Friedrich Wendt
Arbeitsgemeinschaft Media-Analyse
Frankfurt, West Germany

8.3 The AG.MA model

ORGANISATION

The Arbeitsgemeinschaft Media-Analyse was founded in 1954. At the beginning a survey was conducted every two years but since 1960 a study has been published each year. At first, only data for magazines were collected. Other media, such as radio and television, newspapers, supplements and cinema were brought in somewhat later — and the number of magazines measured grew steadily. By now, we have what is nearly an all-media analysis.

The members of the AG.MA are the media (publishing houses, the radio and TV stations), nearly all big advertising agencies (and some middle and smaller ones), as well as a certain number of advertisers.

THE AG.MA'S AIM

The aim of the AG.MA is to deliver a high standard, appropriate and acceptable measure of media consumption on which to base media planning. The basic demand is the dissemination and comparability of valid data for as many media as possible.

The great number of members — about 130 — and the large variety of their interests between them guarantee the development of the best methods available — the best for all, with nobody being damaged.

From the beginning the members' specialist staffs have worked for the realization of the AG.MA aims by permanent cooperation. That is, of course, not easy to do, and is very expensive, especially as over the years the members have asked for many, many experiments.

The main problem has been to gain an overall acceptance, a sort of validity. From very early days in this connection the whole scientific background — not only of media research, but also of our whole profession of market research — has been debated. The polarity of "reproduction of the reality" against "usefulness" has been a matter of permanent discussion.

The result of all these efforts is a valid currency of media consumption, the official German media survey.

The AG.MA is a tool of media planning, which means that it is concerned with the future. Whatever may have happened in the past in the field of media consumption within a given target group, we have to find a way which allows us to forecast the corresponding situation of tomorrow. For this purpose, it is, of course, not necessary to have a precise reproduction of the past. On the contrary: we know with certainty that a simple projection from the situation of yesterday to that of tomorrow — a linear extrapolation with a change rate of zero — is in any case wrong. It is not only not helpful, but is also misleading. Essentially we have to find a forecasting system in which the survey may be located. The survey itself, used in a naive manner, is never sufficient.

WE NEED A MODEL

Unfortunately or not we have — in general, and in particular in our field — no access to reality. Even when discussing our first studies, the readership surveys up to 1960, we arrived at one important consensus: we cannot measure facts. What we get from the interviewees are opinions they have about their own media consumption, not direct information about their real media consumption. Their claims are only their judgements, their images of their own behaviour.

The only thing we are able to do is to lay a net over the reality which gives us the essential of the phenomenon 'media consumption' over a given time period which for reasons of safety should not be too long. That is, in fact, our modest, but realistic pretension with respect to our model.

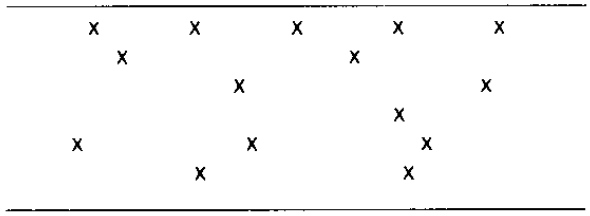
We ourselves have neglected a model in the macro field, as for instance the Linear Programming Model — mainly discussed in the United States in the early sixties — in which advertising effectiveness was measured by calculating in terms of aggregate units. In place of that we decided to use a micro model with the individual as unit. The consequence was that we have had to consider the whole population of the Federal Republic of Germany and West Berlin.

Two general restrictions within the model

Because of the model's purpose we had to restrict ourselves to two general directions:

(1) we reduce the multitude of all characteristics of an individual to a fairly high, but restricted, number of well defined attributes that we want to measure. It remains a big problem to select these appropriate variables, but this is another matter.

Consequently, the model consists of an n-dimensional vector space — with as many dimensions as we selected variables. Or we may say: the model consists of an n-dimensional point space in which — due to the great number of variables — no two individuals are to be found in the same place (not even twins).



Further, the points — our individuals — are not distributed equally over the whole space. There are clusters or clouds of similar individuals which are nearer to each other than to members of other clouds.

(2) the second type of restriction arises from the fact that it is impossible for us to collect information from every individual in the population. That means that for practical work in market research we have to draw samples from the whole population. This is, indeed, a very important part of the model, as it introduces a specific sort of uncertainty into our conclusions — the statistical confidence limits — which can never be neglected.

THE THREE 'PILLARS' OF THE MODEL

Against these general remarks we present the AG.MA model, as a mixture of philosophical background, of some mathematical traits and of the description of the rather complex reality of the AG.MA workshop. We distinguish three pillars within our model:

(1) the field model with the sampling system (this is the second restriction mentioned confronting the reality, so that we may be able to handle representative statistics);

(2) the survey domain;

(3) the data processing field; (these two being the consequences of the first restriction which led us to the n-dimension vector space).

The field model

When starting with the field model we have to distinguish representativity in the three dimensions of space and the one of time.

THE REPRESENTATIVENESS IN SPACE — THE SAMPLING PROCEDURE

In this area, Germany has gone its own way, though we got the first hints about good experiences with area sampling procedures from Politz and Deming. After many years of individual sampling systems — though working in this direction — we have had available since 1970 a uniform sampling system for practically all the big market research institutes (called the Club of 13 because of the number of members, including the AG.MA). It was the AG.MA that initiated and financed the starting phase of this project. We now have the third edition at our disposal, as a renewal of this system takes place every four years.

The area units are about 55,000 polling districts (for election to the Bundestag and the Assembly of Western Berlin) the geographical boundaries of which are available. They define the potential of all sampling points. After some geographical stratifying procedures in the first stage, 210 polling districts are selected at random as sampling points (according to the number of private households they contain) for each of the 'sampling nets', as we call them.

The main point in drawing these nets (each of which is a separate sub-sample of the total population) is that all nets for all participating institutes are selected at the same time by one single procedure. They are then allocated to the institutes, so that there exists no overlapping from one net to the other and no overlapping from one institute to another. All nets are disjunct — with, in consequence, a practically unrestricted possibility of combining nets coming out from the system arbitrarily.

The AG.MA uses 96 nets in 4 years forming a larger unit of representativeness, 24 nets per year, 6 nets per wave. Thus there exists 4 waves per year, although our fieldwork is continuous over nearly the whole year (from January 10th to December 15th). The 6 nets per wave are allocated to 6 market research institutes in an overall optimal way.

We are convinced that, since every market research institute has its own handwriting, its own operating system, the best solution is to employ several institutes (6 as mentioned) for the fieldwork to balance these variations.

The sampling points of one net are distributed proportionately to the density of households over Germany. By accumulation of those nets it is, of course, possible to improve this density of households up to any desired degree of representativeness.

For our purposes it is considered to be sufficient to have — for representativeness as well as for economy — 24 nets of 210 sampling points per year, which means 5,040 sampling points, with 4 addresses. Thus we have 20,160 households at our disposal.

In two regions — Bremen and Saarland — the interviews are augmented disproportionately, because otherwise these would not get enough cases for certain media.

8.3 The AG.MA model

The second stage of building the sample consists of drawing 4 households at random per sampling point. This equal number of households per sampling point is legitimate, since the polling districts — as sampling points — have been drawn proportionately to their number of households. In order to find the 4 households in every selected polling district, the institutes are obliged to list all households in the sampling point according to strict rules about sequence. This is really the point where we take into account all three dimensions of our space: the streets, the houses, the floors and the household-doors on the floors build a well defined system for listing the door-plates or door-bells. The institute gives the interviewer 4 field addresses out of the stock of all addresses drawn at random.

The last stage of random sampling (to find the persons to be interviewed in the household) is done by the interviewer himself, again according to strict rules — a permutation system called in Germany for historical reasons "Swedish key". Only persons of 14 years and older are interviewed. But once more our sampling system is a disproportionate one. Only one person is interviewed in households of one to three persons (14+ years old); in households with four and more persons two individuals are selected and interviewed.

We have to think about this system for the third stage of our sampling procedure. The reason is quite clear. The consequence of our disproportionate system of sampling is that we have to weight the sample. It is not sufficient that every individual has a well known and calculable chance of being drawn. In order to be able to interpret the results of the survey, we have to equalize those different chances to get the same degree of representativeness for every interviewee. Only in this way is it possible to draw conclusions from the sample to the universe.

There is another well known problem of all random sampling surveys — non-response — to which we refer later.

REPRESENTATIVENESS IN TIME

First, we must treat representativeness in time to complete the field model. As already mentioned, the AG.MA gets the data from a survey carried out over the whole year. The model is not dynamic or longitudinal in the sense that it represents a time series. Reflections and considerations exist in this direction — and some steps have already been taken. And the prognostic aspect does provoke such a demand. But so far we have only started.

Those who want to look into the future have to examine the past. In practice, every user of the data compares the most recent results with those of the last year or even longer ago. However, our surveys in their actuality do not give appropriate information for this. Following the necessity always to consider the latest theoretical knowledge in media research, we have always had to decide whether continuity or accuracy should be preferred or, to put it more correctly, where we may find an optimal balance between the two.

As advertising campaigns are planned in time, media planning in the end also has to be done in this way. This gives, of course, the idea of media research as a forecasting instrument.

What we have done in this respect is to spread the interviews equally over the whole year. Nevertheless, there exist limitations. We do not dispose of an exact sample of days, but we try to distribute the interviews fairly over the week. The day of the first visit is fixed. The second visit is made the next day but one. Then the same procedure takes place in the next week, if the interviewer did not succeed in getting an interview. Even a fifth visit may be possible. This is the 'first effort' of visits. If all these endeavours have been in vain, the whole procedure is repeated after four weeks. The maximum number of visits in this way can amount to 8 visits in general — and may continue up to ten!

NON-RESPONSE

By virtue of these efforts we have a non-response rate of about 15% to 20% as maximum. For a fairly good regional representative sample we try to get non-response equally distributed over the whole country.

Approaches like those of Hansen-Hurwitz and others are no more than theoretical crutches. We have conducted experiments in this field, but they cost time and money more than they give practical help. We have to live with this problem in empirical social research, but we have to make further efforts.

There are certain optical methods of dealing with this phenomenon using a limited number of characteristics — the distributions of which, for instance, exist in the Official Statistics — the gaps are filled up either by the interviewer (quota selection) or by a subsequent adapting process. Both procedures suffer from a false hypothesis: that the persons interviewed have the same behaviour with regard to the characteristics not controlled (for instance media consumption) as the persons not reached. And in quota selection there is one more factor, the subjective selection conduct of the interviewer within the cells controlled, irrespective of the doubts of the mathematical statisticians against quota.

From the very beginning the AG.MA decided in favour of a pure random sampling system, and for non-response very soon followed this with an adapting process to data of the Official Statistics. Formerly, punchcards were doubled or removed. Today the process is one of factor weighting in a more elegant manner. This process is called "redressement". It is an

iterative process which allows simultaneously the adoption of the most developed technique in this respect. However, it does not alter anything: there is still the false hypothesis and the operation remains only a cosmetic one. Further, the data in the Official Statistics are very often out of date, subject to a high degree of uncertainty.

The AG.MA is therefore making new experiments on how to complete the sample in a more appropriate way. We call this procedure sampling point weighting. This is a weighting procedure by which the individuals reached in one sampling point render their information to those who could not be reached. We are just approaching the end of the experimental phase, and we already know that the sum of weights — ie. the degree of influence resulting from *this* procedure — is much less than with *any other* procedure, and that by avoiding any exterior influence no artificial or unforeseen distortion is produced.

This ends the first part of the description, the first pillar of our AG.MA model. Now we pass to the second restriction, ie. from the practical unlimited space of characteristics we meet in our human daily life to a well defined n-dimensional space.

The data processing pillar — model reflections

In order to develop our model we searched for an appropriate theoretical base and found it in a particular domain of mathematics — topology. There are several reasons for the use of this instrument, two of them we shall mention immediately.

TWO REASONS FOR THE TOPOLOGICAL APPROACH

(1) The first one results from the nature of the variables describing the individuals. They are of different nature: quantitative, qualitative in categories, ordered and not ordered, and we have to use them all. So a restriction to the metric case is impossible: we have to introduce a topological space.

(2) The next reason comes from the notion of an environment. This must be explained further. The topology or the point-set-theory has many fathers. One of them is the Dutch mathematician L.E.J. Brouwer, born just 100 years ago. He called topology a rubber geometry. And in fact that is what we need.

We started with the assumption that every individual is represented in our model by a vector or a point in the n-dimensional space of variables. But this is no longer true — or at least it is not sufficient. In our model we have to assume that every point is more or less attenuated, spread out in an environment which we cannot demarcate sharply. Why we think that this way to see the individuals is appropriate has to be explained.

No access to the reality

At the beginning we observed that we can measure only opinions, not facts, and we stated that, in principle, the reality is hidden for us. We have good reasons to make this statement.

(1) Firstly and in general, because we can detect with our senses only inexact and partial information, which is combined by our brain to a complete image. Consider the phenomenon of the cinema, where a series of still pictures gives the illusion of movement, through the inertia of our visual capability. Our human perception apparatus is an imperfect instrument.

(2) It is very simple to formulate the general requirement of media planning: efficiency attributable to the media. But how are we to realize it?

In a survey we put questions to the interviewees and we get answers. But the interviewee is conditioned by the whole interrogation process. To explain what we mean, let us go step by step.

First of all, we start with the idea that the efficiency of advertising is a function of several factors including media consumption. That gives us a theoretical formula:

$$\underset{\text{efficiency}}{E} = f\{x_1 \ldots\ldots \underset{\text{media consumption}}{MC} \ldots\ldots x_n\}$$

In current practice we ask for media consumption. But we get an answer which itself is influenced by several factors, only one of which may be media consumption. A practical formula concerning the real answers we get may then be:

$$A = g\{y_1 \ldots\ldots MC \ldots\ldots y_m\}$$

What influences these answers? One set of influences is caused by the interviewer, some of them by the interviewee himself or herself, others by the relationship between both, the mood of the one or the other, their quiet or hurry. And there are external factors: the weather, the presence of other family members and so forth may also influence the answer.

On such a vague base we are obliged to transform the result of the interview (the answer) into an estimate of media consumption — and that is the only thing we have at our disposal:

$$\widehat{MC} = h\{z_1 \ldots\ldots A \ldots\ldots z_r\}$$

and later on to introduce this into the efficiency expression:

$$\widehat{E} = f\{x_1 \ldots\ldots \widehat{MC} \ldots\ldots x_m\}$$

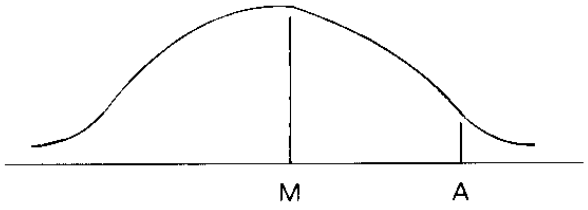
which is now no longer a theoretical, but a practical

formula. Due to these circumstances it is no longer a matter of 'reality or 'precision', when we derive a measure of advertising effectiveness or only of media consumption out of a survey. It is elastic!

But we are not yet at the end of our series of conclusions. For to understand the role of the notion of environment in the inquiry process we have to consider the answer to one question about one characteristic.

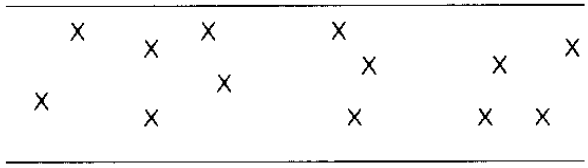
It is rubber — or a detour to come to the mean

The answer possibilities of the interviewee, for every characteristic, are subjected to a distribution — which we may regard as an intra-personal one — with a mean and a standard deviation valid only for this particular individual.

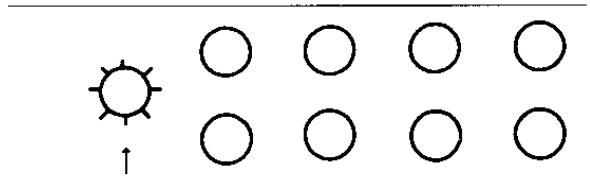


We have no chance to deduce from A — the effective answer — what is the intra-personal mean — M. We have to make a statistical detour. When we now try to get solid answers over a certain period (a limited time), in so far as media planning is valid in the immediate future, then we have to do our best to approximate the mean as well as possible, as a constant value over a not too long time period. We have learnt that this phenomenon is valid for *all* variables.

Our picture with fixed and well defined points



is not quite true. The points blow up to n-dimensional balls or ellipsoids with greater or lesser diameters. They are spread over their environment. Thus we do not know at all where the middle point of the individual is to be found.



This may be our individual in a sample

Nevertheless, however inexact may be the position of single individuals, there exists an environment, and we have the possibility of examining how near, how similar the individuals are — more or less inaccurately.

But as we have learnt that the individuals are spread over the space in an inequal manner, we have to find homogeneous groups or segments of individuals by means of, for instance, a typology analysis (not to be confused with topology). Homogeneous — that means mainly with respect to the media consumption. But, in principle, it may be valid for any other characteristic too.

In this way we are in a position to postulate an enormous proximity of the means of the members of a homogeneous segment of interviewees, in spite of all differences of their effective answers. We then take the mean of the group as the individual mean of all members of the group. In any case, for the group this postulation is true; with regard to single individuals we only have an approximation. However, we need only statistical statements for a sufficient large subgroup. And so we have found a solution of our problem through the environment notion of topology.

THE MEASURE OF MEDIA CONSUMPTION PROBABILITY — AND OF CONTACTS

We profit from these reflections to calculate, in practice, the measure with which we are able to make forecasts for media planning. A reasonable way to do so is to introduce probabilities and mathematical expectation values. The media planner wants to have, today, characteristics about the different media, deduced from information about yesterday, to be able to plan and prepare campaigns for tomorrow.

The central events are the exposures or contacts the persons have with the media, whatever may be the unit of an exposure (an average issue, for instance, or a page, or a spot). Of course, everybody wants to come as near as possible to the point where the advertising effectiveness arises. In a press medium this may be the page. But whether we take into account page traffic or only medium exposure, the essential remains in any case the same. The best adapted measure for calculating the contact distributions which represent a campaign or a media schedule is the media consumption probability.

Segmentation procedure

A good way to calculate these probabilities is to use the maximum of information available — and not only the simple claims of the interviewees about their media consumption. And as we have no direct access to the mean of the intra-personal probability distribution of an individual, we calculate it by means of a segmentation procedure following the Belson method). That is the way

to find homogeneous groups of media users.

In segmentation procedure the target criterion is always readership per issue, as we assume it to be the most actual and accurate information about the media consumption. We have about 300 media units for which we do such segmentations. Men and women are segmentated separately. Thus, we have to calculate every year about 600 segmentations with even up to 32 segments (the size of our sample is nearly 20,000!).

As discriminating variables we use, in general, socio-demographical characteristics and some others, such as the number of visits to attain the interview, the consumption frequency of the relevant medium, and the sum of frequencies of the media classification group (magazines, newspapers, radio, television).

Since introducing in 1971 the segmentation procedure for calculating media probabilities, we have examined experimentally the usefulness of the method. Even since that time we have made experiments concerning the discriminating formula, the characteristics, the stopping criteria and so forth. In addition, we are continually examining the segmentation procedure itself. To date, however, we have seen no reason for changing this procedure fundamentally.

But let us return to the practical aspect of the segmentation procedure. We get segments of homogeneous media consumption and we allocate a media consumption probability to every member of a segment which correspond to the share of readers per issue of this segment. The mean of the group becomes the individual mean.

Conditional probabilities

But there exists a big problem in the background. The media consumption probabilities of today are calculated independently — one title from the other. Only the frequency sums bring in information resulting from other media as well.

The hypothesis of independent media consumption is certainly false. Presumably conditional probabilities are more appropriate. The assumption of independent probabilities, however, is a practical and a useful one in the model, especially when it is possible to consider overlapping media consumption for the calculation of single or special media usage. We are optimistic about this direction and about typological calculations which allow us to group the interviewees, according to their multi-media consumption, in a more homogeneous way.

Apart from these more or less imminent changes in the method of calculating probabilities, there are further lateral ideas.

The dependent criterion for the advertising medium contact is still based upon the notion of the reader per issue — ultimately upon a concept of coverage which refers to the person and his ability to be a reader.

Fuzzy sets

The distinction between 'to be a reader' or 'not to be a reader' is an arbitrary one, made in an abrupt "yes" or "no" manner even when we are using a scale. Our reference points still neglect the fact that neither in theory nor in the measuring field are there clear and disjunctive limits for media consumption. We need only to change the definitions or survey methods to get deviations which are more or less large (mostly they are regarded as dramatic ones — or at least are put forward as such).

In view of this we have obviously reached a stage in media research where we must ask ourselves whether we do not have to revise our customary thinking.

The uncertainties involved in deciding whether an individual belongs to a particular group — to the readership of a magazine, to the audience of electronic media and so forth — are typical problems. We are already making use of this new way of thinking in the field of typological studies, in defining diffuse margin zones and the probabilities of belonging to certain types.

And we can not only look to topology, but can adopt a younger branch of mathematics that has been developed in the last 20 years and is still growing. This deals with what are called 'fuzzy sets'. Whenever it is not possible to grasp a phenomenon accurately, it is nevertheless possible to make precise statements about vague circumstances. We are not obliged to know yesterday's media consumption accurately, but with this "hazy" method we can find stability and hence again underwrite more the prognostic role of our media research.

The error model

Further to elaborate this aspect, we remind you of the fact that interviewees may commit errors as can everybody else, and in ways beyond those previously mentioned. When we tried to approach this phenomenon in 1960 we learnt to distinguish between confusions of time, confusions in the object (the medium) and confusions regarding the conditions or circumstances of media consumption. To overcome those problems, we copied Belson's work. We know what everybody else does: the result of our survey will be that some of those who claimed media consumption do not belong to the readership or audience, while others who did *not* claim media consumption are nevertheless readers or listeners or viewers.

But there is a way of estimating the degree of error:

statisticians will recognize the analogy with errors of the first or second type.

Let p be the proportion of interviewees who claimed media consumption (for instance in the last publication interval). Then the volume of media users, after having made a correction of error probability α (error of first type) and β (error of second type), would be:

$$A = \alpha (1 - p) + (1 - \beta) p$$

And we can see the possibility of calculating α and β, but have not yet finalised it.

Validation aspects — the philosophy behind?

We customarily do media planning on the probability values. They are the only measure on the AG.MA tape for the assessment of media consumption.

But what is our justification? Returning to the basic model idea, we postulate that there is a characteristic in the individual which shapes his media consumption tendency for every medium; a characteristic that remains stable over a certain time, and which may be interpreted as his probability of contact with the medium (at least on contact within the publication period — therefore a 'net' contact). This is a value between zero and one, and its relative constancy is derived from a computation method which considers all possible influencing factors (for the moment, the segmentation method).

But are these probabilities valid, are they really a net laid over the reality in the best-fitting way? Even if we accept that we have no access to the reality itself? The answer comes from rubber geometry.

The first to formulate this, in a very impressive and at the same time plastic way, was Hans-Erdmann Scheler, who remarked that — whatever we do — we are comparing one rubber-band with another one. Data resulting from panels are compared with frequencies, frequencies are compared with probabilities and so forth.

For some time we have been having intensive discussions about 'validation', though there have been many doubts as to the value of such discussion on the one hand and the notion of validation on the other. Is the expression 'valid' a synonym for 'being accepted'? What does 'validation' actually mean?

The term 'validation' comes from the psychological field, to evaluate psychological tests. In market and media research it is not possible to apply this tool in the same way, when considering the relatively enormous sampling surveys set up to provide a base for media planning. What is meant is the effort to meet the different factors of uncertainty in proceeding from collecting the data up to their computation. We have to try to do our best to get stable and robust answers, and that is difficult enough.

Thenceforth, the whole construction of models is thematic, so that we were obliged to reflect upon the whole system, which leads us to very general controversies of an almost theological nature.

My own opinion is, that as Thomas S. Kuhn has said we are facing a "scientific revolution". We are about to replace a paradigm that ruled over media research for decades by another one. This will give rise to a lot of new understanding, but there are many people who will keep their old habits of thinking. Maybe Max Planck will be right in the end with his dictum: "A new scientific truth is not accepted because its opponents are convinced, but because they die in due course and the next generation is confronted with the truth from the beginning".

THE BINOMIAL DISTRIBUTION

But we now have to say how we use these probabilities. For our calculations we had to search for a model of the distribution law of the p-values. At that time (more than 15 years ago) there was vehement discussion and even controversies about which distribution law would be the most appropriate.

Finally, English pragmatism led to the overwhelming acceptance of the Binomial model. With this, the way was free for the transformation of a media plan into an exposure distribution.

That was the moment when media planning involving the computer was introduced. From this development only rudimentary work remained to be done, as a short trip into the world of formulae will show.

For a target group consisting of individuals $i = 1...N$, a media plan may contain the media $j = 1...m$ with insertion figures of b_j $j = 1...m$.

Then every individual $i = 1...N$ has with respect to the medium $j = 1...m$ a media consumption probability p_{ij}. According to the Binomial Model there will be a probability to have k exposures which go from 0 to b_j per medium j:

$$W(k)_{ij} = \binom{b_j}{k} p_{ij}^{k} (1 - p_{ij})^{b_j - k}$$

Due to the assumption of independence these exposure probabilities add up over all media and all individuals to a total exposure distribution:

$$W(k) = \sum_{i=j}^{N} \sum_{j=i}^{m} w(k)_{ij} \qquad k = 0, 1, \ldots \sum_{j=i}^{m} b_j$$

This transformation of individual media consumption probabilities into total exposure distributions (for a target group) is the aim of media planning — the

purpose of the model of media research. Four remarks remain:
(1) in principle, instead of the Binomial distribution any other appropriate distribution could be used, for instance the Beta-Binomial distribution or the hypergeometric distribution. In Germany the hypergeometric distribution has been rejected once. Recently, the Beta-Binomial distribution is again being debated.
(2) it is hardly relevant which method is used for the calculations. In general, the most exact method of convolution is never applied. Simulations by the Monte Carlo method or approximations (as for instance those of Martin Beale) are the common way to handle the practical problem.
(3) the contacts are not taken as being equivalent. Weights for persons, media and/or contacts (in form of the so-called response functions) were introduced very early. Unfortunately, all this is of little use because the appropriate empirical data are missing.

On account of this practically all media selection programmes which *created* media plans have been rejected. But recently, discussion about the quality of media planning has started again: we have still to wait to see what consequences this will have.
(4) model ideas like the existing ones based upon exposure to the medium are likewise applicable to exposure to advertisements. The theoretical tools are at our disposal; even moves towards large-scale comparable surveys.

We have mentioned the efforts of Politz, and may take note over many years of private studies following the Politz method. Some years ago our Dutch friends proved that it is possible to measure page traffic in a comparable way for many magazines, and this was confirmed by experimental work under German conditions last year.

The AG.MA workshop

Before completing the data processing pillar of the AG.MA model we have to be sure that such an ambitious model is supported and highly-developed by an everyday AG.MA 'workshop'. We deliver to the members of the AG.MA not only printed reports in several volumes, but more importantly a magnetic tape containing all the information required. Every year it is a long, laborious process from the raw data coming from the institutes up to the fully audited final tape, including checking and correcting procedures, the weighting procedures such as the proportionalization of the sample, the "redressement" and the segmentations for the calculation of media consumption probabilities. In addition, we prepare special regional characteristics, so that further information from the Official Statistics may be used — for instance conurbations or geographical co-ordinates (the transversal Mercator System).

THE QUESTIONNAIRE MODEL

On the subject of collecting the data we can be relatively short. Most of our efforts of the last nearly 30 years have been focused on the questionnaire — naively, of course, and in the belief that we could approach the truth by questioning. Only gradually did we become acquainted with the idea of models.

In 1954 the survey started with readers per issue and regular readers — as everywhere. At the beginning of the sixties cumulation followed, and somewhat later frequency questions to determine the regularity of media consumption. We do not forget the role that Lester Frankel played in Hamburg in 1963 when introducing the base for this development.

By the end of the sixties the first fundamental pause for reflection began, after many weighty discussions and developments. During many working sessions, supported by extensive experimental work and with the assistance of all the big market research institutes (not only of external experts) the questionnaire model of the seventies has been developed and introduced. Only now are we again beginning to reflect upon the whole complex, and we have again conducted a number of experiments. Eva-Maria Hess, Hans-Erdmann Scheler and Klaus Landgrebe are reporting on developments in this symposium, which is why the subject is not treated more fully in this paper.

Of course, all efforts in the questionnaire field belong to the model. They are, however, more popular than all other components of the model, since everybody believes himself expert in this subject, and political aspects play a certain role in their discussions.

Filter questions, phrasing of questions, and so on, are much easier to understand than any other elements of the model — and are, indeed, of some central importance. They shape the answers which are the original substance. But that they are not the only factors has become more and more clear in the minds of the data users.

We need more and better information

Several views had to coalesce before another important step could be taken. On the one hand, there were the needs of the users for more information, and the AG.MA developed from a pure readership survey to nearly an all-media survey.

On the other hand, the demand for more information describing target groups grew steadily over

the years. For different reasons (and especially on account of divergent interests) the boards of the AG.MA resolved to restrict themselves to media information and to somewhat extended socio-demographical data. That gave competitors and other private institutions the opportunity to provide supplementary information. However, the disadvantages of a single-source survey could not be overcome.

Random answers are the least of our worries. Inadequate questions in single media fields are of greater importance.The electronic media have long been convinced from their own experimental work that there are much better possibilities. Even the newspapers and the magazines have had discussions about improvements in survey methods.

Databank problems

In the end this led to the conviction that we had to search for an optimal balance between the survey and the computation area. Thus methods of marrying data — like the fusion procedure — and other techniques have been taken into consideration, in order to set up a data-bank.

New ways in the media planning need also to be implemented: for instance a combination of decision procedures on the base of macro models (perhaps for the allocation of the budget between the main media fields, as magazines, newspapers, radio and TV), and for decision procedures further into the area of micro or individual models for the allocation of the sub-budgets to the individual media within the media classification groups.

Probably this will progress in the same way as in the data processing field as dialogue techniques are more and more used, so that we may assume that media research and media planning will go in the same direction.

And when we have reached this position the time will be near when media planning may be supported by special simulation procedures which will find their empirical base in a large data bank, including not only the time factor as model component as well as exposure qualities, but finally the whole marketing background, taking in competitors' efforts and other exogenous factors. Castles in the air? We have to wait and see.

PARTNERSHIP MODEL — KONPRESS

A "partnership model" may be handled and prove its worth only when we succeeded in getting the right balance between survey and computation methods (as already mentioned) including all such further considerations as the organisational, judicial and financial ones.

We have just begun to implement this. The first partner of the AG.MA was KONPRESS, the religious press, an association of about 43 more or less regional titles. A study of these titles has been conducted following the same survey method as the AG.MA and this survey has been married with the actual media research findings of the AG.MA.

Meanwhile the activities of the six media research institutes have been largely co-ordinated; for instance the questionnaire and sampling lists are provided by the AG.MA. It was therefore not difficult to use the same survey method for the KONPRESS titles as for the AG.MA media; the institutes are partially identical, and the sampling procedure fully so.

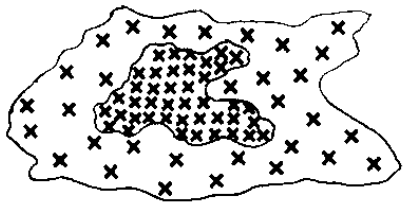
THE FUSION MODEL

The last topic we have to discuss here from the AG.MA system is a consequence of the ideas about the partnership model: how we do fusions and which are the basic ideas for this procedure (which we have been using in the last development phase since 1976).

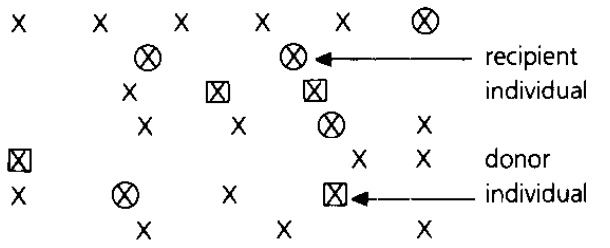
To begin with a general remark, this fusion field gives us a third reason to adopt topological conceptions. In this case we need the invariance property in order to perform transformations.

The fusion procedure is already a type of simulation. We may reduce the essential to a tranformation process, where the information coming from one survey — the donor — has been brought into another survey — the recipient, so that it is if the interviewee of the recipient-survey had answered also to the questions of the donor-survey. (Both of course, have to be samples of the identical universe). The challenge, or the postulate against the result, is that the relationship of data in the donor survey *before* the fusion and the same situation after the fusion in the recipient databank should not differ more than two independent samples differ one from another.

To understand this we must refer to the n-dimensional topological space. In the first step we divide up the population into homogeneous subgroups — with regard to the common variables — let us say in a r-dimensional space, and this by means of a typological analysis. Furthermore we define a so-called 'aureole' or



'halo' about every type, as we follow the concept of "fuzzy sets". The types exist already in the whole population, and in the donor as well as in the recipient databank.
In the environment of every individual — or point or vector — in the universe we find at the same time individuals coming from the donor and from the recipient.



Normally, the recipient-survey is larger than the donor-survey. But what is right for the r-dimensional space of the common variables must be right also for the space of the n – r specific variables and, of course, for the space of the n variables in total — the common *and* the specific ones.

When we now go over from a type of the donor to the corresponding type in the recipient survey, a particular individual of the one survey may be found in the environment of another individual of the other survey. They will be similar — only measured according to their distance in the n-dimensional space. They are close together. That means, however, that the elements of the vectors cannot be completely identical, but only compatible in certain aspects.

As to this phenomenon there exists a wonderful analogous example in the field of molecular biology. Atoms and molecules do not join together in more solid and complex liaisons *only* when they match completely in all their attributes: this happens rather when there is a certain degree of similarity. Nature is much more modest than some of our colleagues, who are always demanding perfection.

In a sufficiently large environment the members of a statistical group of sufficient size are representative of all members of this group in the universe. Thus we do the following. We marry two individuals — the one from the donor and the other from the recipient — by giving the elements of the particular vector with the specific variables of the donor in total to the elements of the particular vector with the common variables of that individual who is the nearest one in the environment of the donor individual.

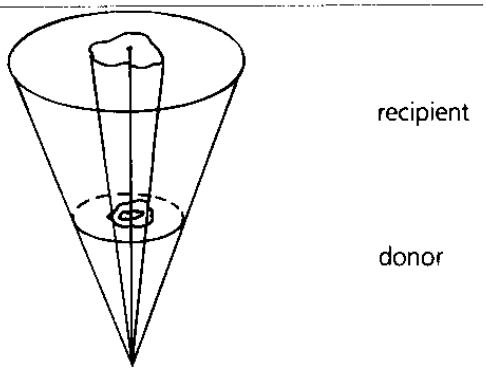
The situation before: optimal similarity in $c_1 \text{ ——— } c_r$

donor-individual: $v_d = \{ c_1 \text{ ——— } c_r, s_{r+1} \text{ ——— } s_n \}$
recipient-individual: $v_r = \{ c_1 \text{ ——— } c_r \}$

The situation afterwards: identity in $s_{r+1} \text{ ——— } s_n$

donor-individual: $v_d = \{ c_1 \text{ ——— } c_r, s_{r+1} \text{ ——— } s_n \}$
recipient-individual: $v_r = \{ c_1 \text{ ——— } c_r, s_{r+1} \text{ ——— } s_n \}$

The simulated respondents out of the receiver-survey then bear the specific characteristics of the remaining n – r dimensions, too.



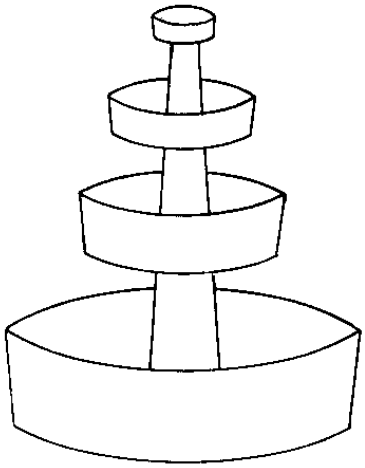
As it is impossible to find exactly for every donor-individual the nearest recipient independently from each other, we have to use an optimization process within every type. We take for this a Linear Programming Model modified for our purposes. And we apply this twice for every type: first, in comparing the type members of the donor with the type members of the recipient and, second, the aureole members of the donor with the type members of the recipient to see whether they could improve the result. The system works well, and we do a lot of controls by means of the α^2 measure to ensure the goodness of fit. Furthermore, we have many experiments to discover the theoretical background of fusion procedures of this kind (partly together with our Dutch friends) which consists of a combination of topological and stochastic elements.

Our thought in using the fusion method in the partnership model is not to bring together as many surveys as possible into one databank once for all. Our idea is to do all surveys in such a manner that they may serve as donors. We envisage the construction of a basic databank with more or less socio-demographic characteristics, and thus to have at our disposal the

8.3 The AG.MA model

potential of a databank from which we may carry out combinations of surveys, combinations of partial vectors, to meet the particular needs of media planning and other purposes.

This whole system of the databank has two general aspects: a static and a dynamic one. We call the static element a "Roman Fountain":

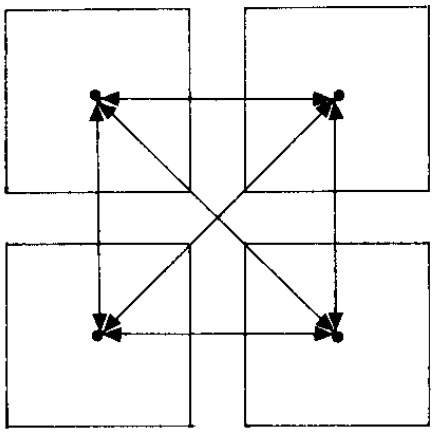


The basin at the bottom contains the socio-demograpic characteristics representing the basic structure of the whole population; the characteristics derive from the official statistics as well as from surveys (for instance the accumulation of several successive annual studies).

The 'upper bowls' represent the different surveys brought into the partnership system — in every case together with the main population in the basic basin. These different surveys are the donors of particular information.

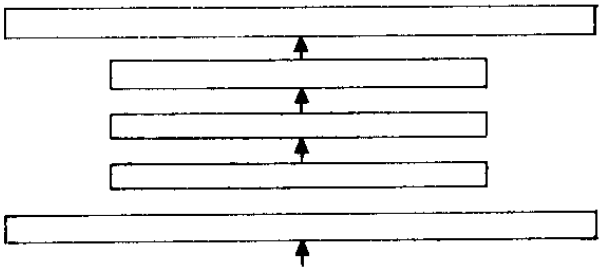
Thus the common denominator is the amount of socio-demographic variables of the interviewees. When a user wants to solve a particular problem, it is possible to combine the prepared elements of the databank according to the needs of this analysis, so that every 'bowl' may serve as donor as well as recipient.

Another way to demonstrate this situation is the following example of four elements of our 'box of bricks':



The dynamic model part can be described in this way. In the first year we have a large basic study, which might be a single source survey or a combined system as mentioned before in the static part of the partnership model.

In the two, three or even four following years it will not be necessary to make the same enormous efforts, but we can look for either supplementary studies and/or smaller pilot surveys give us hints about developments in the meantime. By means of these 'cornerstones' it will be possible to maintain the large system up to the point where it is necessary to renew the whole model:



One last remark: as fusion is a sort of practical topology and as it takes into account many traits of sampling theory, it might be possible that the whole of sampling theory could be interpreted also as a branch of topology. It would be very interesting to reflect upon this. But that is not the task of this contribution.