RAIRO. Recherche opérationnelle

F. A. Lootsma

The french and the american school in multi-criteria decision analysis

RAIRO. Recherche opérationnelle, tome 24, nº 3 (1990), p. 263-285

<http://www.numdam.org/item?id=RO_1990__24_3_263_0>

Recherche opérationnelle/Operations Research
(vol. 24, n° 3, 1990, p. 263 à 285)

THE FRENCH AND THE AMERICAN SCHOOL IN MULTI-CRITERIA DECISION ANALYSIS (*)

by F. A. LOOTSMA (1)

Abstract. — *We address a crucial problem in multi-criteria analysis: the transition from the objective evaluation of the decision alternatives to the subjective weighing. The French school represented by the ÉLECTRE systems of Roy, and the American school represented by Saaty's Analytic Hierarchy Process (AHP), incorporate subjective human judgement in different ways. In the present paper we introduce a thoroughly revised AHP to demonstrate that the American school did not go far enough, so that its potential is somewhat overestimated. Using the persistent pattern of human comparative judgement in many unrelated areas such as history, planning, and psychophysics, we show that there is a natural scale to quantify verbal preferential statements of increasing intensity. Moreover, we analyze the assumptions underlying the logarithmic-regression procedure to compute the impact scores and the criterion weights. Violations of these assumptions imply that we do not merely supply decision support in order to identify the pre-existing consensus in a group of decision makers. We reform the decision process by accelerating the deliberations in the direction of a compromise solution. Sensitivity analysis based on a variety of geometric scales shows that the results of the French and the American school of thinking are unexpectedly close to each other.*

Keywords : Outranking relations; multi-attribute value functions; category scales; geometric progression; psychophysical power law; logarithmic regression; aggregation; preference structure; rank reversal.

Résumé. — *On a affaire à un problème crucial dans l'analyse multicritère : la transition de l'évaluation objective des alternatives à la pondération subjective. L'école française représentée par les systèmes ELECTRE de Roy, et l'école américaine représentée par l'Analytic Hierarchy Process (AHP) de Saaty, traitent les jugements humains subjectifs d'une façon différente. Dans cet article-ci, nous introduisons un AHP changé profondément, pour démontrer que l'école américaine a arrêté la recherche trop tôt, de sorte que son potentiel est surestimé. On se sert des propriétés générales du jugement humain dans quelques terrains isolés comme l'histoire, le planning, et la psychophysique, à fin de montrer qu'il y a une échelle naturelle pour quantifier les déclarations préférentielles d'intensité croissante. En plus, on analyse les hypothèses qui sont à la base de la régression logarithmique utilisée pour calculer les scores des alternatives et les poids des critères. Les violations de ces hypothèses impliquent qu'on n'offre pas une aide à la décision à fin d'identifier le consensus pré-existant dans un gorupe de décideurs. Au contraire, on réforme le processus de la décision en accélérant les délibérations vers un compromis inattendu. Pour terminer, les variations de l'échelle géométrique montrent que les résultats de l'école française et américaine sont plus proches qu'on ne le croyait.*

(*) Received in February 1990.
(1) Faculty of Mathematics and Informatics, Delft University of Technology, P.O. Box 356, 2600 AJ Delft, The Netherlands.

Recherche opérationnelle/Operations Research, 0399-0559/90/03 263 23/$ 4.30

1. INTRODUCTION

It seems to be customary in Europe to distinguish a French and an American school in the field of multi-criteria decision analysis (Schärlig, 1985; Colson and De Bruyn, 1989). The founding father of the French school is B. Roy who developed a series of ÉLECTRE methods (*see* Roy, 1968, 1985, 1989) and prompted many scientists, mainly in French-speaking regions, to design related methods such as PROMÉTHÉE (Brans, Marechal and Vincke, 1984). The American school is inspired by the work of Keeney and Raiffa (1976) on multi-attibute value functions and multi-attribute utility theory. A popular method, typically fitting into this framework, is the Analytic Hierarchy Process (AHP) of Saaty (1980, 1988). Both schools are concerned with the same problem: the evaluation of a finite number of alternatives $A_1, \ldots, A_n$ under a finite number of conflicting criteria $C_1, \ldots, C_m$, by a single decision maker or by a decision-making body. Taking ÉLECTRE and the AHP to represent (as usually) the respective schools, we can easily describe the differences and the similarities.

ÉLECTRE starts with a pairwise evaluation of the alternatives under each of the criteria separately. Using the physical or monetary values $g_i(A_j)$ and $g_i(A_k)$ of the respective alternatives A_j and A_k under a measurable criterion C_i, and introducing certain threshold levels for the difference $g_i(A_j)-g_i(A_k)$, the decision maker may declare that he is indifferent between the alternatives under consideration, that he has a weak or a strict preference for one of the two, or that he is unable to express any of these preference relations. If the alternatives are not measurable under C_i, their performance is expressed on a qualitative scale with increasing values such as 1, 2, ..., 10 assigned to the respective echelons; thereafter, threshold levels are introduced and employed in the same way to elicit the required preference information. Both, indifference between A_j and A_k, as well as a weak or a strict preference for A_j, are summarized in the statement that A_j is at least as good as A_k or, equivalently, that A_j outranks A_k. Thus, under each criterion there is a complete or incomplete system of binary relations between the alternatives, the so-called outranking relations. Next, the decision maker is requested to assign weights or importance factors to the criteria in order to express their relative importance; ÉLECTRE does not really propose a systematic approach to reduce the notorious inconsistency of human beings when they establish such weights. Finally, there is an aggregation step. For each pair of alternatives A_j and A_k ÉLECTRE calculates the so-called concordance index, roughly defined as the total amount of evidence to support the conclusion that A_j globally outranks A_k, as well as the discordance index, the total amount of counter-evidence. The concordance index includes, for instance, the total weight of the criteria where A_j outranks A_k; in the discordance index, the veto thresholds play a major role. Balancing the two indexes, ÉLECTRE finally decides whether A_j outranks A_k, whether A_k outranks A_j or whether there is no global outranking relationship between the two alternatives. Eventually, ÉLECTRE yields a global system of binary outranking relations between the alternatives. Because the system is not necessarily complete, ÉLECTRE is sometimes unable to identify the preferred alternative. It only produces then a core of leading alternatives. Moreover, ÉLECTRE cannot always rank the alternatives completely in a subjective order of preference.

The AHP also starts with a pairwise evaluation of the alternatives under each of the criteria separately. In the basic experiment, where the alternatives A_j and A_k are presented under the criterion C_i, the decision maker is requested to express his indifference between the two, or his weak, strong, or very strong preference for one of them. His verbal judgement (the selected gradation) is subsequently converted into a numerical value $r_{jk}^{(i)}$ on the so-called fundamental scale. Using the matrix $R^i = \{ r_{jk}^{(i)} \}$, the AHP calculates the partial, single-criterion scores $\tilde{v}_i(A_j)$, $j=1, \ldots, n$, also referred to as the impact scores, approximating the subjective values of the alternatives under criterion C_i. It is worth noting that the partial scores are not unique. Because the ratio $\tilde{v}_i(A_j)/\tilde{v}_i(A_k)$ is defined for each pair (A_j, A_k) of alternatives, the partial scores have a multiplicative degree of freedom. They can accordingly be normalized in such a way that

$$\sum_{j=1}^{n} \tilde{v}_i(A_j) = 1; \qquad i=1, \ldots, m. \tag{1}$$

Similar pairwise comparisons and similar calculations yield normalized weights $\tilde{w}(C_i)$, $i=1, \ldots, m$, for the respective criteria. Finally, there is an aggregation step generating the global, multi-criteria scores $\tilde{f}(A_j)$ via the arithmetic-mean rule

$$\tilde{f}(A_j) = \sum_{i=1}^{m} \tilde{w}(C_i)\, \tilde{v}_i(A_j); \qquad j=1, \ldots, n. \tag{2}$$

By these quantities, usually referred to as the final scores, we have a global order (a global preference structure) defined on the set of alternatives. In the terminology of the American school, the partial and the final scores constitute partial value functions and a global value function respectively. In general, each of the alternatives is Pareto-optimal, because alternatives dominated by others can immediately be dropped from further consideration.

At first sight, the AHP yields stronger results than ÉLECTRE. The final scores can be used to identify the preferred alternative, to sort the alternatives into a limited number of categories, to rank the alternatives in a subjective order of preference, and to allocate resources to the respective alternatives on the basis of the relative preferences for them. Sensitivity analysis, however, shows that the rank order of the final scores varies under reasonable deviations from the fundamental scale, so that sorting procedures and resource allocation must be carried out with great care (Lootsma *et al.*, 1990). Moreover, decision makers find it difficult to choose one of the verbal qualifications (indifference, weak, strong, or very strong preference) in order to express their relative preference for one of the two alternatives in a pairwise comparison. This is particularly true in those cases where the performance of the alternatives A_j and A_k under a given criterion C_i can be expressed in physical or monetary values $g_i(A_j)$ and $g_i(A_k)$. When the decision makers carry out the required experiments, they are puzzled by the relation between the physical or monetary values on the one hand and the impact scores $\tilde{v}_i(A_j)$ and $\tilde{v}_i(A_k)$ on the other. In ÉLECTRE, the treatment of measurable criteria is definitely more direct and more transparent. It seems to be easier for a decision maker to accept the AHP when the performance of the alternatives cannot be measured (when the colours of the alternatives are compared, for instance, or the design, the elegance, and the style), and when the criteria are compared on the basis of their relative importance in the actual decision problem (Barda, 1989). Here, ÉLECTRE's facilities are weaker, so that the AHP could be used to fill the gap in actual applications of ÉLECTRE.

This question leads us straightaway to the heart of the matter in the present paper. The physical or monetary values $g_i(A_j)$ and $g_i(A_k)$ are usually obtained by a more or less objective evaluation of the alternatives, that is, by scientific measurement or by cost calculations. The impact scores $\tilde{v}_i(A_j)$ and $\tilde{v}_i(A_k)$ are due to a subjective weighing of the alternatives, via human judgement expressed in verbal terms. We cannot deny that the transition from objective evaluation to subjective weighing, as well as the quantification of verbal judgement, are still poorly explained. Saaty (1980, 1988) introduced on doubtful grounds a "fundamental scale" and an "eigenvector method" to calculate the impact scores and the criterion weights; his arithmetic-mean aggregation rule (2) does not really apply, because only the ratios $\tilde{v}_i(A_j)/\tilde{v}_i(A_k)$ are properly defined [for a critical discussion, *see* also French (1988)]. In recent years, we introduced a class of geometric scales and a geometric-mean aggregation rule (Lootsma, 1987, 1988; Lootsma *et al.*, 1990), but in real-life applications we are still not satisfied with the underlying theory: we cannot properly explain it to the decision makers. This prompted us to carry out additional research on the nature of comparative human judgement. The results may be found in the sections to follow.

The organization of the present paper is as follows. Section 2 provides a heuristic introduction to illustrate the transition from car prices to the subjective judgements whereby cars are referred to as "cheap", "somewhat more expensive", "more expensive", or "much more expensive". In fact, we subdivide a given price range into a number of price categories (intervals) which are felt to be of the same order of magnitude, and we use the corresponding echelons (levels) to establish ratios of price increments expressing what we mean by "somewhat more", "more", and "much more". Section 3 shows that human judgement leads in many unrelated areas to the same categorization of intervals: there are roughly four major categories, the echelons constitute a sequence with geometric progression, and the progression factor is roughly 4. We use these results in Section 4 to propose a natural geometric scale for the quantification of verbal, comparative judgement: a scale with major as well as threshold echelons, and the progression factor 2. Moreover, we calculate the impact scores and the criterion weights via logarithmic regression, whereby we cover applications in groups of decision makers with non-negligible power relations. In Section 5, we are concerned with the geometric-mean aggregation rule and with the notorious phenomenon of rank reversal. We show that the weights of the criteria as well as the final scores of the alternatives are rather insensitive to variations of the scale parameter, the logarithm of the progression factor which characterizes the geometric scales. Nevertheless, rank reversal is an almost inevitable phenomenon in our revised version of the AHP (rank reversal due to scale sensitivity is ignored in Saaty's original version). Finally, Section 6 concludes the paper with some general remarks and comments on the subjective weighing via multi-criteria methods of the French and the American school. We shall particularly discuss the question of whether the proposed decision support is welcome in actual decision making. It is our experience that advisory councils and consulting companies may reject the support, so that the decision is ultimately left to authorities who do not always have the time to digest the flood of information presented to them.

2. CATEGORIZATION OF A PRICE RANGE

We start with the example which is frequently used to illustrate multi-criteria analysis: the evaluation and the selection of a car. Usually, low costs are important for the decision maker so that he carefully considers the consumer price, and possibly the annual expenditures for maintenance and insurance.

The consumer price as such, however, cannot tell us whether the car in question would be more or less acceptable to him. That depends on his spending power and on the alternative cars which he seriously has in mind. In general, there is a minimum price C_{min} which he is prepared to pay, and a maximum price C_{max} which he can afford and which he does not really want to exceed. Intuitively, he will subdivide the price range (C_{min}, C_{max}) into a number of price categories by the introduction of subjectively distinct price levels partitioning the range into subintervals which are felt to be of the same order of magnitude. We take $e_0, e_1, e_2, \ldots$ to stand for the so-called echelons of the category scale under construction, and $C_{min}+e_0$, $C_{min}+e_1, \ldots$ as the associated price levels. In order to model the requirement that the sub-intervals must subjectively be equal, we recall Weber's law (1834) in psychophysics, stating that the just noticeable difference Δs of stimulus intensities must be proportional to the actual stimulus level s. The just noticeable difference is a step of the smallest possible order of magnitude when we move from C_{min} to C_{max}; we assume that it is practically the step carried out in the construction of our model. Thus, taking here the price increment above C_{min} as the stimulus intensity, that is, assuming that the decision maker is not really sensitive to the price as such but to the excess above the minimum price C_{min} which he has to pay anyway, we set

$$e_\delta - e_{\delta-1} = \varepsilon\, e_{\delta-1}, \qquad \delta = 1, 2, \ldots,$$

which yields

$$e_\delta = (1+\varepsilon)\, e_{\delta-1} = \ldots = (1+\varepsilon)^\delta e_0.$$

Obviously, the echelons constitute a sequence with geometric progression. The initial step is e_0, and $(1+\varepsilon)$ is the progression factor. It is important to observe that the number of categories is rather small, because our linguistic capacity to describe the categories unambiguously in verbal terms is limited. We can introduce, for instance, the following qualifications to identify the subsequent price categories:

cheap,

cheap/somewhat more expensive,

somewhat more expensive,

somewhat more/more expensive,

more expensive,

more/much more expensive,

much more expensive.

Thus, we have four major, linguistically distinct categories: cheap, somewhat more, and much more expensive cars. Moreover, there are three so-called threshold categories between them, which can be used if the decision maker hesitates between the neighbouring qualifications.

The next section will show that human beings follow the same pattern in many unrelated areas when they categorize an interval. They introduce three to five major categories, and the progression factor $(1+\varepsilon)^2$ is roughly 4. By the interpolation of threshold categories, they have a more refined subdivision of the given interval. Then there are six to nine categories, and the progression factor $(1+\varepsilon)$ is roughly 2. In the present section, we will use these results in advance, in order to complete the categorization of a price range. Let us, for instance, take the range between Dfl 20,000 (écu 9,000) for a modest Renault 5 and Dfl 40,000 (écu 18,000) for a well-equipped Renault 21 in the Netherlands. The length of the range is Dfl 20,000. Hence, setting the last price level $C_{\min}+e_6$ roughly at $C_{\max}$ we have

$$e_6 = C_{\max} - C_{\min},$$

$$(1+\varepsilon)^6 e_0 = 20{,}000; \qquad 1+\varepsilon = 2,$$

$$e_0 = 20{,}000/64 \approx 300.$$

It is sometimes more convenient to associate the above-named qualifications, not with the sub-intervals, but with the price levels. Thus, cheap cars are roughly found at the price $C_{\min}+e_0$, somewhat more expensive cars at $C_{\min}+e_2$, etc. This will eventually lead to the following subdivision:

$C_{\min}+e_0$, Dfl 20,300: cheap cars.

$C_{\min}+e_1$, Dfl 20,600: cheap/somewhat more expensive cars.

$C_{\min}+e_2$, Dfl 21,200: somewhat more expensive cars.

$C_{\min}+e_3$, Dfl 22,500: somewhat more/more expensive cars.

$C_{\min}+e_4$, Dfl 25,000: more expensive cars.

$C_{\min}+e_5$, Dfl 30,000: more/much more expensive cars.

$C_{\min}+e_6$, Dfl 40,000: much more expensive cars.

We leave it to the reader to decide whether he would in principle agree with the price levels assigned to the respective qualifications. Because we have been concerned with price increments above the lower bound $C_{\min}$, we can now give a more precise interpretation for the qualifications. A somewhat more expensive car has a price increment e_2, which is 4 times the price increment e_0 of a cheap car, etc. We will use this observation to identify the so-called modifiers "somewhat more", "more", and "much more" with ratios 4:1, 16:1, and 64:1 respectively. They refer to increments above a certain minimum level. Via the progression factor and the number of qualifications, these ratios are also related to a certain maximum level. Note that, by this convention, a car of Dfl 25,000 is somewhat more expensive than a car of Dfl 21,2oo because the price increments also have the ratio 4:1. By the same token, a car of Dfl 21,200 is somewhat cheaper than a car of Dfl 25,000. We ignore the possibility of hysteresis when we invert the orientation of comparative judgement.

It will hopefully be clear now, why we have taken ratios of price *increments* to model the intensity of modifiers such as "somewhat more", etc. The ratios of the prices themselves are practically 1:1, at least at the lower end of the range under consideration. They do not properly model the strength of the corresponding feelings.

Because the alternative cars are judged under the consumer-price criterion, the target is at the lower end C_{min} of the interval of possible prices. From this point the decision maker looks at less favourable alternatives. That is the reason why the above categorization, in principle an asymmetric subdivision of the interval under consideration, has been oriented from the lower end: the upward direction is typically the line of sight of the decision maker, at least under the given criterion.

When the cars are judged under the reliability criterion, the orientation is downwards. Numerical data to estimate the reliability are usually available. Consumer organizations collect information about many types and models of cars which follow the prescribed maintenance procedures, and they publish the frequencies of technical failures in the first three or five years. Let us suppose that the decision maker only considers cars with a reliability of at least 95%, so that we are restricted to the interval (R_{min}, R_{max}) with $R_{min}=95$ and $R_{max}=100$. Following the mode of operation just described, we obtain the major echelons

$R_{max}-e_0$, 99.9%: reliable cars.

$R_{max}-e_2$, 99.7%: somewhat less reliable cars.

$R_{max}-e_4$, 98.7%: less reliable cars.

$R_{max}-e_6$, 95.0%: much less reliable cars,

because $e_0=(100-95)/64 \approx 0.08$, with the progression factor such that $e_6/e_0=64$.

It is important to note that the alternatives are usually compared with respect to a certain target. The relative performance is inversely proportional to the distance from the target. The reader can easily verify this in the two examples just given. If we take R_j and R_k to denote the reliability of the alternative cars A_j and A_k, for instance, then the inverse ratio

$$e_k/e_j = (R_{\max} - R_k)/(R_{\max} - R_j)$$

represents the relative performance of A_j and A_k under the reliability criterion. The qualifications "somewhat cheaper" and "somewhat more reliable" imply that the *inverse ratio* of the distances to the target (the echelons) is 4 : 1.

3. CATEGORY SCALES IN OTHER AREAS

It is surprising to see how consistently human beings categorize certain intervals of interest in totally unrelated areas. They use echelons with geometric progression because the subsequent intervals are felt to be of the same order of magnitude. Both, the progression factor and the number of categories or echelons, are so uniform that we confidently use them in multi-criteria analysis to establish a natural relationship between verbal comparative judgement on the one hand and a particular geometric scale on the other. In this section we present some examples to show, for instance, how human subjects partition certain ranges on the time axis, and how they categorize sound and light intensities.

a. Historical periods

The written history of Europe, from 3,000 BC until today, is subdivided into a small number of major periods. Omitting the recent years which may be the beginning of a new period, and looking backwards from 1985, we distinguish the following turning points marking off the start of a characteristic development:

1947, 38 years before 1985: cold war, decolonization.
1815, 170 years before 1985: industrial and colonial dominance.
1500, 500 years before 1985: world-wide trade, modern science.
450, 1,550 years before 1985: middle ages.
−3000, 5,000 years before 1985: ancient history.

These major echelons, measured by the number of years before 1985, constitute a geometric sequence with the progression factor 3.3. We obtain a more refined subdivision when we introduce the years

1914, 71 years before 1985: world wars.
1700, 300 years before 1985: modern science established.
1100, 900 years before 1985: high middle ages.
−800, 2,800 years before 1985: Greek/Roman history.

With these turning points interpolated between the major ones, we find a geometric sequence of echelons, with progression factor 1.8.

b. Planning horizons

Let us now turn towards the future, and let us concern ourselves with industrial planning activities. In this area, we usually observe a hierarchy of planning cycles where decisions under higher degrees of uncertainty and with more important consequences for the company are increasingly prepared at higher management levels. The planning horizons constitute a geometric sequence, as the following list readily shows:

1 week:		weekly production scheduling.
1 month,	4 weeks:	monthly production scheduling.
4 months,	16 weeks:	ABC planning of tools and labour.
1 year,	52 weeks:	capacity adjustment.
4 years,	200 weeks:	production allocation.
10 years,	500 weeks:	strategic planning of company structure.

The progression factor of these major horizons is 3.5. We do not see any good reason to interpolate planning horizons between the major ones, because they do not seem to occur in practice.

c. Size of nations

The above categorization is not only found on the time axis, but also in spatial dimensions. In order to illustrate this, we categorize the nations on the basis of the size of their population. The major echelons in the list to follow reveal a somewhat European bias. Omitting the very small nations with less than one million inhabitants, we have:

Small nations,	4 million:	DK, N, GR.
Medium-size nations,	15 million:	NL, DDR.
Large nations,	60 million:	D, F, GB, I.
Very large nations,	200 million:	USA, USSR.
Giant nations,	1,000 million:	China, India.

We find again a geometric sequence, with progression factor 4.0. It seems to be reasonable to interpolate the following threshold echelons:

small/medium size,	8 million:	A, B, H, S,
medium size/large,	30 million:	E, PL,
large/very large,	110 million:	Japan,

because the respective nations fall typically between the major echelons. The refined sequence of echelons has the progression factor 2.0.

d. Loudness of sounds

Vigorous research in psychophysics has revealed that there is a functional relationship between the intensity of physical stimuli (sound, light, ...) on the one hand and the sensory responses (the subjective estimates of the intensity) on the other. Psychophysics starts from Weber's law (1834), stating that the just noticeable difference Δs of stimulus intensities must be proportional to the actual stimulus level s itself. In Fechner's law (1860), the sensory response $\Delta\Psi$ to a just noticeable difference Δs is supposed to be constant, which implies that $\Delta\Psi$ is proportional to $\Delta s/s$. Integration yields a logarithmic relationship between Ψ and s. Additional experience has finally shown that Fechner's law does in general not hold. Brentano (1874) suggested that the sensory response $\Delta\Psi$ might be proportional to the response level Ψ, so that $\Delta\Psi/\Psi$ would also be proportional to $\Delta s/s$. By integration, one obtains that Ψ would be a power function of s. Empirical evidence in many areas of sensory perception prompted Stevens (1957) eventually to postulate the power law as a general psychophysical law. Thus, with s_1 and s_2 representing intensity levels of a particular stimulus such as sound or light, the sensory and the physical intensity ratios are connected by

$$\frac{\Psi(s_1)}{\Psi(s_2)} = \left(\frac{s_1}{s_2}\right)^{\beta}. \tag{3}$$

The exponent β has been established for many sensory systems under precisely defined circumstances. For a 1,000 Hz tone it is roughly 0.3. It is customary in acoustics to use a *dB*-scale for sound intensities. Thus, the intensity s with respect to a reference intensity s_0 is represented by

$$dB(s) = 10 \log(s/s_0).$$

A difference of 10 *dB* between sound intensities s_1 and s_2 can henceforth be written as

$$dB(s_1) - dB(s_2) = 10,$$

which implies

$$s_1/s_2 = 10,$$
$$\Psi(s_1)/\psi(s_2) = (s_1/s_2)^{\beta} \approx 2.$$

In other words, by a step of 10 *dB* the sound intensity is felt to be doubled. The interesting result for our purposes is that the range of audible sounds has roughly been categorized as follows:

40 *dB*: very quiet; whispering.

60 *dB*: quiet; conversation.

80 *dB*: moderately loud; electric mowers and food blenders.

100 *dB*: very loud; farm tractors and motorcycles.

120 *dB*: uncomfortably loud; jets during take-off.

Although the precision should be taken with a grain of salt because we have a mixture of sound frequencies at each of these major echelons, we obviously find here a geometric sequence of subjective sound intensities with the progression factor 4.

e. Brightness of light

Physically, the perception of light and sound proceed in different ways, but these sensory systems follow the power law with pratically the same value of the exponent β. Hence, a step of 10 *dB* in light intensity is felt to double the subjective brightness. The range of visible light intensities has roughly been categorized as follows:

30 *dB*: star light.

50 *dB*: full moon.

70 *dB*: street lightning.

90 *dB*: office space lightning.

110 *dB*: sunlight in summer.

Under the precaution that the precision should not be taken too seriously because we have at each of these major echelons a mixture of wave lengths, we observe that the subjective light intensities also constitute a geometric sequence with the progression factor 4.

For a more detailed documentation on psychophysics we refer the reader to Marks (1974), Michon, Eykman, and de Klerk (1978), Roberts (1979), Zwicker (1982), and Stevens and Hallowell Davis (1983). The reader will find that the sensory systems for the perception of tastes, smells, and touches follow the power law with exponents in the vicinity of 1. We did not see a categorization such as for loudness and brightness so that we neither have additional evidence nor counter-evidence for the geometric progression described in the above examples.

4. A NATURAL SCALE FOR RELATIVE PREFERENCES

In a basic experiment of pairwise-comparison methods for multi-criteria analysis, two stimuli S_j and S_k (two alternatives A_j and A_k under a particular criterion, two cars under the minimum-price criterion, for instance) are presented to the decision maker whereafter he is requested to express his indifference between the two, or his weak, strict, strong, or very strong preference for one of them. We assume that the stimuli have unknown subjective values V_j and V_k for him, possibly inversely proportional to the distances from a certain upper limit of attractiveness. The purpose of the basic experiments and the subsequent analysis is to approximate these values under the assumption that they have been normalized. The verbal comparative judgement, given by the decision maker and converted into a numerical value r_{jk} is taken to be an estimate of the ratio V_j/V_k. The conversion is based on the results of the preceeding sections, that is, we use a geometric scale to quantify the verbal statements. Such a scale is conveniently characterized by a scale parameter γ, the logarithm of the progression factor $(1+\varepsilon)$. Thus, we set

$$r_{jk}=\exp(\gamma\,\delta_{jk})$$

where δ_{jk} is an integer designating the gradation of the decision maker's judgement as follows:

0: indifference between S_j and S_k.

+2: weak (mild, moderate) preference for S_j versus S_k.

−2: weak (mild, moderate) preference for S_k versus S_j.

+4: strict (definite) preference for S_j versus S_k.

−4: strict (definite) preference for S_k versus S_j.

+6: strong preference for S_j with respect to S_k, etc.

Obviously, weak (somewhat more) preference for S_j with respect to S_k is converted into $\exp(2\gamma)=(1+\varepsilon)^2$, strict (more) preference into $\exp(4\gamma)=(1+\varepsilon)^4$, etc. When the progression factor $(1+\varepsilon)$ is set to 2, we have precisely the ratios for comparative judgement announced at the end of Section 2. It is easy to understand why we set δ_{jk} to 1 if the decision maker hesitates between indifference and weak preference for S_j, etc. In summary, we use the even values of δ_{jk} to designate the major echelons (the major gradations) of comparative judgement, and the odd values for the threshold echelons (the threshold gradations).

It is a matter of course that the results of Section 3 prompt us to propose a geometric scale with $(1+\varepsilon)=2$ and $\gamma=0.7$ as a natural scale for the quantification of the gradations just mentioned. We do not see any reason to maintain the fundamental scale of Saaty (1980). In earlier experiments, we have used a short, normal scale ($\gamma=0.5$) and a long scale ($\gamma=1$), for reasons to be explained at the end of this section. Those scales are still recommended for a sensitivity analysis. Remember that the progression factor of the refined sequences of major and threshold echelons in Section 3 has a progression factor which was *roughly* equal to 2!

We approximate the vector $V=(\ldots, V_j, \ldots, V_k, \ldots)$ of subjective stimulus values via logarithmic regression, that is, we approximate V by the normalized vector $\bar{v}$ which minimizes the expression

$$\sum_{j<k} (\ln r_{jk} - \ln v_j + \ln v_k)^2, \tag{4}$$

where the summation is further restricted to the pairs (j, k) judged by the decision maker. He does not really have to consider each pair of stimuli, an advantage which the eigenvector method of Saaty (1980) signally fails to offer. Minimization of (4) is carried out by solving the associated, linear system of normal equations, with variables $w_j = \ln v_j$. Obviously, the w_j have an additive degree of freedom. The v_j will accordingly have a multiplicative degree of freedom, which is used to single out the normalized vector $\bar{v}$, with components summing up to unity.

By this procedure we calculate stimulus weights for an individual decision maker. We obtain a vector $\bar{v}$ of group weights, possibly a compromise, by minimizing

$$\sum_{j<k} \sum_{d \in D_{jk}} (\ln r_{jkd} - \ln v_j + \ln v_k)^2, \tag{5}$$

where D_{jk} stands for the set of decision makers who judged the pair (j, k), and r_{jkd} for the estimate of V_j/V_k expressed by decision maker d. We are clearly assuming that the values V_j and V_k of the respective alternatives are uniform for the group of decision makers, an assumption that will be further discussed in Section 6. We solve the variables $w_j = \ln v_j$ from the associated, linear system of normal equations, and we use the multiplicative degree of freedom in the v_j to obtain the normalized minimum solution of (5). It is interesting to note that the calculations remain unchanged when each term in (5) is multiplied by the factor p_d, the relative power of decision maker d, for instance the relative size of the state or the constituency which he represents (*see* Lootsma, 1987; Lootsma *et al.*, 1990). Then we minimize the expression

$$\sum_{j<k} \sum_{d \in D_{jk}} (\ln r_{jkd} - \ln v_j + \ln v_k)^2 p_d \tag{6}$$

by solving a linear system of normal equations. We have the impression that the power game in groups has hitherto been ignored in multi-criteria analysis. This might explain why the decision makers sometimes reject the proposed, formalized approach.

It is interesting to note that the rank order of the calculated stimulus weights does not depend on the scale parameter γ. The leading stimulus remains number one. When γ tends to zero, the calculated weights tend to be mutually equal. When γ goes to infinity, the weight of the leading stimulus increases to 1, and the remaining weights converge to 0.

The above procedure is applied m times to calculate the normalized impact scores $\bar{v}_i(A_j)$, $i = 1, \ldots, m$, of the alternatives A_j, $j = 1, \ldots, n$, and only once to calculate the normalized weights $\bar{w}(C_i)$, $i = 1, \ldots, m$, of the criteria (it will be obvious that the criteria can also be taken to stand for stimuli which are considered in pairs). For the time being, we assume that the scores $\bar{v}_i(A_j)$ approximate certain subjective values $V_i(A_j)$ which could be deeply hidden in the mind of the decision makers. Similarly, the $\bar{w}(C_i)$ approximate unknown subjective criterion values $W(C_i)$. Thus, each decision maker carries out at most $m[(1/2)n(n-1)]$ pairwise comparisons to judge the alternatives under the respective criteria, and at most $(1/2)m(m-1)$ comparisons to assess the criteria themselves. As we have seen, not every possible pair has to be presented to each decision maker, but in order to reduce the notorious inconsistency of human judgement, they should consider as many pairs as possible.

In Section 1, we noted that the decision makers find it difficult to choose a gradation for their comparative judgement, particularly when the performance of the alternatives under the given criterion is expressed in physical or monetary units. The categorization of Section 2 will help them to carry out the task properly. In many real-life applications we did observe that the decision makers intuitively turn to such a procedure. They classify the alternatives in a small number of groups (the good ones, the bad ones, and an intermediate group) on a vaguely defined range of attractiveness, whereafter they judge them in pairs via inspection of the classification. The alternatives outside the range are practically dropped from further consideration.

We conclude this section with a note on our earlier choice of a value for the scale parameter γ. In real-life experiments with groups of decision makers (Légrády *et al.*, 1984; Lootsma *et al.*, 1986) we used the verbal statements (indifference, weak, strong, very strong preference) in two different ways: (*a*) we converted them into numerical values on various geometric scales, with trial values assigned to γ, whereafter we applied logarithmic regression [formula (5)] to calculate stimulus weights, and (*b*) we converted weak, strict, strong, and very strong preference into preference without further gradations, whereafter we calculated the stimulus weights via the method of Bradley and Terry (1952) which does not have a particular scale. For practical purposes, the results of (*a*) and (*b*) were sufficiently close when γ varied between 0.5 and 1. The idea is obvious. Method (*a*) can be used when we have a single decision maker only who provides preference information in gradations. For method (*b*), we need a group of 10 or more members, each providing a very limited amount of information in every pairwise comparison: indifference between the stimuli under consideration, or just preference for one of the two. If we assume that the members are in principle subject to identically distributed perturbations and that they have the same stimulus values in the back of their mind, we may compare the results of (*a*) and (*b*) in order to match the scale parameter γ. The analysis of the present paper, however, enables us to choose γ more precisely. We generate the natural scale by setting γ to the value of 0.7. Sensitivity analysis is carried out via the short scale ($\gamma=0.5$) and the long scale ($\gamma=1$).

5. AGGREGATION, FINAL SCORES, AND RANK REVERSAL

Aggregation is a delicate operation in multi-criteria analysis. It is a mathematical operation which may present unexpected results to the decision makers, when the underlying assumptions are ignored.

First, we have to find a common nominator for the operation. In what follows, we shall be assuming that the decision makers express their relative *preference* for the alternatives, under each of the respective criteria. Thus, they are not supposed to choose the qualification "somewhat cheaper" when they compare cars under the consumer-price criterion, but "weak preference" for one of them, etc. Next, we assume that the preference ratios designated by expressions such as "weak preference", strict preference", and "strong preference", correspond to inverse ratios of echelons in a *common* range (D_{min}, D_{max}) on the one-dimensional axis of desirability of the alternatives. The orientation of the categorization is downwards from the maximum desirability D_{max} in the actual decision problem. Thus, taking $D_j = D_{max} - e_j$ and $D_k = D_{max} - e_k$ to denote the desirability of stimuli S_j and S_k respectively, we model the preference for S_j with respect to S_k as

$$\frac{V_j}{V_k} = \frac{e_k}{e_j} = \frac{D_{max} - D_k}{D_{max} - e_j}.$$

In practice, it is difficult to verify whether the intensities of the feelings vary over a *common* interval under each of the criteria, but it is easy to identify situations where the assumption is violated. In the heated debates about the choice of a strategy for the national electricity supply (Lootsma *et al.*, 1991), we found that some participants could reasonably discuss the alternatives under various criteria, but their feelings were intensified to a disproportionate order of magnitude as soon as the safety criterion came under study. The critical issue was nuclear safety! It was obvious that the strength of their feelings fell outside the common range of desirability under the other criteria.

The above assumptions enable us, however, to operate with preference ratios, because they are all defined in terms of distances from the target D_{max} on the common interval (D_{min}, D_{max}). We consider two alternatives A_j and A_k with their calculated profiles, the vectors $\bar{v}_i(A_j)$, $i = 1, \ldots, m$, and $\bar{v}_i(A_k)$, $i = 1, \ldots, m$, respectively. For each i the ratio

$$\bar{v}_i(A_j)/\bar{v}_i(A_k) \tag{7}$$

expressing the relative preference for A_j with respect to A_k under criterion C_i, is unique. Since we are dealing with ratios, it is natural to model the global preference for A_j with respect to A_k by the expression

$$\prod_{i=1}^{m} (\bar{v}_i(A_j)/\bar{v}_i(A_k))^{\bar{c}_i}, \tag{8}$$

where $\bar{c}_i$ simply denotes the calculated weight $\bar{w}(C_i)$ of the i-th criterion. In an attempt to express the global preferences for the respective alternatives by final scores $\bar{f}(A_j)$ and $\bar{f}(A_k)$, we set the ratio

$$\bar{f}(A_j)/\bar{f}(A_k) \tag{9}$$

to (8), whence

$$\bar{f}(A_j)=\prod_{i=1}^{m}(\bar{v}_i(A_j))^{\bar{c}_i}. \tag{10}$$

The final scores have a multiplicative degree of freedom. They can accordingly be normalized to sum up to unity. We tacitly hope, of course, that the final scores approximate the subjective global values $F(A_j)$, $j=1, \ldots, n$, of the alternatives (provided that they exist).

The geometric-mean aggregation rule (10) has the interesting property of "infinite compensation for zero preference". Suppose that the decision maker is indifferent between A_j and A_k, which implies that the ratio (9) is roughly equal to 1. Imagine now that A_j remains fixed but that A_k can be varied continuously. If we take one of the impact scores in the profile of A_k to converge to 0, then indifference between A_j and A_k can only be maintained if at least one of the remaining scores of A_k goes to infinity. It is easy to verify that the arithmetic-mean aggregation rule (2) has "finite compensation for zero preference" only. Hence, the geometric-mean rule does not make it urgent to introduce a veto mechanism as in ÉLECTRE, which rules out certain alternatives with an extremely poor performance under one of the criteria.

A peculiar form of rank reversal in the AHP has been observed by Belton and Gear (1983). They noted that the addition of a new alternative may change the rank order in a set of consistently assessed alternatives. It is easy to verify, however, that the reversal in their example disappears as soon as the arithmetic-mean aggregation rule (2) is replaced by the geometric-mean rule (10) (D. Akkermans, A. Zwagemakers, *TU Delft*, private communication and *M. Sc. thesis*).

It is also important to note that the geometric-mean aggregation rule (8) is based on ratios, which do not depend on the units of measurement! It would be difficult, of course, to measure desirability as such, but in the present method this is not a point of major concern. The ratios are typically dimensionless quantities.

TABLEAU I

Weights and rank order of the criteria, final scores and rank order of the alternative energy research programs in a budget-reallocation study. Comparative human judgement has been encoded on a short scale ($\gamma=0.5$), a natural scale ($\gamma=0.7$), and a long scale ($\gamma=1.0$) to show the scale sensitivity of the final scores and their rank order.

		$\gamma=0.5$		$\gamma=0.7$		$\gamma=1.0$	
Scale values	Indifference	1.0		1.0		1.0	
	Weak Preference	2.7		4.0		7.4	
	Strict Preference	7.4		16.0		54.6	
	Strong Preference	20.1		64.0		403.4	
Weights and rank order of criteria	Security of Energy Supply	26.1	2	27.7	2	29.4	2
	Energy Efficiency	13.3	4	10.8	4	7.7	4
	Long-Term Contribution	27.7	1	30.2	1	33.2	1
	Environmental Protection	25.1	3	26.2	3	27.2	3
	Suitability for Comm.Action	7.8	5	5.1	5	2.6	5
Final scores and rank order of alternatives	Photovoltaic Solar Energy	11.8	3	12.6	3	13.7	3
	Passive Solar Energy	6.9	9	6.0	9	4.7	9
	Geothermal Energy	8.7	7	8.0	7	7.0	7
	Advanced Energy Saving	13.2	2	14.1	2	15.4	2
	Saving in Industry	13.6	1	15.3	1	18.0	1
	Hydrocarbons	9.1	6	8.5	6	7.6	6
	New Energy Vectors	7.7	8	6.8	8	5.5	8
	Biomass Energy	5.6	10	4.4	10	3.0	10
	Solid Fuels	11.7	5	12.0	5	12.0	5
	Wind Energy	11.8	4	12.4	4	13.1	4

Table 1 shows how the weights and the rank order of the criteria as well as the final scores and the rank order of the alternatives, in a projet reported by Lootsma *et al.* (1990), vary with the scale parameter γ. The scale variations are considerable, but the sensitivity of the weights and scores remains within reasonable limits. The rank order of the final scores does not change here. When rank reversal occurs in a real-life project, however, by a sensitivity analysis which shows the weights and scores on the natural scale ($\gamma=0.7$) and on two neighbouring scales ($\gamma=0.5$ and $\gamma=1.0$), one has to warn the decision maker that the rank order of the alternatives has not been established beyond reasonable doubt. In doing so, we are close to the results of the French school of thinking in multi-criteria analysis, which is not always able to rank the alternatives completely in a subjective order of preference. It is important to realize this, because rank reversal is a frequently occurring phenomenon when we vary the numerical scale for comparative judgement.

6. CONCLUDING REMARKS

The main theme of this paper is clearly the transition from the objective evaluation to the subjective weighing of the alternatives in a multi-criteria decision problem. The French school models subjective human judgement via partial systems of binary outranking relations between the alternatives and via a global system of outranking relations. The American school builds partial value functions on the set of alternatives as well as a global value function. Although the American school yields results which are easier to handle in actual decision making, the foundations presented so far are unnecessarily weak. This is what we have shown in the present paper by the introduction of a revised AHP with a natural geometric scale to quantify comparative human judgement.

Via logarithmic-regression analysis we calculate impact scores and criterion weights to approximate hypothetical, subjective values hidden in the heads of the decision makers. The impact scores approximate the values $V_i(A_j)$ of the alternatives under the respective criteria; these values are supposed to be uniform for the group of decision makers. Similarly, the criterion weights approximate the values $W(C_i)$, uniformly valid for all decision makers involved in the actual decision problem. If some of these values do not exist, or if they are heavily dependent on the decision makers, we cannot really provide decision support. We cannot guarantee that we help the decision makers to identify their pre-existing subjective values, but we manipulate them towards an (unexpected) compromise solution. In those cases we merely have a mathematical and computational technique which reforms the decision proces. It structures and accelerates the deliberations. Decision reform cannot immediately be rejected on moral grounds. We suggest the reader to imagine how a decision should be made when the underlying assumptions (in fact, existence of a consensus) are violated.

Aggregation is based on the strong but reasonable hypothesis, that the desirability of the alternatives varies over a common interval. What we implicitly presuppose, is a group of even-tempered decision makers. Under such an assumption, the final scores approximate the common global preferences deeply hidden in their heads. The hypothesis also reveals the limited scope of multi-criteria analysis. In the words of French (1988), decision theory is the mathematics of rationality. Irrationality is the driving force behind human decision making, however, and that is not incorporated in the methods for decision support.

The transition from objective evaluation to subjective weighing is not merely a technical question for experts in multi-criteria analysis. On several occasions we found that mathematical and computational tools to support or to reform the subjective weighing are not welcome at all. The RAND Corporation, for instance, a large American consulting company commissioned with the PAWN project (Policy Analysis of Water Management for the Netherlands, *see* the PAWN reports, 1981-1983) by the Dutch water management authority Rijkswaterstaat, designed a large number of alternative strategies for surface-water control. They rejected multi-criteria analysis for the final selection of a particular strategy, on the ground that the decision makers would have to agree explicitly on the criterion weights and on the impact scores. Although such an agreement is not necessary (the decision makers are completely independent in the execution of pairwise comparisons, and they only have to agree on the calculated compromise solution), the RAND Corporation still prefers to display the consequences of the alternative strategies (expressed in their original physical or monetary units) on coloured cards which would help the decision makers to see the comparative strengths and weaknesses of the strategies (W. Walker, private communication). The subjective weighing should unconditionally be left to the decision makers themselves, regardless of whether they are able to digest the flood of information without a structured multi-criteria analysis. Incidentally, the example shows that American management consultants do not unanimously follow the American school. Conceptually, the RAND Corporation is closer to the French school of thinking which typically concentrates on measurable criteria.

We have frequently proposed the structured approach of multi-criteria analysis as a tool for advisory councils to come up with a unanimous advice. The General Energy Council in the Netherlands, however, rejected the suggestion (*see* Lootsma *et al.*, 1991) because they did not want to arrive at such a group compromise solution. They agreed that a unanimous advice would be stronger than a set of divided recommendations, but they insisted that the final decision should be taken by the responsible authorities, confronted with the rich variety of views and opinions in the Council. This implies that the subjective weighing is largely left to these authorities, regardless of whether they have the time and the capacity to do so properly.

The fortunes of these projects, however, show that there is an urgent need to improve the subjective weighing in actual decision making. The American school did not go far enough; in exploring its potential we found that the results are not always far away from what the French school would produce.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge Bernard Roy and LAMSADE at the University of Paris-Dauphine for the stimulating discussions and the excellent hospitality (December 17, 1989-January 17, 1990). I am also indebted to Prof. dr. ir. F. A. Bilsen, Dr. ir. J. Raatgever (Faculty of Technical Physics), Mrs. Ir. M. Kwak and Mr. F. A. Vos (Faculty of Mathematics and Informatics, Delft University of Technology) for their valuable support and criticism.

REFERENCES

O. H. Barda, Recherche d'une méthode appropriée pour l'étude des problèmes de choix de sites pour l'implantation des centrales thermiques, Document 54, LAMSADE, Université de Paris-Dauphine, 1989.

V. Belton and A. E. Gear, On a Shortcoming of Saaty's Method of Analytical Hierarchies. *Omega*, 1983, *11*, pp. 227-230.

R. A. Bradley and M. E. Terry, The Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons, *Biometrika*, 1952, *39*, pp. 324-345.

J. B. Brans, B. Marechal and P. Vincke, Prométhée: a New Family of Outranking Methods in Multi-Criteria Analysis, *in* J. P. Brans Ed., Operational Research 84, *North-Holland*, Amsterdam, 1984, pp. 408-421.

G. Colson and C. de Bruyn, Models and Methods in Multiple Objective Decision Making, *Math. Compt. Mod.*, 1989, *12*, pp. 1201-1211.

S. French, Decision Theory, an Introduction to the Mathematics of Rationality, *Ellis Horwood*, Chichester, 1988.

R. Keeney and H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs, *Wiley*, New York, 1976.

K. Légrády, F. A. Lootsma, J. Meisner and F. Schellemans, Multi-Criteria Decision Analysis to Aid Budget Allocation, *in* M. Grauer and A. P. Wierzbicki Eds., Interactive Decision Analysis, *Springer*, Berlin, 1984, pp. 164-174.

F. A. Lootsma, Modélisation du jugement humain dans l'analyse multi-critère au moyen de comparaisons par paires, *R.A.I.R.O. Recherch. Oper.* 1987, *21*, pp. 241-257.

F. A. Lootsma, Numerical Scaling of Human Judgement in Pairwise-Comparison Methods for Fuzzy Multi-Criteria Decision Analysis, *in* G. Mitra Ed., Mathematical Models for Decision Support, *Springer*, Berlin, 1988, pp. 57-88.

F. A. Lootsma, J. Meisner and F. Schellemans, Multi-Criteria Decision Analysis as an Aid to the Strategic Planning of Energy R & D, *E. J. Oper. Res.*, 1986, *25*, pp. 216-234.

F. A. Lootsma, T. C. A. Mensch and F. A. Vos, Multi-Criteria Analysis and Budget Reallocation in Long-Term Research Planning, *E. J. Oper. Res.*, 1990 (to appear).

F. A. Lootsma, P. G. M. Boonekamp, R. M. Cooke and F. Van Oostvoorn, Choice of a Long-Term Strategy for the National Electricity Supply via Scenario Analysis and Multi-Criteria Analysis, *E. J. Oper. Res.*, 1991 (to appear).

L. E. Marks, Sensory Processes, the New Psychophysics, *Academic Press*, New York, 1974.

J. A. Michon, E. G. J. Eijkman and L. F. W. de Klerk Eds., Handboek der Psychonomie (in Dutch), *Van Loghum Slaterus*, Deventer, 1976.

The PAWN reports I-XX. The RAND Corporation, Santa Monica, California, USA/Rijkswaterstaat, Hooftskade 1, The Hague, The Netherlands, 1981-1983.

F. S. ROBERTS, Measurement Theory, Addison-Wesley, Reading, Mass., 1979.

B. ROY, Classement et choix en présence de points de vue multiples, R.A.I.R.O. *Rech. Opér.*, 1968, *2*, pp. 57-75.

B. ROY, Méthodologie multicritère d'aide à la décision, Collection Gestion, *Economica*, Paris, 1985.

B. ROY, The Outranking Approach and the Foundations of ELECTRE Methods, Document 53, LAMSADE, Université de Paris-Dauphine, 1989.

Th. L. SAATY, The Analytic Hierarchy Process, Planning, Priority Setting, Resource Allocation, *McGraw-Hill*, New York, 1980.

Th. L. SAATY, What is the Analytic Hierarchy Process? *in* G. MITRA Ed., Mathematical Models for Decision Support, *Springer*, Berlin, 1988, pp. 109-122.

A. SCHÄRLIG, Décider sur plusieurs critères, *Presses Polytechniques Romandes*, Lausannes, 1985.

S. S. STEVENS, On the Psychophysical Law, *Psychol. Rev.*, 1957, *64*, pp. 153-181.

S. S. STEVENS and M. D. HALLOWELL DAVIS, Hearing, its Psychology and Physiology, American Institute of Physics, New York, 1983.

E. ZWICKER, Psychoakustik, *Springer*, Berlin, 1982.