DETLEF LAUGWITZ

MOTIVATIONS AND LINEAR ALGEBRA*

1. MOTIVATIONS IN MATHEMATICAL INSTRUCTION

I choose the field of Linear Algebra as an example for some very general considerations on the role of motivations in mathematical instruction, especially in the teaching of those parts of mathematics which are (believed to be) applicable. I shall always have in my mind the pupil or student for whom mathematics is a requirement and who does not take it for, say, 'cultural' reasons; referring to college students S. K. Stein recently spoke of 'the captured student' [1], and similar remarks may be made for the same reasons with respect to most pupils on the secondary level.

For them, we must permanently use motivations for the mathematics we want them to learn, and these motivating examples should be taken from their environment, and they should be accessible for their age and their knowledges. These remarks are almost trivial. Nevertheless, I have to stress them. Almost all of the dozens of textbooks in Linear Algebra for the secondary level as well as for colleges and the lower university level I know, and even those written by engineers for engineering students, by economists for economics students, make the same mistake: They give a more or less thorough and complete presentation of Linear Algebra or what the authors think are the essential parts of this field (either matrices including rows and columns, or 'genuine' coordinate-free vector space theory), and at the end there follow some 'applications'. If motivations are given, or at least examples for the use of Linear Algebra, they are mostly from geometry [2].

But motivating examples from geometry, like the well known calculation of the distance of two straight lines in 3-dimensional space or of the volume of a tetrahedron, or of the axes of an ellipsoid, are either dull and tedious, or they look far-fetched to everybody who is not interested in mathematics in itself and who is not yet aware of the usefulness of such techniques for applications. I have observed the change of reactions by engineering and economics students to geometry for more than 15 years: Their interest in geometry has been permanently decreasing, and this process is going on [3]. On the other hand, Linear Algebra has to be taught now at an earlier age than before. In Germany (the Federal Republic) it is now a part of the mathematics schedule at the higher secondary level, and some preliminaries are postulated for the 14 or 15 years of age. What can be done to arouse interest in Linear Algebra?

Educational Studies in Mathematics **5** (1974) 243–254.

Of course, there is a very strong intrinsic mathematical motivation for linearity as a whole: Linear functions are simple, calculus relies on linear approximations of functions, linear problems occur everywhere in differential and integral equations, and even the second approximations to 'general' functions, by quadratic functions, lead to bilinear forms. There are most important motivations from mechanics and physics, think of the always present principle of superposition. But all this is far beyond the reach of the school level, and even beyond the first university year. (Until recently I postponed linear algebra to the second university year, when the students had enough background from physics, mechanics, engineering, and differential equations, and even then I restricted it to matrices, except with physics students who had already had a glimpse into quantum mechanics).

So we are in a situation we often have in mathematics teaching: The motivation for the teacher who knows the final aim of the instruction is different from the motivation which is accessible for the learner. The essential motivation for learning Linear Algebra is a *global* one: The field is a necessary prerequisite for other mathematical fields and for important applications; and the motivation for genuine (i.e. coordinate-free) Linear Algebra is mostly an *inner-mathematical* one, its use in almost all applications (save quantum mechanics) provides 'natural' coordinates. The motivations for the learner who is not primarily interested in mathematics will have to be *local*, that means, every new concept, every new method has to be motivated; and they should be, if possible, *extra-mathematical*, there should be direct applications in the fields which are mentally accessible for the learner. Moreover, these motivations have to be *convincing*; to take an example from another mathematical field: Maybe for the extension of the field of rationals to the real field inner-mathematical motivations will be sufficient at the age we use to teach this extension; but the proof that there is no rational number with square 2 is not convincing. This and similar examples lead to the algebraic numbers. Endless decimal numbers are probably a better motivation, and nobody will be astonished at the fact that not every decimal number is periodic. (It should rather be a surprise that the rational numbers are characterized by the periodicity property.) Of course, for the 'genuine' mathematician the irrationality of $\sqrt{2}$ is most exciting, with its Greek history and all that. But I have in my mind the consumer of mathematics who has to be won for learning just the mathematics of which the teacher knows that it will be useful for him; if we tell him about how exciting mathematics can be and how important it is for culture etc., he will in many cases prefer music or football or chess. Of course, it is legitimate that the teacher make use of such preferences of his pupils to win them for the mathematics he has to teach them.

The task of the teacher is more difficult in mathematics than in any other field: He has to give local motivations, where he himself prefers a global one. This corresponds to the fact that at the earlier stages of learning local organizing of mathematics, as Freudenthal calls it, will be the right way of mathematical instruction. And the teacher must know the world of the learners well enough to find suitable and convincing motivations which only in the minority of cases will be innermathematical, and which will only very rarely coincide with the motivations that are used in contemporary mathematical research.

2. Examples for motivations of linear algebra

I shall now discuss some examples.

A. *Mechanics*

The age-honoured motivations for vector addition (parallelogram of forces), of the inner product (work), and of the vector product (moment) should of course be mentioned. They are still useful, if only for euclidean three-dimensional space.

B. *Communication Systems*

Let there be a party of n persons $P_1, \ldots, P_n$ who speak and understand one or more of the m languages $L_1, \ldots, L_m$. We arrange the informations in a matrix A, such that $a_{rs}=1$ if P_r knows L_s, and $a_{rs}=0$ otherwise; as an example take

$$\begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{array} \begin{array}{c} \begin{array}{cccc} \text{Engl.} & \text{French} & \text{Germ.} & \text{Ital.} \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} = A.$$

You get motivations for *transposition* of matrices, and for *multiplication*:

$C = A \cdot A^t$ is a 'communication' matrix which gives the number of languages in which two persons can talk to each other; the diagonal elements indicate the numbers of languages a person knows. We have

$$C = \begin{pmatrix} 3 & 1 & 0 & 2 & 1 \\ 1 & 2 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 2 & 2 & 1 & 3 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

This matrix is *symmetric* which is to be expected from its interpretation. (Is the product $A \cdot A^t$ always symmetric?) If you only want to know which (different) persons can make direct contact to each other, you will take the reduced communication matrix R, with $r_{ii}=0$, $r_{ik}=0$ for $c_{ik}=0$, and $r_{ik}=1$ else. In the example,

$$R=\begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then, $S=R^2$ has entry s_{ik} if $P_i \neq P_k$ can speak to each other in just s_{ik} ways by use of at most one interpreter, and the main diagonal entry s_{ii} gives the number of other persons to which P_i can speak:

$$S=R^2=\begin{pmatrix} 3 & 1 & 2 & 1 & 0 \\ 1 & 2 & 1 & 2 & 1 \\ 2 & 1 & 2 & 1 & 0 \\ 1 & 2 & 1 & 3 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

After reducing S and squaring again, you will see that any two persons of our party can speak to each other by the aid of at most two interpreters.

(Find the necessary and sufficient conditions for A, that there exists a number d, such that every pair of persons can speak to each other by the aid of at most d interpreters! What is the maximum of d, if d exists?)

There are models of communication systems whose matrices are not symmetric. Take n stations (communication centers), and let C be the matrix with entries $c_{ik}=1$ if station i can speak to station k, and $c_{ik}=0$ otherwise. By definition, let $c_{ii}=0$. Take as an example

$$C=\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$

Here, stations 2 and 4 can speak in both directions, whereas 2 and 3 cannot have any direct contact. But what about using a relay station? Taking

$$C^2=\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \end{pmatrix}.$$

you see, that 4 can speak to 1 in two different ways by using exactly one relay station. The *sum*

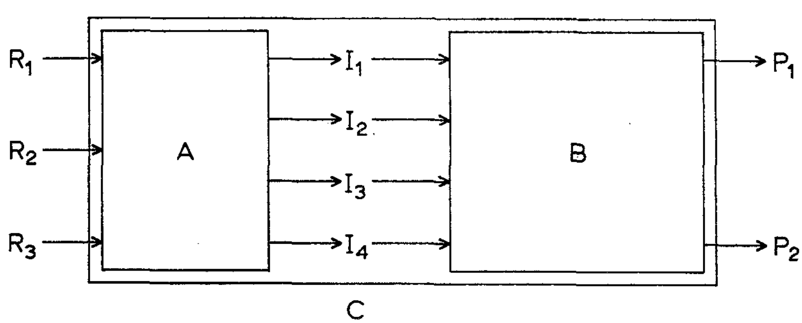
$$C+C^2=\begin{pmatrix}0&1&1&1\\1&1&1&2\\1&0&0&1\\2&1&1&1\end{pmatrix}.$$

shows that only 3 cannot speak to 2, and that each other station can speak to every other station either directly or by using at most one relay station. The sum $C+C^2+C^3$ has only positive entries; at most 2 relay stations are needed. (Why is this the worst result to be expected, if any?)

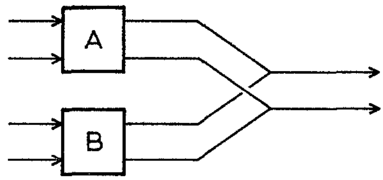
Here we get a motivation for the *addition* of matrices.

C. *Black Boxes*

The concept of 'black box' gives a very useful introduction to mathematical modeling by the use of matrices. Its origin seems to be in electrical engineering, but there are useful models which require no physics. Let C be a factory which uses three types of raw material, P_1, P_2, P_3 for the production of two types of final products, P_1, P_2. The factory C consists of two departments, A and B. Department A transforms the raw materials into four types of intermediate products, $I_1, \ldots, I_4$, which are needed in department B for the production of P_1, P_2. Assume that a_{ij} units of R_i are needed for the production of one unit of I_j, and b_{jk} units of I_j for the production of one unit of P_k. Then $c_{ik}=\sum_j a_{ij}b_{jk}$ units of R_i are needed for the production of 1 unit of the final product P_k. This model can be varied in many ways, and we get a very intuitive motivation for matrix *multiplication*, $C=A\cdot B$:



If there is some knowledge of electricity, the usual models for n-ports are useful, with currents and voltages. Matrix *addition* is easily motivated by using 'parallel' black boxes,



Moreover, in our example above we see that the column **R** can be calculated uniquely, if the production column **P** is known, but in general not vice versa. There are hidden hints to inverse matrices, linear systems, and linear optimization.

Models from economy give examples for *inner products*: If p_k is the price for one unit of product number k, and you want to buy a_k units of this product, then $\mathbf{a}^t\mathbf{p} = \sum_k a_k p_k$ is the amount you have to pay when you leave the shop.

D. *Coding and Decoding*

This field provides very nice examples which will appeal to many pupils at a certain age. Unfortunately, it does not directly motivate why we multiply matrices just in the way we do. But if matrix multiplication is already known we can make it more interesting by a little cryptography.

As usual, we start by numbering the alphabet, using 0 for all signs which are not letters, as free spaces:

A	*B*	*C*	*D*	*E*	*F*	*G*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	...
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	...

Our message reads LINEAR ALGEBRA and becomes

10, 8, 12, 5, 1, 16, 0, 1, 10, 7, 5, 2, 16, 1;

as a 'coding matrix' I take an invertible matrix, C,

$$C = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix}.$$

We write the coded message as a matrix with three rows, M, and take $C \cdot M$:

$$C \cdot M = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 10 & 5 & 0 & 7 & 16 \\ 8 & 1 & 1 & 5 & 1 \\ 12 & 16 & 10 & 2 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 14 & -9 & -8 & 15 & 18 \\ 30 & 22 & 11 & 14 & 17 \\ 48 & 28 & 12 & 26 & 34 \end{pmatrix} = M'.$$

The columns of M' are transmitted as a coded message. The receiver knows the decoding matrix $D=C^{-1}$, and he has to calculate

$$M = D\cdot M' = C^{-1}(C\cdot M).$$

In our case,

$$D = C^{-1} = \begin{pmatrix} -1 & -4 & 3 \\ 1 & 3 & -2 \\ 0 & 2 & -1 \end{pmatrix}.$$

Several little problems arise. For instance, it is certainly useful that both the coding and the decoding matrix have integers as their entries; find conditions for this! (Since determinants will not be available at this stage, take 2–2-matrices only.) Life might be easier if $D=C$; is this possible with non-trivial matrices? (It is a nice problem to determine all 2–2-matrices with this property.) The occurence of negative numbers is a certain handicap; can it be avoided? Moreover, are there matrices which together with their inverses have no negative entries, and are these matrices useful? (I shall return to the last question.)

E. *Population Matrices*

Though eigenvalues and eigenvectors are most important in geometry and higher mechanics and physics there can be found only few motivations at school level.

Consider a population of (hypothetical) beetles on a tropical island. The following facts are known: A beetle will live at most two months. Let f_j be the number of new beetles created per old beetle of age j (months). It is observed that $f_0=0$, $f_1=1$, $f_2=3$ under laboratory conditions similar to those in freedom. And it is known that the population on the island is stable. Let p_j be the probability that a beetle of age j at the month k will survive to month $k+1$, and $p_0=\frac{1}{2}$ is known from the laboratory observations. Due to unknown influences on the island which cannot be simulated in the laboratory the probability p_1 is not known. Can it be calculated from the data?

Let a_{jk} be the number of living beetles of age j months at time k (months), and consider the column

$$\mathbf{a}_k = \begin{pmatrix} a_{0k} \\ a_{1k} \\ \vdots \\ a_{nk} \end{pmatrix}$$

which gives the age distribution of the population at time k, with n as the maximum possible age. Then we have

$$\mathbf{a}_{k+1} = T\mathbf{a}_k$$

with the transition matrix

$$T=\begin{pmatrix} f_0 & f_1 & f_2 \cdots f_n \\ p_0 & 0 & 0 \ldots 0 \\ 0 & p_1 & 0 \cdots 0 \\ 0 \cdots 0 & p_{n-1} & 0 \end{pmatrix}.$$

The condition that the population is stable is $\mathbf{a}_{k+1}=\mathbf{a}_k$ or

$$T\mathbf{a}_k = \mathbf{a}_k,$$

hence, 1 has to be an eigenvalue of the matrix T. In our special case, we have

$$T=\begin{pmatrix} 0 & 1 & 3 \\ \frac{1}{2} & 0 & 0 \\ 0 & p_1 & 0 \end{pmatrix}.$$

If determinants are known, we get the characteristic equation $0=-\lambda^3+ +\frac{1}{2}\lambda+\frac{3}{2}p_1$, which has root 1 iff $p_1=\frac{1}{3}$.

Determinants can be avoided by direct computation of the non-trivial population distribution, from the linear equations with $x=a_{0k}=a_{0,k+1}$, $y=a_{1k}=a_{1,k+1}$, and $z=a_{2k}=a_{2,k+1}$:

$$\begin{aligned} y+3z&=x \\ \tfrac{1}{2}x&=y \\ p_1 y&=z, \end{aligned}$$

from which $p_1=\frac{1}{3}$ is easily calculated under the assumption $z\neq 0$, and moreover the 'eigenvector' $x=6z$, $y=3z$, $z=z$ is obtained.

One can ask several questions, for instance: Are the f_j's uniquely determined if you know the age distribution and the p_j's? (In fact, they are not: $f_0=f_1=0$, $f_2=6$ gives another stable population in our example).

F. *Stochastic Matrices*

Supposed there is some previous knowledge of elementary probability, then call a row vector a probability vector if its entries are non-negative and have sum 1. A stochastic matrix is a matrix all of whose rows are probability vectors.

As an example take the following method for voting predictions. Let there be three parties, S(ocialists), C(onservatives), L(iberals). At an election the following shift of voters is observed

to:		S	C	L
	S	60%	20%	20%
from:	C	30%	60%	10%
	L	30%	20%	50%

with the transition matrix

$$P = \begin{pmatrix} 0.60 & 0.20 & 0.20 \\ 0.30 & 0.60 & 0.10 \\ 0.30 & 0.20 & 0.50 \end{pmatrix}.$$

which is stochastic. If the trend of the voters is assumed to remain unchanged until the next election, the transition matrix P^2 permits a prediction.

Problems: It is to be expected from the interpretation that the products of stochastic matrices are again stochastic.

Or: Are there stochastic matrices whose inverse matrices have the same property? Find them! (I shall return to this question).

In our case,

$$P^2 = \begin{pmatrix} 0.48 & 0.28 & 0.24 \\ 0.39 & 0.44 & 0.17 \\ 0.39 & 0.28 & 0.33 \end{pmatrix}.$$

That means: 48% of the people who voted for S in the first election will do the same in the third election, etc. If you want to calculate the distribution of votes, you have to take the transposed matrix P^t; e.g., if each party got 100 votes in the first election, the distribution will be

$$P^t \begin{pmatrix} 100 \\ 100 \\ 100 \end{pmatrix} = \begin{pmatrix} 120 \\ 100 \\ 80 \end{pmatrix}$$

after the first election, and

$$(P^t)^2 \begin{pmatrix} 100 \\ 100 \\ 100 \end{pmatrix} = P^t \begin{pmatrix} 120 \\ 100 \\ 100 \end{pmatrix} = \begin{pmatrix} 126 \\ 100 \\ 74 \end{pmatrix}$$

after the second election,

$$(P^t)^3 \begin{pmatrix} 100 \\ 100 \\ 100 \end{pmatrix} = \begin{pmatrix} 127.8 \\ 100.0 \\ 72.2 \end{pmatrix}$$

after the third election, etc.

(At a higher stage of instruction, many interesting questions arise. For instance, it is 'easily' seen that 1 is always an eigenvalue of P, with (1, 1, 1) as an eigenvector. From the determinant condition or otherwise one sees that 1 is also an eigenvalue of P^t. Hence, for every P there should exist stable distributions of voters!)

In our example, the eigenvectors are proportional to (9; 7; 5), for instance after multiplying with $\frac{100}{7}$: (128.6; 100; 71.4). This is approximated by the results of consecutive elections! Is this a general fact?)

There are more realistic examples: The stochastic matrix of survival probabilities from one state to another state, or of probabilities of a disease progressing from one state to another at different periods of diagnosis[4].

G. *Linear Optimization*

This field is well known, but it has not yet become a subject which is generally accepted as useful mathematics at school level. Concepts and methods of rows, columns, and matrices are used, and even such concepts as basis and transformations occur, if only at a comparatively higher level.

What makes the subject very attractive is that the concepts of row-and-column linear algebra are used in permanent connection with other mathematical subjects, as inequalities and convexity. I need not go into details, though I feel contempted to discuss the subject because it appears to be a most impressive example for the usefulness of mathematics instruction nowadays.

3. Linear Algebra without Geometry?

Just up to the discussion of optimization I have always tried in this paper to avoid connections of linear algebra and geometry. As I said at the start, I am aware of the fact that – though linear algebra once emerged from analytic geometry, and analytic geometry had its roots in 'genuine' geometry – geometry appears to be driven into the backwoods by contemporary mathematics instruction. Though much of the material I have discussed are examples of row-and-column linear algebra for organizing and reorganizing numerical data, there were some hints where I see geometry creeping in.

Let me discuss a few examples for the use of geometric intuition and methods of thinking in those subjects which appeared here without any connection to geometry.

We saw that it might be useful to have square matrices A with the following properties: (a) A and A^{-1} have non-negative entries, (b) the entries are non-negative integers. [The question (a) was recently proposed as Elementary Problem E 2379, *Amer. Math. Monthly* **79**, (1972) 1033 by H. Kestelman.] Try (a) by purely algebraic methods; you will find it extremely tedious and probably get nowhere.

Now consider the problem in R^n with the non-negative 'cone' C consisting of the columns all of whose entries are non-negative. Then obviously $AC \subseteq C$ and $A^{-1}C \subseteq C$, hence A is a 1–1-mapping (or bijection) of C. Any vector from C which is a sum of two linearly independent vectors from C will have an image with the same property; hence, A permutes the positive half-axes! A has exactly one positive element in each row as well as in each column, all other entries being 0. This condition is sufficient. If the entries of A and A^{-1} are integers, according to (b), the only solutions are permutation matrices. As coding matrices they are not very good.

By the way, we have also solved the problem of finding all stochastic matrices whose inverses have the same property: Only permutation matrices are solutions.

Consider the transposed P^t of a stochastic matrix P. Let Δ be the intersection of the hyperplane $x_1 + \ldots + x_n = 1$ of R^n with the non-negative cone C. P^t is a contraction mapping of Δ into itself and by geometric reasoning the fixed point theorem may be made plausible. The iterating process of 2F above gives an example for the use of this fixed point theorem for contraction mappings; this theorem is useful also in analysis (Newton's approximation method of zeros).

This solution of a purely algebraic problem by approximation is unsatisfactory to a mathematical purist. But it is a nice example for good instruction. Now the remark will arise: We saw easily that P has an eigenvector $(1, 1, \ldots, 1)$ and eigenvalue 1; we concluded without explicit use of this fact that also P^t has 1 as an eigenvalue. Of course, this could have been established by the use of determinants. The question is: Can we establish by *geometric* reasoning, that an eigenvalue λ of a square matrix A is always an eigenvalue of A^t? Or, letting $B = A - \lambda E$, does $Bx = 0$ with an $x \neq 0$ imply, that a vector $y \neq 0$ exists such that $B^t y = 0$? $Bx = 0$ means that the columns of B are linearly dependent, these vectors lie in a hyperplane. Let y be a vector perpendicular to this hyperplane, and you get $B^t y = 0$.

In linear optimization geometric thinking is especially useful. The graphic solution methods for problems in two variables contain a lot of good mathematics which can be instructed at a rather early stage: Linearity, convexity, intersection of point sets. Thus, the algebraic methods which are needed for large numbers of variables and restrictions can be made transparent by a geometric anticipation of the essential steps.

Such examples give glimpses into a very broad and interesting field of modern mathematics for applications: Functional analysis, which combines the methods of linear algebra and geometry (like convexity, mappings, fixed points, etc.). I think here is an answer to those who postulate that mathematics instruction should be 'modern' and who wish to see the gap filled between school mathematics and university. This cannot be done in an anti-didactic inversion (Freudenthal) by introducing so called fundamental notions like the language of sets or of the algebra of vector space at an early stage of instruction, where no motivation for these can be presented.

Modernization is really effective if it is oriented at the *contents* of useful modern mathematics – in our example functional analysis – which may give the global motivation for the teacher. What I wanted to show was how we can give local motivations for the learner, even saving fields like geometry (which some people say to have died) from complete recess by showing how their methods are useful for present day problems.

Note added in proof: When giving this lecture I was not aware of the excellent book by T. J. Fletcher: *Linear Algebra Through Its Applications*, Van Nostrand Reinhold, London, etc. 1972. There are several aspects in my lecture which have been treated in Fletcher's book.

Darmstadt

NOTES

* Invited lecture held at the 3e Seminaire d'Echternach, I.C.M.I., 4–9, June 1973, on 'New Aspects of Mathematical Applications at School Level'.

[1] S. K. Stein, 'Mathematics for the Captured Student', *Amer. Math. Monthly* **79**, (1972), 1023–1032.

[2] A book which tries to avoid these mistakes and which has been helpful for the preparation of this paper is: Hugh G. Campbell, *Linear Algebra with Applications*, New York 1971. This book contains many useful references to applications of matrices.

[3] The reasons for the permanent recess of geometry in mathematical instruction (not in mathematical research, by the way!) are not quite clear. 'New Math' has evidently not much affinity to genuine geometry, and there is a temporary misinterpretation of 'rigour' in Mathematics which permits only trivialities to be rigorous mathematics. See: H. Freudenthal, *Mathematics as an Educational Task*, Dordrecht 1973, pp. 38ff., pp. 402ff.

[4] At the 3rd Seminar at Echternach John H. Durran gave a conference on Markov Chains in which he used the three possibilities of British weather (fine, dull, wet) for the motivation of transition probabilities (changes of weather from one day to the next) and of stochastic matrices.