

## Die Realität in der Mediaforschung <sup>1</sup>

Mehr als in jedem anderen Gebiet der Marktforschung haben wir im Bereich der Mediaforschung das Problem, einen Bezug zwischen den ermittelten Daten und der Wirklichkeit herzustellen.

Wenn es um Daten zu Besitz und Verbrauch geht, dann besteht in vielen Fällen die Möglichkeit, die durch eine Erhebung herausgefundenen Daten mit statistischen Daten oder mit solchen aus relativ verlässlichen anderen Erhebungen zu vergleichen. Ohne Zweifel wird von dieser Möglichkeit viel zu wenig Gebrauch gemacht; aber dieses Problem muss uns an dieser Stelle nicht beschäftigen.

In der Medienforschung sieht es in weiten Breichen gänzlich anders aus. Sofern es sich um den Besitz von z. B. Fernseh- oder Rundfunkgeräten handelt, kann man die Zahlen mit Statistiken vergleichen. Wenn es um Vertriebsuntersuchungen geht, kann man die Resultate mit Abonnements oder Verkaufszahlen in Beziehung setzen und damit überprüfen. Wenn es aber um den Medienkonsum geht, - also um das Zuschauen beim Fernsehen, das Radiohören, das Lesen von Zeitungen oder Zeitschriften, etc. - dann fehlen Maßstäbe für die Richtigkeit der erzielten Ergebnisse.

Es gibt Bemühungen von verschiedenen Ausgangspunkten her, die Realitätsnähe oder Verlässlichkeit der Daten zu sichern, soweit es möglich ist:

1. Man versucht im Bereich der Stichprobenbildung und der Erfassung aller Teile der Stichprobe so exakt wie möglich vorzugehen;
2. man bemüht sich in der Ermittlung des Sachverhaltes mit dem Maßstab der Plausibilität das Verhalten so gut, wie man es sich vorstellen kann, zu erfassen.
3. Im Bereich der Datenverarbeitung bemüht man sich darum, durch Verrechnung oder durch Berechnungen innerhalb der Daten oder durch in - Beziehung - setzen mit externen Daten deren Richtigkeit zu überprüfen.

Allen diesen Bemühungen sind Grenzen gesetzt, und so kommt es, dass seit Anfang an, mindestens jedoch in den letzten 35 Jahren, in denen die Problematik der Erfassung immer deutlicher wurde, die Zweifel an der Korrektheit der Daten bestehen geblieben sind, zeitweilig in geringem Ausmaße, zeitweilig mit großer Heftigkeit umstritten.

Es hat gelegentlich Versuche gegeben, die Erhebungsmethodik, oder wesentliche Teile davon, auf experimentellem Wege zu überprüfen. Oder man hat, wie z. B. Belson, durch Nachbefragungen herauszufinden versucht, wie gut oder wie abweichend die Antworten der Befragten von der zuvor objektiv festgestellten Realität sind. Diese experimentellen Untersuchungen haben das Problem des Vertrauens in die Daten nicht beseitigt, man könnte sogar sagen, dass die Zweifel bestärkt worden sind.

Diese frustrierende Dauersituation der Mediaforschung hat sie in zwei Lager getrennt:

---

<sup>1</sup> Entwurf eines unveröffentlichten Aufsatzes vom 13. Dezember 1983

- Das eine Lager bemüht sich immer wieder und mit immer wieder verschiedenen Verfahrensweisen, eine bessere Erfassung der Realität zu versuchen; im Großen und Ganzen noch immer nicht mit einem befriedigenden Ergebnis.
- Eine zweite Gruppe findet sich mehr oder minder mit der Tatsache ab, dass die Daten einen unbestimmten und unbestimmbaren Bezug zur Realität haben; und sie postulieren, dass es nunmehr primär darauf ankommt, bestimmte Konventionen zu formulieren, unter denen die Ergebnisse als bestmögliche Lösung des Problems angesehen und in der Praxis der Mediaplanung gebraucht werden.

Die Trennungslinie zwischen diesen beiden Gruppen ist sicherlich nicht so scharf, wie es bei dieser Beschreibung den Anschein hat. Die erste Gruppe begnügt sich nolens volens mit dem, was man zur Zeit erreichen kann; und ihre Anhänger erklären die Ergebnisse nicht als schlichtweg unzuverlässig und für die praktische Planung unbrauchbar, sondern geben sie für die Verwendung frei. Insofern gibt es bei dieser, sagen wir einmal „idealistischen“ Gruppe durchaus keine Abneigung gegen ein pragmatisches Vorgehen.

Die zweite Gruppe, die wir dann vielleicht die „Pragmatiker“ nennen dürfen, stehen den erhobenen Daten ohne Zweifel nicht naiv und vorbehaltlos gegenüber, sondern sehen sie durchaus kritisch und suchen ebenfalls nach Möglichkeiten zu ihrer Verbesserung.

Der Unterschied liegt mehr in den Tendenzen, vielleicht auch in der Betonung der anzustrebenden Ziele, auch in der Forschungspolitik. Nach meiner Überzeugung hat diese unterschiedliche Haltung gewisse Folgen, vor allem dann, wenn die Vorstellungen der einen oder der anderen Gruppe in verkürzter Form vielleicht als Schlagworte geäußert werden, und dann die allgemeine Haltung zu den Forschungsvorhaben und zu den Ergebnissen beeinflussen.

In jeder wissenschaftlichen Disziplin kommt sicherlich an bestimmten Punkten die Stunde der Wahrheit. Es sind Punkte, an denen die Unzulänglichkeit der Verfahrensweisen, der Modellvorstellungen, der Ziele oder Ideale in irgendeine kritische Phase geraten sind. Diese kritische Phase kann mancherlei Ursachen haben, von „politischen“ Ursachen der Betrachtung der Ergebnisse durch die Umwelt oder die Benutzer, oder geprägt von irgendwelchen wesentlichen Schritten in der Erkenntnis.

Es ist mein persönlicher Eindruck, dass die Marktforschung generell, damit aber auch die Mediaforschung derzeit in eine solche kritische Phase getreten ist. Diese ist auf der positiven Seite verursacht durch neue Verfahrensweisen, aber auch durch Ansätze für neue Erkenntnisse. Auf der negativen Seite finden wir eine Welle der Unzufriedenheit oder des Misstrauens in die Untersuchungen, die zu mehr oder minder berechtigter Kritik geführt haben.

Der Ansatzpunkt für meine Fortführung der seit Jahr mehr oder minder öffentlichen Diskussion ist der kürzlich abgedruckte Vortrag von Friedrich Wendt über „Das AG.MA.-Modell“. Dies ist eine hervorragende, sehr prägnante und konkrete Darstellung der Vorstellungen, die sich bei einer Reihe von Kollegen gebildet haben. Im Gegensatz zu früheren Äußerungen anderer Kollegen bietet sie genügend Ansatzpunkte, um eine die Sache hoffentlich fördernde Diskussion auszulösen.

## Was ist „Realität“?

Eine der früheren Beiträge zu diesem Thema stammt vom Kollegen Dr. Klaus Landgrebe. In einer umfassenden Darstellung und kritischen Würdigung der verschiedenen Verfahrensweisen, insbesondere der Heftvorlage zur korrekten Ermittlung von Leserschaften, kommt er zu der berechtigten Vorstellung, dass selbst die derzeit als optimal angesehene Methode der Vorlage konkreter Ausgaben, wie sie vor allen Dingen mit dem Namen Alfred Politz assoziiert wird, auch ihre Probleme hat, und wohl auch nicht ohne weiteres als eine durchweg zutreffende Reflexion der Realität liefert. Als Abschluss dazu vertritt Landgrebe die Auffassung, dass es uns Menschen wohl nie vergönnt ist, die Wahrheit und Wirklichkeit zu erfassen. Um dies zu unterstreichen zitiert er das „Höhlengleichnis“ von Platon:

*„Es sitzen Menschen in einer Höhle mit dem Rücken zum Eingang, und sehen auf eine Feldwand, auf der das, was draußen vor der Höhle vor sich geht nur als Schatten auf der Wand erscheint.“*

So werden wir Menschen niemals die Dinge an sich sehen, sondern immer nur ihre „Schatten“.

Die von Landgrebe gezogene Schlussfolgerung ist von der modernen Erkenntnistheorie korrigiert worden, jedenfalls in einem Bereich, der für die praktische Markt- oder Mediaforschung von Belang ist. Aber abgesehen von dieser Korrektur, die sicherlich Landgrebe heute selbst vornehmen würde, ist der Sinn des platonischen Höhlengleichnisses ein ganz anderer gewesen. Das Gleichnis geht ja weiter:

*„Wenn jemand käme und die Höhenbewohner veranlassen würde, durch den Eingang der Höhle die Realität in der hellen Sonne zu betrachten, so würden die Menschen dies nicht wollen und denjenigen, der sie dazu veranlasst, umbringen“*

Diese vollständige Fabel bezieht sich auf Sokrates und sein Schicksal, er wollte den Menschen DIE Realität statt ihrer schemenhaften, schattenhaften Vorstellungen bringen, - und sie haben ihn umgebracht.

So gesehen hat die platonische Fabel auch heute noch Relevanz, nur in einem gänzlich anderen Sinne, als in der Beschreibung unserer Beziehung zur Realität.

Etwa so alt wie diese Anmerkung von Landgrebe stammt das Wort, welches Friedrich Wendt zugeschrieben wird: *„Was schert mich die dumme Wirklichkeit!“*

Diese Aussage, mag sie so gefallen sein, und - was auch immer der Zusammenhang gewesen sein mag - sie war von Wendt etwas anders gemeint worden, als es von der breiten Fachöffentlichkeit verstanden wurde:

Er hat mit diesem sehr griffigen Wort in der Essenz gemeint, dass naive Antworten auf naive Fragen, in naiver Art und Weise verarbeitet, nicht das rechte Verfahren wäre. Er ist natürlich gänzlich anders verstanden worden:

Dass ihn die Wirklichkeit, die ja doch nicht zu erfahren sein, gar nicht so sehr interessierte, sondern das er sie eigentlich nur als „Rohmaterial“ für eine Verarbeitung aufgrund höherer Einsicht genommen wird.

Eine dritte Stellungnahme stammt von dem eminenten holländischen Kollegen Dr. Jan Noordhoff. Er sagte einmal auf einer Präsentation seiner Erkenntnisse:

*„Wir sind nicht in der Lage die Realität zu erfassen. Darauf kommt es auch gar nicht an, sondern es kommt darauf an, dass die Mediaplaner Daten erhalten, mit denen Sie arbeiten können.“*

Die vierte Stellungnahme stammt von Hans Erdmann-Scheler, der in der Vollversammlung der AG.MA 1982 sagte, und dies in einem Aufsatz wiederholte:

*„Es gibt gar keine Wirklichkeit, sondern es gibt Millionen von Wirklichkeiten.“*

Auf Nachfrage stellte sich heraus, dass er sich auf einen Physiker stützte, Fridtjof Capra.

In dem betreffenden Buch stellt Capra die Erkenntnisse und Hypothesen der modernen Physik vor - übrigens in einer für mich außerordentlich gut verständlichen Art und Weise, so dass gewisse Aspekte von mir erstmalig verstanden werden konnten. Außerdem zieht er Vergleiche zur buddhistischen oder östlichen Mystik, die man jedoch getrennt von der physikalischen Darstellung betrachten und bewerten kann, - oder nach meiner Auffassung: getrennt betrachten und bewerten muss.

Ich habe Capra in einem Vortrag über ein späteres Buch selbst erlebt und dabei die Erfahrung gemacht, dass seine Aussagen viel stärker naturwissenschaftlich geprägt sind und viel „vernünftiger“ sind, als es seine damalige Zuhörerschaft zu sein schien.

Diese Zuhörerschaft setzte sich offenbar nur zu einem Teil aus wissenschaftlich oder sachlich interessierten Zuhörern zusammen, und zu einem anderen Teil aus engagierten Ökologen, Sektierern und Mystikern; die weißen Roben der Anhänger irgendwelcher indischen Sekten waren nicht zu übersehen. Auch in dem zitierten Buch zieht Capra am Ende ein Fazit, dass schon im Duktus wie eine ausdrückliche Korrektur von Vorstellungen aussieht, die durch seine Darstellung physikalischer Vorgänge geweckt worden sind.

Seine Darstellung der Hypothesen hinsichtlich der für uns unbegreiflichen, nur hypothetisch erfassbaren Vorgänge im ganz großen und im ganz kleinen Bereich – im Klartext: im Bereich der Weltraumphysik und im Bereich der kleinsten Teilchen – sind für den „mittleren“ Bereich des menschlichen Lebens auf der Erde ohne Bedeutung. Man kann sicherlich die Materie im Weltraum als „vabande Wahrscheinlichkeit“ auffassen, wo Masse und Energie, Teilchen und Wellen abwechselnd auftreten - wenn man überhaupt von „auftreten“ sprechen darf.

Einen Bezug zum Kauf von einem Pfund Butter oder einer illustrierten Zeitschrift hat dies sicherlich nicht, denn dies sind konkrete, vom Menschen direkt zu beobachtende Vorgänge.

## **Beschreibung des AG.MA.-Modells von Wendt**

Auf dem Kongress der Leserschaftsforscher in Montreal im Jahre 1983 hat Friedrich Wendt in einem sehr grundlegenden und fachkundigen Referat die Philosophie der Deutschen Media-Analyse vorgestellt. Dabei hat er sowohl die theoretischen oder hy-

pothetischen Grundlagen erörtert, wie auch die praktischen Konsequenzen geschildert, die gezogen worden sind.

Damit bietet sich, man kann sagen: erstmalig, die Chance zu einer grundlegenden Auseinandersetzung über dieses Gebiet und seine Grundlagen.

## **Erfassung der Realität**

Wendt schreibt, in etwas verkürzter Form zitiert: „Wir haben im allgemeinen schon und erst recht im besonderen Fall unseres Forschungsbereiches keinen direkten Zugang zur Realität. Sehr früh in der Agla/AG.MA-Geschichte – Anfang der 60-iger Jahre – gewannen wir eine bedeutsame Erkenntnis – und es gab dazu voll Konsens: wir sind nicht in der Lage, Fakten zu messen. Was wir von den Befragten erfahren, sind Meinungen, die diese über ihren eigenen Mediakonsum haben, nicht aber direkt die Information über ihren „realen“ Mediakonsum. Was sie ausdrücken, sind ihre Schätzungen, ihre Vorstellungen über ihr eigenes Verhalten.“

An einer zweiten Stelle zitiert er Begründungen dafür in folgender, wiederum etwas geraffter Form zitiert:

*„Und wir mussten feststellen, dass uns prinzipiell die Realität verborgen bleibt. Für diese Behauptung gibt es gute Gründe.“*

- 1. Können wir mit unseren Sinnen nur ungenaue und unvollständige Informationen empfangen, die sich in unserem Hirn zu geschlossenen Eindrücken umformen.*
- 2. Es ist sehr einfach die allgemeine Anforderung der Mediaplaner zu formulieren. Sie richtet sich auf Werbewirkung; oder genauer: den Fall der Werbewirkung, den man den Medien zuweisen kann. Ich möchte im Moment die Frage ausklammern, ob die Anforderung der Mediaplaner sich tatsächlich auf „Werbewirkung“ bezieht.“*

Zur Ermittlung des Mediakonsums sagt er, dass die Antwort der Befragten von relevanten Einflussfaktoren beeinflusst würde:

*„Ein Teil wird vom Interviewer erzeugt, ein anderer Teil vom Befragten, andere von den Beziehungen zwischen beiden, den Stimmungen des einen oder anderen, ihre Ruhe oder Eile. Auch externe Faktoren spielen mit: das Wetter, die Gegenwart anderer Familienmitglieder usw.“*

*Auf einer derart unsicheren Grundlage sind wir nur verpflichtet, das Resultat des Interviews [die Antwort] in eine Schätzung über den Mediakonsum zu transformieren.“*

Und er schließt darauf:

*„Aufgrund dieser Umstände können wir nicht länger von „Realität“ oder „Präzision“ reden, wenn wir ... auch nur ein Maß für den Mediakonsum aus einer Erhebung ableiten.“*

Anschließend stellt Wendt eine Hypothese vor, die er schon in einer früheren Untersuchung entwickelt hatte: es handelte sich um die von uns für RTV durchgeführte Untersuchung 1975. Es geht dabei um folgendes:

*„Die Antwortmöglichkeiten eines Befragten in Bezug auf jedes Merkmal ist einer Verteilung unterworfen, die wir als intra-individuelle Verteilung einer Zufallsvariable ansehen können, mit einem Mittelwert und einer Standardabweichung, die nur für dieses Indivi-*

*duum gültig ist. Wir haben keine Möglichkeit, direkt von A = der effektiven Antwort – auf M = den intra-individuellen Mittelwert zu schließen.“*

Daraus schließt er folgendes:

*„Was wir brauchen, sind solide Auskünfte, die über einen bestimmten (begrenzten) Zeitraum gültig bleiben, gerade so lange, wie die Mediaplanung sich in die nächste Zukunft richtet. Dafür haben wir unser Bestes zu tun, den Mittelwert wenigstens annähernd als zeitliche Konstante zu bestimmen.“*

In diesen Zusammenhang gehört auch das, was Wendt zum Thema „Validierung“ sagt, denn die Validierung, d. h. die Überprüfung der Gültigkeit irgendwelcher Aussagen oder Daten setzt ja voraus, dass man irgendein reales Maß, also eine Darstellung der Realität hat, an der man die Daten oder die Ergebnisse überprüfen kann. Und hierzu sagt Wendt:

*„Seit einiger Zeit werden in Deutschland intensive Diskussionen über die so genannte Validierung geführt ... ist der Ausdruck „Validierung“ ein Synonym für akzeptabel, brauchbar? Was meint Validierung eigentlich? Der Ausdruck stammt aus der Psychologie und bezieht sich darin auf die Testtheorie. Wegen der völlig anders gelagerten Verhältnisse ist es nicht möglich, diese Betrachtungsweise auch auf die Markt- und speziell Mediaforschung zu übertragen.“*

Zu dieser Darstellung und der dahinter stehenden Auffassung sind nach meinem Dafürhalten eine ganze Menge von Dingen zu sagen, zum Teil grundsätzlicher, zum Teil spezieller Art.

Am leichtesten lässt sich vielleicht über diese Dinge dadurch sprechen, dass man die Übereinstimmung an den Anfang stellt und sie damit aus der Diskussion heraus nimmt.

Wendt hat in richtiger Weise auf den Konsensus hingewiesen, der vor rund 20 Jahren hinsichtlich der Erkenntnis gefunden wurde, dass die *Aussagen* der Befragten über ihren Medienkonsum keine **korrekten** Aussagen über den Medienkonsum darstellten, sondern Meinungen oder Schätzungen, die von sehr unterschiedlichen Dingen, wie Zeit und Umständen beeinflusst worden sein können.

Eine zweite Vorstellung oder Arbeitshypothese haben wir ebenfalls ohne weiteres als plausibel und für das weitere Denken sehr brauchbar akzeptiert: dass es nicht nur inter-individuelle Unterschiede gibt, sondern auch intra-individuelle. Dies hängt, wenn ich es recht sehe, mit der erstgenannten Vorstellung zusammen, und formuliert sie nur in etwas prägnanterer Form.

Diese Vorstellung finden wir in etwas einfacher verständlicher Form schon in den Äußerungen, die von Alfred Politz zu Produkttests gemacht worden sind. Er bezog sich vor rund 20 Jahren insbesondere auf solche Produkte, die wie Genuss- und Nahrungsmittel die Geschmacks- und Geruchsorgane ansprechen. Und diese sind außerordentlich „elastisch“. Damit soll gesagt werden, dass sie von Zeit und Umständen außerordentlich stark beeinflusst werden. Dies hat zur Folge, dass ein und dasselbe Testprodukt von ein und derselben Person zu verschiedenen Zeitpunkten sehr unterschiedlich beurteilt werden kann.

Diese Erkenntnis hat Politz veranlasst, zwei Konsequenzen zu ziehen:

1. Er ließ die Testpersonen das Produkt nicht einmal ausprobieren, sondern gab ihnen die Möglichkeit, z. B. eine Stange Zigaretten über eine Woche hinweg zum Probieren. Damit wird pro Person eine Verminderung der intra-individuellen Meinung zu dem Produkt, dieser Zigarette herbeigeführt. Damit aber muss die Verlässlichkeit der Aussage erheblich steigen.
2. Er hat sich nicht für die Meinungen von Individuen oder von kleinen Gruppen interessiert, sondern hat für den Produkttest immer so umfangreiche Gruppen benutzt, dass damit die inter-individuellen Unterschiede ebenfalls möglichst repräsentativ, also wie in der Realität vertreten sind.

Auf diese Art und Weise gelang es ihm, selbst in schwierigen Bereichen zu Ergebnissen zu kommen, die damals für nicht möglich gehalten worden waren.

An diesem sehr einfachen Fall, der für die Marktforschung aber eine erhebliche Bedeutung bekommen hat, lässt sich diese Arbeitshypothese von Wendt illustrieren. Es gibt auch in der Medienforschung Experimente, welche auf dieses Phänomen der Unterschiedlichkeit der Antworten der Befragten zu ein und demselben Sachverhalt über kurze Zeit hinweg demonstriert worden sind.

In einem dritten Punkt kann ich teilweise Übereinstimmung konstatieren, und diese bezieht sich auf die Verwendung des Begriffes „Validierung“. Es mag sein, dass dieser Begriff aus der Test-Psychologie stammt, und wir ihn in der Marktforschung nur sinngemäß oder in nicht sehr deutlich bedachter Art und Weise gebraucht haben.

Wenn ich die Wendt'sche Vorstellung richtig verstanden habe, und er die Testtheorie in verkürzter Form auch richtig dargestellt hat, so geht es dabei darum, dass ein durch einen Test ermittelter Sachverhalt für die einzelne Testperson auch zutrifft.

Nun hat man insbesondere in der Motivforschung der 50-iger und 60-iger Jahre, zum Teil aber auch noch viel später die Vorstellung gehabt, dass die Auskünfte von Befragten über sich selbst, oder die aus der Auskünften abgeleiteten Daten zu einem Individuum tatsächlich zutreffen. Es gab bei einer Reihe von Marktforschern, zu denen nach meiner Kenntnis Herr Wendt gehörte, so wie ich auch, welche mit dieser Betrachtungsweise nicht glaubten, viel anfangen zu können. Wir waren immer der Meinung, so stellt es sich mir in der Erinnerung dar, dass es bei den Aussagen über relevante Gruppen in der Bevölkerung geht - und nicht um Individuen.

Um dies an einem Beispiel zu illustrieren, das ich nicht mehr konkret zitieren kann, das aber von einem Kollegen genannt worden ist:

In einer Untersuchung, die auch aus zwei Wellen bestand,

- wurde in der ersten Welle eine Kaufabsicht für irgendein Produkt von z. B. 50 % festgestellt;
- bei der zweiten Befragung stellte sich heraus, dass tatsächlich etwa 50 % aus ursprünglich nicht die Absicht habenden Leuten und solchen die die Absicht gehabt hatten, zusammen setzten.

Im Sinne der Validität der individuellen Aussage war dies ein höchst unbefriedigendes Ergebnis, weil sowohl die eine, wie auch die andere Gruppe je zur Hälfte etwas tat, was der ursprünglichen Aussage nicht entsprach. Vom Standpunkt der statistischen Betrachtung von Gruppenaussagen ist dieses Ergebnis keineswegs beunruhigend,

denn angesichts so vieler möglicher intervenierender Variablen zwischen dem ersten und dem zweiten Zeitpunkt kann es durchaus vorkommen, dass die Meinungen und Absichten in der Zwischenzeit verändert wurden, nach der einen, wie nach der anderen Seite.

Wenn wir die Dinge so betrachten, dann stimme ich mit Wendt darin überein, dass die Validierung im Sinne der Testpsychologie die Markt- und Mediaforschung vermutlich nicht grundsätzlich weiter bringen wird.

Aber Halt! Als diagnostische Überprüfung der Methoden ist der Versuch der Validierung im Einzelfalle wahrscheinlich nicht verzichtbar. Ich denke hier besonders an die „Studies in Readership“ von William Belson. Er hat, wie wir wohl einstimmig feststellen würden, mit seinen Bemühungen, die Richtigkeit der Antworten und die Gründe für Diskrepanzen bei den einzelnen Befragten der Leserschaftsforschung sehr wesentliche Hinweise gegeben. Und es gibt ähnlich gelagerte Untersuchungen in Deutschland, die ebenfalls wichtige Hinweise geliefert hatten. Konsequenzen daraus zu ziehen, heißt jedoch, dass man sich nicht mehr auf Individuen konzentriert, sondern auf relevante statistische Kollektive.

Der Begriff der Validierung ist jedoch auch in der Mediaforschung schon sehr früh verwendet worden, wobei die Richtigkeit der individuellen Aussage letztendlich zurücktrat hinter der Betrachtung der Daten, welche von dem jeweils an dem Test teilnehmenden Kollektiv geliefert worden.

Die erste mir bekannt gewordene Untersuchung (was nicht ausschließt, dass es frühere gibt) stammt aus dem Jahre 1941: damals berichtete der Curtis-Verlagsmarktforscher Ludeke über experimentelle Verfahrensweisen, welche das Ziel hatten, bestimmte Methoden zur Ermittlung der Beachtung von redaktionellen Beiträgen zu überprüfen.

Wir finden ähnliche Versuche bei den berühmt gewordenen Experimenten von Alfred Politz, in denen die von ihm und seinen Mitarbeitern entwickelte Fragestellung zur Ermittlung der Doppelseiten-Kontakte durch ein experimentelles Verfahren überprüft wurde. Auch in diesem Falle konnte man durch das Experiment die Realität recht genau darstellen und diese Realität der Heftnutzung und des Aufschlagens von Doppelseiten mit den Antworten der Befragten vergleichen. Auch hier zeigten sich im Bereich der teilnehmenden Individuen mehr oder minder große Diskrepanzen. Jedoch auf der Ebene der Gruppe stellte sich heraus, dass die Fehler auf der einen oder anderen Seite also der Über- oder Unterschätzung des Verhaltens sich die Waage hielten. Dies betrachtete Politz als eine Validierung der Fragestellung, und weder habe ich seinerzeit gezögert dies so zu betrachten, noch sehe ich heute einen Grund, hierfür den Begriff Validierung nicht mehr zu verwenden.

Wenn auch Friedrich Wendt in diesem seinem Vortrag den Begriff der Validierung offenbar als nicht besonders nützlich betrachtet, glaube ich aufgrund nun fast 30-jähriger Bekanntschaft und Kenntnis seines Standes sowie Sachverstandes, dass er im Grunde mit mir einer Meinung ist.

### **Kennen wir die Realität?**

Viel bedeutsamer scheint mir das Problem zu sein, welches darin besteht, dass selbst die gescheitesten Mediaforscher, wie beispielsweise Alfred Politz, für repräsentative

Untersuchungen keinen direkten Bezug zur Realität und damit keine direkte Validierung der Ergebnisse herstellen konnten. Und es ist sicherlich Eines, in einer experimentellen Situation die Beziehungen zwischen tatsächlich beobachtetem Leseverhalten und späterer Befragung zu überprüfen, und es ist ein Anderes, dies in einer großen Repräsentativ-Erhebung der üblichen Art zu versuchen. Letzteres ist nach meinem Wissen bisher nicht gelungen.

Auch Alfred Politz hat sich damit begnügen müssen, einerseits alle wesentlichen Vorkehrungen zu treffen, um von der Stichprobe der Ausschöpfung und der Erhebungsmethodik her Ergebnisse zu liefern, die so gut sind, wie man sie zur Zeit liefern konnte. Und zum anderen hat er besonders große Anstrengungen unternommen, um hinterher, bei Vorliegen der Ergebnisse, den Nachweis zu führen, dass die Untersuchung in sehr vielen, sehr wichtigen und möglichst relevanten Punkten die aus anderen Quellen bekannte Realität reflektierte.

Ein charakteristisches Beispiel ist die Untersuchung „A Study for Four Media“ für die Zeitschrift LIVE, in der auch andere Zeitschriften und sonstige Medien untersucht worden sind. Die Zeitschrift LIVE hatte damals einen Zeichentrickfilm hergestellt, der sich in sehr amüsanten und ebenso sachlich fundierter Art und Weise nur mit der Untersuchungsmethode und den Grundlagen befasste; er erhielt also keinerlei Ergebnisse.

Und in eben dieser Untersuchung hat Politz im Methodenteil nicht nur das sorgsame Vorgehen sehr ausführlich und offen dargelegt, sondern auch alle möglichen Prüfungen angestellt, um die Richtigkeit der Ergebnisse zu untermauern. Alle diese Daten bezogen sich auf Sachverhalte, die etwas über die Repräsentativität der Stichprobe aussagten, mehr oder minder einen Einfluss oder eine Beziehung zum Medienkonsum haben konnten; und es gab nur einen einzigen Punkt, der mit dem Medienkonsum also der Nutzung oder dem Lesen der Zeitschriften direkt einen Zusammenhang hatte: das war die Schätzung der Auflagen aus der Untersuchung im Vergleich zu den tatsächlichen Auflagen wie sie von der Prüfstelle ermittelt worden waren.

Demographische Merkmale und Merkmale wie der Besitz von Autos oder von bestimmten Automarken sind Indizien dafür, dass die Stichprobe stimmt; oder umgekehrt: wenn wesentliche demographische Gruppen oder das Vorhandensein bestimmter Gebrauchsgüter oder Besitz über offensichtlich falsch repräsentiert sind, dann besteht Grund zu der Annahme, dass das damit korrelierende Mediaverhalten auch teilweise falsch dargestellt worden ist.

Aber dies bezieht sich immer nur auf den Aspekt der Richtigkeit der Repräsentanz der Stichprobe. Die Schätzung der Auflagen geht hierüber hinaus. Sicherlich ist hiermit auch ein statistischer Aspekt verbunden, wie es auch von Politz deutlich gesagt wurde:

- wenn die Untersuchung hinsichtlich der Auflagen irgendeine Schiefe aufweist, so ist anzunehmen, dass diese Schiefe in der Stichprobe sich ebenfalls als eine Schiefe in der Mediennutzung zeigen müsste.

Aber diese Prüfung geht nach Auffassung von Politz, die ich teile, darüber hinaus: er sagte, dass wenn die Befragten zu ihrem Kauf oder Abonnement falsche Auskünfte geben, sie also über- oder unterschätzen, dann ist zu erwarten, dass diese Tendenz sich auch hinsichtlich des Medienkonsums, also des Lesens zeigen würde. Er sagte und dies kann man wohl nicht bezweifeln, dass es kein Merkmal in einer Untersuchung der bisherigen Art gibt, welches mit dem Lesen von Zeitschriften enger zusammenhängt als Kauf und Abo.

Es ist außerordentlich betrüblich, dass keine ernsthaften Versuche gemacht worden sind, in diesem Sinne wenigstens teilweise die ermittelten Leserschaftsdaten zu überprüfen. Ja, es geht noch weiter: in der deutschen Media-Analyse gibt es seit Jahren eine Frage danach, woher die Befragten die zuletzt gelesene Ausgabe einer Zeitschrift bezogen haben. Dieses Ergebnis wird auch ausgewertet, jedoch nur im Allgemeinen, als einen Hinweis auf die Bindung der Leser zur Zeitschrift betrachtet: Abonnenten oder Käufer haben eine engere Beziehung zur Zeitschrift als Mitleser oder gar gelegentliche Leser von Lesezirkelexemplaren im Wartezimmer.

Aber schon vor einer ganzen Reihe von Jahren hat der Heinrich Bauer Verlage diese Daten einmal unter die Lupe genommen und dabei festgestellt, dass die Aussagen bezüglich Kauf und Abonnement zusammengenommen und hochgerechnet von den Auflagen der Zeitschriften mehr oder minder stark abweichen, und zwar in anscheinend charakteristischer Weise.

Im Auftrage des Bauer Verlages haben wir dieses Phänomen mit zwei Untersuchungen der AG.MA. weiter untersucht und ganz bemerkenswerte Zusammenhänge zwischen den Fehlschätzungen der Auflagen aus der Untersuchung und den Lesierzahlen aus dieser Untersuchung gefunden. Dies ist den zuständigen Gremien der AG.MA. und übrigens auch Herrn Wendt zur Kenntnis gegeben worden. Es hat jedoch nicht den Effekt gehabt, dass man sich mit diesem Phänomen nun eingehend und mit verbesserter Fragestellung intensiv befasst, sondern ganz im Gegenteil: vom Jahr 1984 an wird diese Frage gänzlich weggelassen. Politisch ein voller Erfolg, denn nun können missliebige Ergebnisse nicht mehr produziert werden. Sachlich oder anspruchsvoller gesagt, wissenschaftlich ist dies ein Schlag ins Gesicht.

Wenn man die verschiedenen Versuche in der Medienforschung zur Überprüfung der Korrektheit der Angaben der Befragten auf bestimmte Fragestellungen hin überprüfen lässt, so hat man den Eindruck, dass es doch möglich sei, auf einem etwas verzwickten Wege zu besseren Überprüfungen und hoffentlich zu besseren oder verlässlicheren Ergebnissen zu kommen.

Es muss ja nicht sein, dass die jetzt ermittelten Ergebnisse samt und sonders falsch sind, sondern es besteht nur die berechtigte Befürchtung, und durchaus in weiten Kreisen der Nutzer der Daten, dass die Ergebnisse möglicherweise teilweise verkehrt sind. Und es wäre immerhin beruhigend, wenn man wüsste, *ob* sie falsch sind, *welche* falsch sind, und welche *richtig* sind. Und dann könnte man gleich daran gehen, um sie verlässlicher und richtiger zu gestalten.

Der Weg ist eigentlich vorgezeichnet. Es hat, wie schon gesagt, auch in Deutschland Experimente gegeben, um die Beziehung zwischen dem beobachteten Leseverhalten in einer experimentellen Situation und mit den Ergebnissen einer später erfolgten Befragung in Beziehung zu setzen.

Die bekannteste Untersuchung dieser Art wurde vom Hamburger Psychologischen Institut durchgeführt. Sie hatte damals zur Folge, dass größere Titelformen zur besseren Identifikation des einzelnen Titels verwendet wurden. Auch davon ist man abgekommen und verwendet jetzt ganz kleine Titelformen.

Ausgehend von einem Experiment oder einer Serie von Experimenten wäre es das Ziel verschiedenartige Fragestellungen zu überprüfen, und zu sehen, ob irgendwelche davon, oder weiter verbesserte in der Lage sind für die unterschiedlichsten Zeitschriften

die Realität in gleicher Art und Weise zu reflektieren. Dies muss nicht für jedes Individuum korrekt sein, und es muss auch nicht vollkommen korrekt sein in der Höhe der Prozentangaben, sondern es kommt vielmehr darauf an, dass die Relationen der Zeitschriften untereinander der Realität entsprechen.

Man würde ja mit einem konstanten Fehler oder einer konstanten Missweisung durchaus leben, und diese ggf. kompensieren können. Solche Experimente sind in der letzten Zeit überhaupt nicht mehr durchgeführt worden. Man hat nur noch Experimente durchgeführt, bei denen man verschiedenartige Methoden miteinander verglich, ohne den Versuch zu machen, diese in irgendeiner Weise mit dem tatsächlichen Leseverhalten in Beziehung zu setzen. Solche Methodenexperimente haben natürlich auch ihren Wert, vor allen, wenn es darum geht, das überlastete Interview zu verkürzen, komplizierte Abfolgen und Fragen raus zu werfen, um dann zu sehen, ob in etwa dieselben Ergebnisse wie mit den bisherigen Verfahrensweisen erzielt werden. Nur darf man dann nicht den Anspruch erheben, wie es Scheler tut, dass die neue Verfahrensweise, die etwas gefälliger ist, „validere“ Ergebnisse zu liefern im Stande sei.

Es hat solche und ähnliche Experimente in der ganzen Historie der Agla/AG.MA. gegeben. Ein besonders bemerkenswertes Experiment war das von ASK–Knuth durchgeführte, das sich auf verschiedene Arten von Titelvorgaben bezog. Dieses Experiment schwebte ebenfalls im luftleeren Raum, wie Knuth in seinem Kommentar mehrfach ausdrücklich feststellte. D. h. es gab keine externen Erkenntnisse auf Grund derer man die erzielten Resultate hätte bewerten können. Stattdessen begnügte man sich expresso verbis damit, jene Titelvorgaben als „besser“ zu betrachten, die in möglichst gleichförmiger Weise über alle einbezogenen Zeitschriften hohe Ergebnisse lieferte.

Als politisches Kriterium ist so etwas natürlich verständlich, und von den Verlagen her gesehen durchaus akzeptabel. Es sei denn, dass auf diese Weise gewisse Titel eine zu hohe Reichweite zugebilligt bekommen. Bei diesem Experiment fiel interessanterweise ein Verfahren hinten rüber: Bei diesem wurden die Titelzüge der Zeitschriften auf einem Tableau vorgelegt.

Die Ergebnisse für diese Methode fielen durch starke Abweichungen von allen anderen Methoden auf. Sie wurde abgelehnt. Und bei einer gänzlich anderen Gelegenheit begutachtete Dr. Klaus Landgrebe:

*„Aus Experimenten zur optimalen Gestaltung von Identifikationshilfen bei Media-Analysen ist bekannt, dass bei der Vorgabe von Marken auf Tableaus die Befragten die einzelnen Marken nicht gleich stark beachten. Breitet man mehrere Untersuchungsgegenstände auf einer Fläche aus, dann wirkt jede Frage nach diesen Gegenständen wie eine Aufforderung zu einer Art Greiftest oder Deutetest, auf jeden Fall entsteht ein Reiz zur Auswahl ...“*

Die ist eine Interpretation, welche bei bestimmten Fragestellungen und Zusammenhängen ihre Berechtigung haben mag. Bei der Ermittlung von Leserschaften erscheint sie mir weniger plausibel als eine andere Interpretation der damaligen Ergebnisse. Wenn man nämlich die damaligen Ergebnisse auf Mittelwerte und Streuungen untersucht, - was damals unterlassen wurde -, so stellt sich heraus, dass die Darbietung der Titel auf dem Tableau zu einer deutlich höheren Streuung der Ergebnisse geführt hatte. Und daraus schliesse ich, dass die Befragten gerade durch das Nebeneinanderliegen der Titel viel besser in der Lage waren, zwischen den einzelnen Titeln zu unterscheiden, sich zu orientieren, und sich ggf. selbst zu korrigieren.

Es gibt in der Markenartikelforschung ein Ergebnis, welche diese meine Interpretation zu stützen scheint. Und zwar hat unser Kunde Colgate-Palmolive die Zuordnung von Eigenschaften zu Zahnpasta-Marken einmal in einer Omnibus-Befragung mit Interviewern, und parallel dazu als schriftlich-postalische Befragung in unserem Panel durchführen lassen. In der Interviewer-Befragung wurden die Marken nacheinander anhand der Eigenschaften beurteilt; in der Panel-Untersuchung standen die Marken und Eigenschaften auf einem Tableau. Es mögen noch andere Einflussfaktoren wirksam gewesen sein, um Unterschiede bei diesen beiden Methoden hervorzurufen.

Als wesentlichste Ergebnisse kamen heraus:

1. Die Niveau-Unterschiede waren erheblich, aber die Korrelationen ziemlich hoch. D. h., es wurden bei beiden Verfahren grundsätzlich dieselben Sachverhalte ermittelt.
2. Bei der schriftlich erfolgten Befragung im Panel, mit dem Tableau von Marken und Eigenschaften, war die Differenzierung zwischen den Marken sehr viel schärfer als bei der Interviewer-Befragung, wo ja die Marken einzeln für sich abgefragt wurden.
3. Auch die Differenzierung innerhalb der Marken war wesentlich schärfer bei der schriftlichen Befragung als bei der Interviewer-Befragung.

Diese schärfere Differenzierung in beiden Dimensionen hat uns dazu veranlasst, bei Interviewer-Umfragen, wenn es sich um eine Differenzierung handelte, ebenfalls mit Tableaus zu arbeiten. Dies bezieht sich zum Beispiel auf die Zuordnung von Funktionen zu Architekten- und Bauingenieur-Fachzeitschriften in der „agla a+b 1978“. Dies hatte zur Folge, dass die Zeitschriften in außerordentlich klarer Weise von einander unterschieden wurden. Und sowohl Verlage wie Inserenten wie auch wir hielten aufgrund der Kenntnisse dieser Fachzeitschriften diese Ergebnisse für korrekt, also nicht etwa in irgendeiner Weise verzerrt.

Meine Interpretation der damaligen Ergebnisse der ASK–Knuth Untersuchung erhebt nicht den Anspruch eines Dogmas. Es würde mir auch nichts ausmachen, wenn zu einem gewissen Prozentsatz Dr. Landgebe's Interpretation, zu einem anderen Prozentsatz meine Interpretation, und zu irgendwelchen Prozentsätzen noch andere Interpretationen richtig sind, - d.h., es kann durchaus verschiedene Möglichkeiten der Reaktionsweisen der Befragten geben.

Alles, was ich in diesem Zusammenhang ausdrücken wollte, ist folgendes: wenn man Experimente veranstaltet, bei denen ein Bezug zur Realität nicht herzustellen versucht wird, dann sind sachliche Interpretationen nur mit beschränkter Haftung möglich. Mit wissenschaftlicher Forschung hat so etwas nichts zu tun.