haben ihre Umfrageergebnisse solange umgewichtet, bis die vorherrschende Meinung herauskam.^(*)

^(*) Auf der Basis von 1000 Interviews besteht nämlich eine Chance von weniger als 1%, daß die Resultate von sieben *unabhängigen* Meinungsforschungsinstituten einen Spielraum von lediglich 45,5% bis 47% aufweisen. Selbst bei 2000 Interviews liegt diese Chance noch deutlich unter 5%

Erste-Hilfe-Kurs in Statistik



Siamesische Kür

frei nach Mannheimer Morgen

Danksagung

Diese Arbeit wurde durch eine Anfrage der Zeitschrift für Markt-, Meinungs- und Zukunftsforschung angeregt. Die Durchführung wurde durch ein vorzeitig gewährtes Forschungs-Freisemester ermöglicht. Die Arbeit stellt das Resultat einer Teamarbeit unter meiner Anleitung dar. Mein besonderer Dank geht an *Edith Achiman* und *Ullrich Tesche*.

Für Korrekturen und redaktionelle Vorschläge danke ich Thomas Hochkirchen und Frank Sauder. Bernd Wallaschkowski bemühte sich um die Textverarbeitung mit seinem privaten PC.

Ohne die zahlreichen Gespräche mit Nichtstatistikern - u. a. mit W. Huncke, G. Pinkernell und meiner Frau - die mich mit ihren Fragen oft zur Verzweiflung trieben, hätte ich nicht versucht, den statistischen Jargon über Bord zu werfen und die Auslosung des repräsentativen Querschnittes zum Mittelpunkt meiner Darstellung zu machen.

Ich danke meinen Kollegen und Studenten G. Heindl, H. Hoffmann, H. Scheid, E. Scholz, D. Vogt und A. Zieschang sowie meiner Frau und Herrn M. Kroemer für die kritische Durchsicht der verschiedenen Fassungen des Manuskriptes. Durch ihre Fragen und Vorschläge wurde die Gestaltung der Arbeit verbessert.

Ich danke Herrn Witte vom Uni-Rechenzentrum für seine Kooperation bei der Durchführung der vielen Simulationen.

Ich danke Yolanda Aguado, Ruth Horine und Martine Mérigeau sowie Agustin Blanco-Roiz und Gilbert Liébray für die Übersetzung der Zusammenfassung ins Französische und Spanische. Durch ihre Rückfragen und die daraus resultierenden Gespräche wurde die Gestaltung wesentlich verbessert.

Mein Dank geht auch an den Demokrit Verlag und an Herrn G. Wickert für die gewährte Unterstützung und für ihre Engelsgeduld.

Der Lotteriebertrag des repräsentativen Querschnittes

Wahlprognosen und Meinungsumfragen und der Ablaßhandel mit Prozentzahlen

Zusammenfassung

Das Ziel dieser Arbeit besteht darin, dem Nicht-Statistiker mit Hilfe von frappanten Beispielen und zugespitzten Formulierungen klar zu machen, daß die Meinungsforscher in der Bundesrepublik ihre hochtrabenden Versprechen betreffend Wahlprognosen seit über dreißig Jahren aus statistischen Gründen nicht einhalten können. Zunächst wird gezeigt, daß der fromme Spruch vom „repräsentativen“ Querschnitt - im Sinne eines Spiegel- oder Miniaturbildes der Bevölkerung - ein reiner Etikettenschwindel ist. Es ist aus mathematischen Gründen unmöglich, Miniaturbilder zu realisieren, die aus 1000 oder 2000 Personen bestehen. Sogar mit 10000 Personen geht es nicht, aber dann wären Meinungsumfragen wenigstens „unbezahlbar“, und die meisten Probleme würden sich erst gar nicht stellen. Selbst wenn Miniaturbilder eines Tages durch höhere Gewalt auf die Welt gebracht würden, könnten sie die völlig unrealistischen Vorstellungen und Erwartungen der Meinungsforscher und ihrer Kunden nicht erfüllen.

Durch das Suggestivwort „repräsentativer“ Querschnitt wird bei den Konsumenten und Auftraggebern von Meinungsumfragen der Eindruck erweckt, die gelieferten Daten seien glaubwürdig und zuverlässig. Die Autorität des „repräsentativen“ Querschnittes löst bei vielen eine Art Verpflichtung und Resignation aus, sich mit der gelieferten Information abzufinden, als hätte die große Mehrheit der Bevölkerung ganz persönlich zu ihnen gesprochen. Jeder Branchenkundige weiß jedoch, daß es die Zufallsstichprobe ist - d.h. eine per Lotterie getroffene Auswahl - welche die mit dem Werbeslogan „Repräsentativumfrage“ gemachten Versprechen einlösen soll (von den Quotenstichproben sehen wir fürs erste ab ...). **Die Auswahl per Lotterie liefert zwar viel bessere Resultate, als der Laie sich vorstellen kann, aber bei weitem nicht so gute, wie die Prognoseindustrie und ihre stille, aber äußerst geschickte Lobby seit Jahrzehnten behaupten.** Mit Computersimulationen läßt sich zeigen, daß die Auswahl per Lotterie in „repräsentativen“ Querschnitten größere Abweichungen verursacht, als bisher angenommen wurde. Dies macht es praktisch unmöglich, auf die Fragen nach den politischen Kräfteverhältnissen in der Bundesrepublik verbindliche Antworten zu geben, weil relativ geringe Verschiebungen die politische Situation grundlegend verändern können.

Im Gegensatz zu anderen Ländern bewegen sich hier die Änderungen der relevanten Parteistärken in engen Grenzen. Seit 1965 lagen beispielsweise die Resultate für die CDU/CSU zwischen 44,3% und 48,8%, für die SPD zwischen 37,0% und 45,8% und für die FDP zwischen 5,8% und 10,6%. Zählt man ab 1983 bei der SPD die Grünen hinzu, die zu einem guten Teil von der SPD „abgewandert“ sind, dann reduziert sich der Bereich der SPD auf 39,3% bis 45,8%. Im Vergleich dazu weisen die üblichen Vertrauensintervalle - deren Aufgabe bekanntlich darin besteht, Aufschluß darüber zu geben, in welchem Umfang der „repräsentative“ Querschnitt durch auslosungsbedingte Abweichungen verzerrt sein kann - ein Ausmaß von $\pm 2\%$ bis $\pm 4\%$ auf. Wie man sieht, liegen die Vertrauensintervalle in der Größenordnung der historischen Schwankungsbreiten der Parteistärken und übersteigen diese sogar gelegentlich. **Der „repräsentative“ Querschnitt kann also wegen der auslosungsbedingten Fehler nicht mehr Information liefern als die vorangehenden Wahlresultate, oft sogar weniger.**

Bei dieser Sachlage bleibt den Wahlforschern „nur“ der Ausweg, die auslosungsbedingten Abweichungen unter den Teppich zu kehren. Die Bekanntgabe von aktuellen Wahlprognosen in statistisch vertretbarer Form, d.h. unter Angabe der auslosungsbedingten Abweichungen wie z.B.

CDU/CSU 42% - 50%, SPD 35% - 43%, FDP 5% - 10%, Grüne 4% - 9%

würde so lächerlich wirken, daß sie nur in Witzblättern oder als Comic-strip möglich wäre. Wer würde für eine solche Banalität 50000 Mark hinblättern? Noch ärgerlicher wäre der laufende Einnahmeausfall, denn für den gleichen faulen Witz kann man nicht jede Woche abkassieren.

Also tun die Wahlforscher so, als hätten sie das statistische Ei des Columbus entdeckt: Eine magische Formel, welche die auslosungsbedingten Fehler in nichts auflöst. Mit Hilfe einer Umgewichtungszprozedur, deren alchimistischer Gehalt unverkennbar ist, wird den Vertrauensintervallen die überschüssige Luft abgelassen, und alsdann schlüpfen „exakte Zahlen“ aus der Retorte, bis hin zur Stelle nach dem Komma. Dieses Vabanquespiel basiert auf einer Umgewichtung der aktuellen Umfrageergebnisse in Abhängigkeit vom früheren Wahlverhalten (sogenannte Recall-Frage). Der statistische Gehalt dieser „Einkommens-umverteilung“ im Prozentbereich beruht auf dem folgenden „Gedankengang“: Falls die Wähler die Absicht haben, für die gleiche Partei zu stimmen wie bei der letzten Wahl vor vier Jahren, und falls sie dies dem Interviewer anvertrauen - unsere Wahlforscher hoffen also, daß die Wähler ihre Gedächtnislücken selbst „reparieren“, auch diejenigen, die damals noch nicht wahlberechtigt waren oder seither gestorben sind - ja, dann funktioniert die Umgewichtung statistisch tadellos. Abgesehen davon braucht man unter diesen Annahmen weder Meinungsumfragen noch eine Wahl durchzuführen, um Resultate zu erhalten. Man kopiere einfach die Resultate der letzten Wahl

Tragischerweise entpuppt sich das Ei des Columbus unserer Wahlforscher in der Praxis als Kuckucksei. Die alchimistische Recall-Formel produziert häufig surrealistische Resultate, die nicht zu vermarkten sind. So sind weitere Meinungsklima-Kuren vonnöten, deren phantasievolle Nomenklatur den höchsten Ansprüchen genügt. Ein Schönheitsfehler mag darin bestehen, daß das Endresultat ganz anders ermittelt wird: **Eine Prognose wird einfach so gemacht, daß das letzte Wahlresultat in Abhängigkeit von der vermeintlichen politischen Windrichtung Pimal Daumen ein bißchen nach oben oder unten korrigiert wird.** Die aktuellen Umfrageergebnisse spielen dabei eine untergeordnete Rolle. Sie dienen allenfalls dazu, die vermeintliche politische Windrichtung zu ermitteln. In Anbetracht der unkontrollierbaren auslosungsbedingten Abweichungen und Interviewfehler erfordert dieses Abenteuer jene Mischung von grenzenlosem Selbstvertrauen und unerschütterlicher Ahnungslosigkeit, welche sonst zum Rüstzeug von Politikern gehören. **Die laufend feilgebotenen Prognosezahlen für die sogenannte Sonntagsfrage**

„Wie würden Sie wählen, wenn am nächsten Sonntag Bundestagswahl wäre?“

reflektieren also ausschließlich die Spekulationen der Wahlforscher, oder, wie sie es salbungsvoll ausdrücken: ihre Gewichtungskunst.

Bei dieser Sachlage versteht es sich von selbst, daß Trendangaben keinen Bezug zur Realität haben, weil sie durch auslosungsbedingte Pseudo-Trends bis zur Unkenntlichkeit entstellt sind. Um es boshaft auszudrücken: Die „Gunst“ des Zufalls beschert den Wahlforschern laufend Neuigkeiten und Schlagzeilen. Was sich wirklich abspielt, das weiß kein Mensch, und die von den Meinungsforschungsinstituten gelieferten Daten über das zeitliche Auf und Ab der

Parteistärken - z.B. CDU/CSU -2,8%, SPD +1,9%, FDP +1,3%, Grüne -0,4% - haben ungetrübten Horoskop-Charakter. Sie täuschen Veränderungen vor, die quantitativ mit an Sicherheit grenzender Wahrscheinlichkeit falsch sind und die häufig nicht einmal qualitativ richtig sind.

Diese Probleme werden dadurch potenziert, daß in der Meinungsforschung ein falsches statistisches Modell zur Anwendung kommt: die Binomialverteilung und ihre Approximation - die Normalverteilung. Beide sind bei den empirischen Sozialforschern auf der ganzen Welt populär. Diese Verteilungen belohnen den glücklichen Gläubigen mit den kleinstmöglichen Vertrauensintervallen - die von der Auslosung des „repräsentativen“ Querschnittes verursachten Abweichungen werden also minimal. Was will er mehr? Er wird nicht belohnt, wenn er der Tatsache ins Auge sieht, daß dieses Modell in keiner Weise eine statistisch legitime Beschreibung seiner Meinungsumfrage darstellt, sondern der Ausdruck von reinem Wunschdenken ist, das von einer Generation von Lehrbüchern auf die nächste vererbt wird, wie ein Fluch bis ins dritte und vierte Geschlecht. Das Modell unseres Gläubigen setzt voraus, daß die von ihm organisierte Meinungsumfrage aus einer einzigen Frage besteht, die mit „ja“ oder „nein“ zu beantworten ist. Jede Hausfrau weiß, daß es nicht genügt, einmal „nein“ zu sagen, wenn der Interviewer vor der Tür steht. Es ist ihr sonnenklar, daß der ungebetene Gast - der sich an einem schönen Nachmittag unvermittelt in ihrem Wohnzimmer festsetzt - ihr mit seinem seitenlangen Fragebogen ein Loch in den Bauch fragen wird und daß sie vor lauter Möglichkeiten oft nicht weiß, welche Antwort sie geben soll. Statistisch bedeutet dies, daß das korrekte Modell für eine Meinungsumfrage nicht aus einer Binomialverteilung besteht, sondern aus dutzenden von Multinomialverteilungen mit verschiedenen Freiheitsgraden, die gleichzeitig betrachtet werden müssen, was - wenn überhaupt - meist nur mit einer Computersimulation möglich ist. Die Berechnung von Vertrauensintervallen bei Umfragen mit Hilfe der Normalverteilung erinnert an jenen Metzger, der eine Wurst in ein Rudel von Hunden wirft und dann jedem Besitzer den Schmaus in Rechnung stellt. Gewiß kann kein Besitzer beweisen, daß es nicht sein Hund war, der die Wurst erwischte. Dennoch ist der Metzger kein Modell für Integrität. Solche „Vertrauensintervalle“ haben die Funktion von statistischen Alibi-Übungen und Milchmädchenrechnungen.

Computersimulationen von Zufallsstichproben zeigen, daß solche Vertrauensintervalle bis zu 100% oder mehr vergrößert werden müssen, damit die gewählte statistische Sicherheit eingehalten werden kann. Je umfangreicher die Meinungsumfrage, desto größer der Zuschlag. Die Größe der tatsächlichen Vertrauensintervalle hängt von der Anzahl der Fragen und der Antwortmöglichkeiten ab und nicht allein von der Anzahl der Interviews und der statistischen Sicherheit, wie bisher angenommen wurde. Auf den Punkt gebracht: Unter der Glut der Simulations-Sonne schmelzen Berge von angeblich so harten Daten wie Butter.

Wenn der berühmte Psychologe Carl Gustav Jung - der Entdecker und Erforscher des kollektiven Unbewußten - noch lebte, würde er sich wohl gezwungen sehen, seine lange Liste kollektiver Inhalte durch einen weiteren Archetyp zu ergänzen: Zahlengläubigkeit.

Die Alchimisten vergangener Zeiten würden ihre heutigen Kollegen wohl beneiden. Was jenen während Jahrhunderten mangels geeigneter Stoffe und Know-how versagt blieb, ist für unsere Meinungsforscher zur Routine geworden:

aus Prozentzahlen und Aberglauben eine Goldgrube zu machen.

The lottery aspect of the representative sample

Election forecasts and opinion polls

The traffic with percentage points

The purpose of this paper consists in demonstrating to non-statisticians by means of striking examples that for statistical reasons pollsters in the Federal Republic of Germany cannot deliver the goods they have promised the public for the last thirty years. The pious wish of a “representative” sample in the sense of a mirror image or miniature image is shown to be a public relations gimmick. For mathematical reasons it is impossible to construct such images based on 1000 or 2000 people. With 10000 it won't work either, but at least opinion polls would become so costly that most problems would not exist in the first place. Even if miniature images were to be created by an act of God, they could not fulfil the totally unrealistic ideas and expectations of the pollsters and their customers.

The highly suggestive term “representative” sample sounds so reassuring. It brainwashes people into believing that election forecasts and opinion polls must be accurate and reliable. The moral authority of the “representative” sample provokes in many people a feeling of obligation and resignation. They swallow the numbers with which they have been fed, as if the silent majority had spoken to them in person. Everybody in the business knows however that it is random sampling - i.e. the selection by lottery - that is supposed to back up the promise behind “representative” sample, whatever the latter may mean. (As for quota samples, we prefer to look the other way.) **No doubt the lottery principle leads to much better results than a layman expects, but they are not nearly as good as the forecasting industry keeps claiming ever since it “incorporated” as a pressure group.** By means of computer simulations it can be shown that random sampling causes larger errors than has been assumed. These errors make it virtually impossible to provide politically relevant information in the Federal Republic of Germany by means of “representative” samples, because relatively minor changes in party strengths can lead to a completely different political situation.

For in Germany, unlike other countries, the strengths of the dominating parties have traditionally varied in narrow ranges of a few percentage points. For instance, since 1965 the CDU/CSU has varied between 44,3% and 48,8%, the SPD between 37,0% and 45,8% and the swing party FDP between 5,8% and 10,6%. If, as of 1983, one adds the Greens to the SPD (one may consider the former to a large extent as a “spin off” from the latter), then the range of SPD narrows to 6,5% (from 39,3% to 45,8%). By comparison the customary confidence intervals amount to $\pm 2\%$ to $\pm 4\%$. (Recall that their job is to keep you posted on the extent to which the information you get from your sample is distorted by lottery effects.) From this it is obvious that confidence intervals have the same order of magnitude as the historical variations of party strengths and sometimes they are even bigger. **This means that “representative” samples provide no more information than the previous elections and often even less.**

For this reason the “only” way out for the forecasters is to sweep the lottery effect under the rug. For if they were to divulge their forecasts in a form which could be backed up statistically, i.e. by clearly indicating the errors caused by random sampling, for example

CDU/CSU 42% - 50%, SPD 35% - 43%, FDP 5% - 10%, Greens 4% - 9%,

then their numbers would be eligible for the funny pages. Who would fork out thirty thousand dollars for such a banality? Even worse would be the continual loss of revenue because you can't cash in every week for the same bad joke.

So the pollsters act as if they had discovered the magic formula in statistics which makes random errors evaporate: By means of a weighting procedure, the alchemical origins of which are unmistakable, they first let out the surplus air from the confidence intervals, whereupon "exact figures" down to decimal points hatch out of the retort. Their gamble is based on modifying current voter preferences according to how people voted in the previous election. The statistical gist of their "reallocation" program is the following: If the voters plan to cast their ballots in the same way as they did in the election four years previously and if they confide their intention to the interviewer - our forecasters thus hope that the voters "make up" for their memory lapses, including those who were too young to vote then or have died since - , yes, then everything is perfect in the best of all possible worlds ... and even the statistics of their reallocation system work impeccably. Never mind the fact that under these assumptions you neither need polls nor an election to get results. Just copy the ones from the previous elections ... and there you are.

Tragically the pollsters magic formula frequently produces surrealistic results which they don't dare to sell because they just don't look right. To cover up the further doctoring of half-baked results our pollsters have developed a highfalutin nomenclature which exceeds any standard one might ever have. The minor blemish is that they arrive at the final results in a completely different way: **To produce a forecast they merely adjust the results of the last election a little bit upwards or downwards according to the alleged direction of the political trend.** They largely ignore current 1poll results and use them, if at all, to determine the alleged direction of the trend. Considering the uncontrollable nature of lottery and interview errors, this adventure requires that blend of unlimited self-confidence and unshakeable ignorance with which usually only politicians are endowed. **As for the famous "Sunday-question":**

"For whom would you vote if elections were held next Sunday?",

the upshot is that pollsters' results merely reflect their speculation, or, as they put it unctuously - their art of weighting.

From this it is obvious that the trends they report have no bearing on reality whatsoever. For the real trends are so distorted by pseudo-trends resulting from the lottery effect that they become unrecognisable. To put it sarcastically: The favours of chance continually provide our pollsters with news and headlines. In reality nobody knows what is going on. The numbers the pollsters keep dishing out - for instance CDU/CSU -2,8%, SPD +1,9%, FDP +1,3%, Greens -0,4% - contain no more information than a horoscope. They fake changes which are almost surely quantitatively false and are often qualitatively incorrect as well.

These problems are magnified by a wrong statistical model used in demoscropy: the binomial distribution and its substitute, the normal distribution. Both are popular among social scientists throughout the world. They reward the happy believer with the smallest possible confidence intervals - thus the errors due to random sampling become minimal. What more does he want? He gets no rewards for facing the fact that his model is not a legitimate statistical description of the poll he conducts. Instead it is merely a product of wishful thinking which is being handed down from one generation of textbooks to the next like a curse "unto the third and fourth generation".

For this model assumes that an opinion poll consists of a single question which is to be answered by either “yes” or “no”. Every housewife knows that it is not enough to say “no” once, when the interviewer knocks at her door. It is very clear to her that the uninvited guest is going to install himself comfortably in her living room for the afternoon, bombarding her with dozens of multiple choice questions so that she often is at a loss for an answer. Statistically this means that the correct model for an opinion poll does not consist of a single binomial distribution but of dozens of multinomial distributions with different degrees of freedom which have to be taken into account simultaneously. If at all, this can be done only by computer simulations. Computing confidence intervals for opinion polls by means of the normal distribution is reminiscent of that butcher who throws one sausage into a pack of dogs and then charges each owner for the “feast”. No owner can prove that it was not his dog that devoured the sausage. Still the butcher is not a model of integrity. Such “confidence” intervals merely serve as a statistical alibi. They have no bearing on reality whatsoever.

Computer simulations of random sampling show that confidence intervals based on the normal distribution have to be increased by 100% or more in order to maintain the chosen probability level. The longer the questionnaire, the greater the increase. The real confidence intervals depend on the number of questions asked and their multiple choice complexity, and not only on the number of people polled and the chosen probability level as has been assumed. In short: under the heat of the simulation-sun vast quantities of allegedly hard data melt down like butter.

If the famous psychologist, Carl Gustav Jung - the discoverer and explorer of the collective unconscious - were still alive, he would feel compelled to extend his long list of collective contents by another archetype:

the happy believer in numbers.

The alchemists of past times would envy their colleagues of today. What they tried to do in vain for centuries for lack of suitable resources and know-how has become routine for our pollsters:

to transform percentage points and superstition into a goldmine.