VI.B. A Delphi Study of Factors Affecting the Quality of Life

NORMAN C. DALKEY

1. Earlier Related Studies

In a previous publication [1] several Delphi studies are reported using respondent groups to identify and estimate linear weights for those aspects of experience, which they judged to be important in determining the quality of life or sense of well-being of an individual. The procedure was first to ask each respondent to list a certain number of aspects of 'experience which he thought were most important in determining the quality of life of an individual. This number varied from five `to ten over the different exercises--depending mainly on the resources available to process the lists, In general, this open-ended round produced a long list of nominated factors-typically two hundred to three hundred. Inspection of the lists showed a fair amount of similarity among some of the items, but very few identical pairs. These long lists were compacted by the experimental team by aggregating items that appeared to have high similarity, producing an intermediate list of about fifty items. A matrix was formed and each respondent was asked to judge the similarity of each item to every other item. (In a 50 X 50 matrix, this required 1225 estimates, since the similarity relationship was presumed to be symmetrical.) The combined similarity matrices were then analyzed by the hierarchical clustering routine developed at Bell Telephone Laboratories by S. C. Johnson [2] and the hierarchy was truncated at lists of twelve or thirteen clusters.[1]

The aggregated lists were then presented to the respondents for judgments of relative importance. In general, the exercises indicated that group relative importance ratings produce reasonable ratio scales, and that the reliability. of such judgments across randomly selected groups is high.

---

[1]In this procedure objects are clustered according to the (judged) similarity between them. The objects within a cluster are more similar to one another than to objects belonging to a different cluster. In addition, the procedure merges' similar clusters into larger clusters in a step-wise fashion until all the objects are, placed into a :single cluster. This hierarchy of clusters can be used in its entirety, or, depending on the purposes of the experiment, truncated at some appropriate level to furnish a list of aggregated items.

Research reported herein was conducted under Contract Number F30602-72-C-0429 with the Advanced Research Projects Agency, Department of Defense.

2. Models of the Quality of Life

The general model of individual well-being, underlying these studies is that quality of life is a function of the individual's location in a "quality space"; that is, the sense of well-being enjoyed by an individual depends upon the extent to which his experiences exemplify several basic characteristics. Inherent in this view is the existence of trade-offs among these characteristics; two different individuals can enjoy about the same level of quality of life with highly different patterns of experiential rewards. One can be comfortable, receiving a great deal of love and affection, and living a routine existence; the other can be living an exciting life, with a high sense of achievement, and lonely, and each report about the same overall level of well-being.

What is not spelled out in this general view is how the contribution of each component is reflected *in* overall quality of life. There are two broad possibilities. One, by far the simpler of the two, is that overall quality of life is simply a weighted arithmetical sum of the individual's status on each component. In symbols

$$Q(i) = \sum_j q_{ij} w_j + C_i, \tag{1}$$

where Q(i) is individual i's *overall* quality of *life*, q_{ij} is status on quality j w_j is the relative importance of quality j, and C_i is a constant added to balance differences of scaling on Q and the q_{ij}. The essence of this model is that each quality makes a fixed relative contribution to' overall Q, independently of where the individual stands on all the other qualities.

The second model assumes' that the relationship is more complex, and the contribution of any one quality depends *on* where the individual stands with respect to all the others. That is, the relative importance relationships are not fixed, but are functions of the individual's location *in* the quality space. These two different hypotheses can be illustrated (using, for simplicity, only two of the thirteen qualities) as in Fig. 1 and Fig. 2. What is diagrammed in each case is the set of lines of equal quality of life. In both' cases, quality of life (QOL) increases with *a* simultaneous increase in sense of achievement and in affluence. But in Fig. 2, the amount of increase from achievement depends on how affluent the individual is, whereas this is not the case for Fig. 1.

Although existing data are not very helpful in deciding between these two types of models, it seems likely that a nonlinear model is, at least in principle, the more accurate representation. In Rokeach's survey study of American values [3], the low-income group ranked "comfort" at level 6, whereas the high-income group ranked it at level 15 (lower numbers indicating higher importance with a total range of 1-18); conversely, the low-income group ranked a sense of accomplishment at 12, and the high-income group ranked it at level 5. A possible interpretation of these data is expressed in Fig. 2 where the low-income individual P would gain more by a small increase in affluence (vertical arrow) than he would by a small increase in achievement (horizontal arrow), whereas the reverse is true for the high-income individual R.

Although the nonlinear model seems likely to be more correct, the linear model may be a fairly good approximation. From previous studies we can say that there is a large amount of uncertainty concerning the relative importance of qualities, and it appears reasonable to assume a fair amount of uncertainty in the individual's judgment of his own well-being. Under conditions of uncertainty on all parameters, a linear model often gives as good an approximation as a potentially more precise model where uncertainty in the data obscures the more complex interactions.[2]

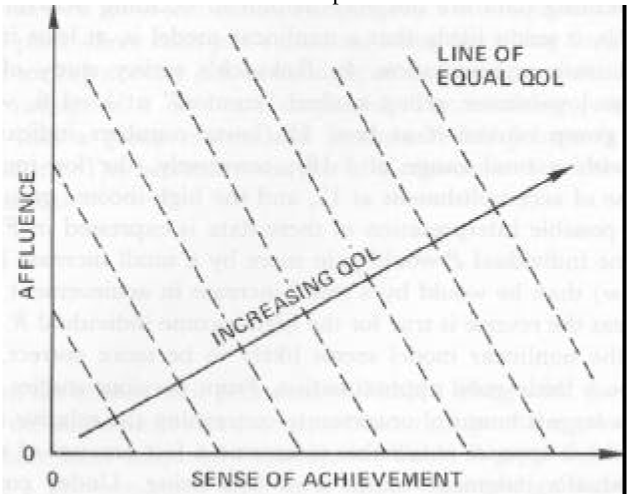

Fig. l. Illustration of linear model of quality of life.

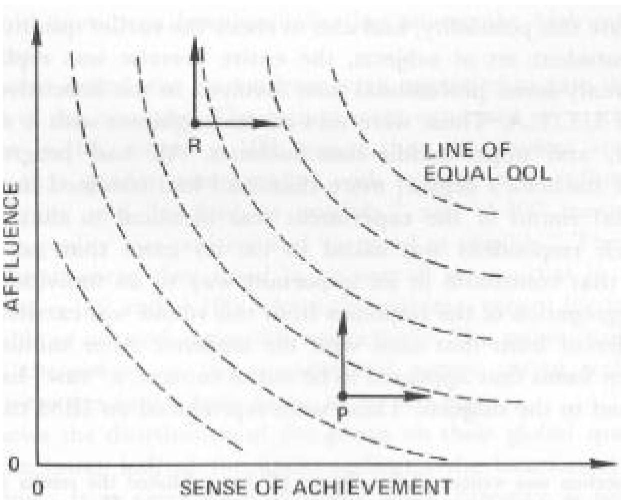

Fig. 2. illustration of nonlinear model of quality of life.

2 Since this section was written, Robin Dawes [41 has published the results of several studies showing that a much stronger statement can be made in this regard. This is taken up more fully in the final section of this paper.

3. Procedure

To investigate this possibility, and also to check the earlier quality of life studies with a nonstudent set of subjects, the entire exercise was replicated with a group of twenty-seven professional men involved in the Executive Engineering Program at U.C.L.A. These were mid-career engineers with a median age of thirty-seven, and upper-middle-class incomes. All had progressed at least through the bachelor's degree; more than half had obtained master's degrees.

The initial round of the experiment was identical to that of the earlier studies; each respondent was asked to list no more than seven aspects of experience that contribute in an important way to an individual's quality of life. The aggregation of the responses from this round was carried out in a way slightly different from that used with the students. After minimal editing to remove a few items that appeared to be out of context, a "raw" list of 227 items was presented to the subjects. These were reproduced on IBM cards, and each subject had his own set of 227 cards. The subjects were requested to sort the set of cards into piles where each pile contained items describing roughly the same quality. The sorting activity can be -thought of as a truncated similarity rating, where only ratings of 0 or 1 are employed. The resulting judgments were transformed into a group matrix and processed with Johnson's hierarchical clustering routine. The hierarchy was cut at a convenient level to furnish a list of twelve clusters.

4. Results

The resultant list of clusters is presented in Table 1 along with the medians and quartiles of the subsequent relative importance ratings of the group. The items listed are those most frequently mentioned. Table 2 gives the list of clusters identified in a previous study with forty upper-class and graduate students at U.C.L.A.

There is not a great difference between the two either in terms of clusters, or in terms of judged relative importance. Privacy and Sex do not appear as separate items in Table 1, and in Table 1 Achievement is separated from Respect and Prestige, which are combined in the student list. Other minor differences are apparent. There is probably not much to be gained by attempting to probe these differences. One evident difference is the number of ties for the top rating by the engineer group. There is no clear indication from the data why they should have been less discriminating among the high-rated clusters than the students.

In a subsequent round, the respondents were requested to rate their present status on each of the qualities. The ratings were expressed on a scale ranging from -100 to + 100, where -100 meant that the present status of the respondent on that quality was negative with respect to its influence on his sense of well-being and "as bad as possible," and + 100 meant that the influence of the quality' was positive and "as good as possible." They also were requested to assess where they stood in an overall sense-that is, to assign a number between -100 and + 100, where the extremes meant life in general is as bad as possible or as good as possible respectively. The respondents were also asked to rate themselves on the conventional survey rating scale, namely, whether they were *very happy*, *fairly happy*, or *not very* happy.

Figure 3 shows the distribution of the group on their global quality of life ratings, and the lower half of the figure indicates the location of the verbal happiness ratings. It is clear that the *fairly happy* category is not very discriminating, using the numerical ratings as a standard.

Table 1
Quality of Life Clusters and Relative Importance Ratings, Engineer Group

	Relative Importance		
Quality	Median	Lower Quartile	Upper Quartile
Self-confidence, self-esteem, self-knowledge, pride	80	70	100
Security, peace of mind	80	70	97
Sense of achievements, accomplishment, success	80	75	90
Variety, opportunity, freedom	80	60	90
Receiving and giving love and; affection	80	50	90
Challenge, intellectual stimulation, growth	75	60	80
Comfort, congenial surroundings; good health	70	50	80
Understanding, helping and accepting others	60	30	75
Being needed by others, having friends	60	40	70
Leisure, humor, relax ation	53	50	75
Respect, social acceptance, prestige	40	20	70
Dominance, leadership, aggression	30	10	50

The items are ordered first on the median-then on the upper quartile; and then on the lower quartile in case of ties.

Table 2

QOL Factors, Student Group

		Relative Importance (Median)
1	Self-respect, self-acceptance, self-satisfaction; self-confidence, egoism; security; stability, familiarity, sense of permanence; self-knowledge, self-awareness, growth	100
2	Love, caring, affection, communication, interpersonal understanding; friendship, companionship; honesty, sincerity, truthfulness; tolerance, acceptance of others; faith, religious awareness	96
3	Peace of mind, emotional stability, lack of conflict; fear, anxiety; suffering, pain; humiliation, belittlement; escape, fantasy	91
4	Challenge, stimulation; competition, competitiveness; ambition, opportunity, social mobility, luck; educational, intellectually stimulating	80
5	Achievement, accomplishment, job satisfaction; success; failure, defeat, losing; money, acquisitiveness, material greed; status, reputation, recognition, prestige	79
6	Sex, sexual satisfaction, sexual pleasure	78
7	Individuality; conformity; spontaneity, impulsive, uninhibited; freedom	76
8	Social acceptance, popularity; needed, feeling of being wanted; loneliness, impersonality; flattering, positive feedback, reinforcement	75
9	Involvement, participation; concern, altruism, consideration	72
10	Comfort, economic well-being, relaxation, leisure; good health	63
11	Novelty, change, newness, variety, surprise; boredom; humorous, amusing, witty	61
12	Dominance, superiority; dependence, impotence, helplessness; aggression, violence; hostility; power, control, independence	58
13	Privacy	55

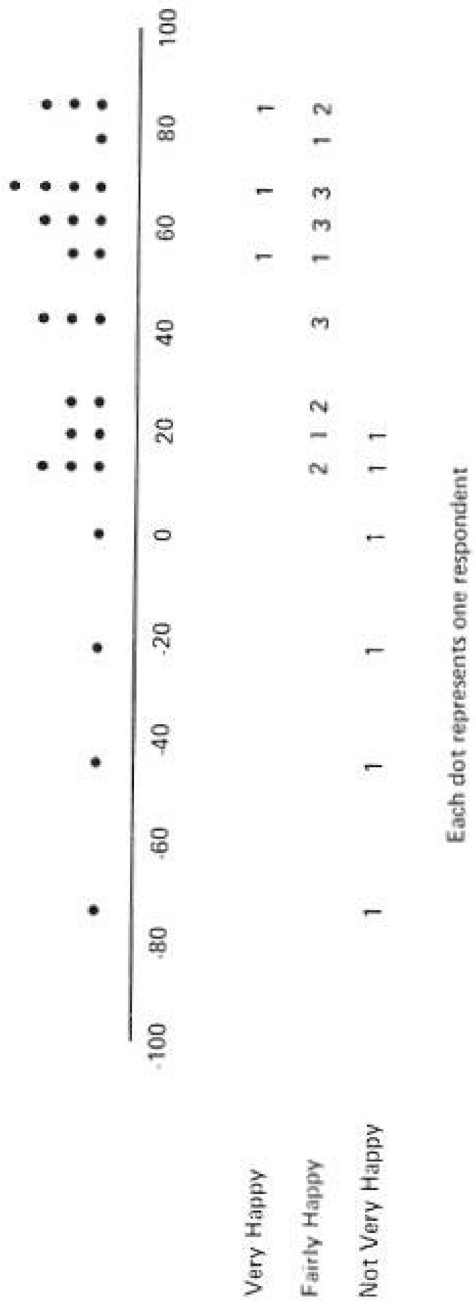

Fig. 3. Distribution of overall scores (present status judgments) and happiness ratings.

The engineer group is somewhat divergent from the general public in their verbal assessments. In particular, the proportion reporting *very happy* is definitely smaller than for the general public. [5] On the other hand, the numerical ratings appear to give a different picture --- 40 percent of the group rated themselves at 70 or higher. It would appear that a much more careful investigation is needed of the relationship between numerical and verbal ratings on the happiness question.

Given the individual self-ratings on qualities and on overall quality of life, several forms of the linear model can be tested. The general form of the test is to predict the overall ratings from the ratings on qualities. This can be done, using formula (1) where we can use for the w_j's either the individual's own relative importance ratings, or the group relative importance ratings. In addition, empirical weights can be computed by determining the best linear prediction of the overall ratings from the ratings on qualities.

When the overall ratings are computed using the individual's own relative importance numbers, the correlation between actual and computed scores is .57. When the same computation is made using group relative importance ratings, the correlation between actual and computed scores is .58. These correlations are not particularly impressive. They indicate roughly 33 percent (the square of the correlations) of the variation in the overall ratings has been accounted for by the predicted ratings. In addition, the group relative importance ratings do not give a better prediction than the individual relative importance ratings.

The results of the linear estimation computation are displayed in Table 3. The multiple correlation associated with this set of linear estimation weights is .84, hence the proportion of the variance accounted for is about 70 percent. This would generally be considered a favorable degree of accordance of the empirical linear model with the data. However, a number of considerations are raised by the computation. The weights bear little relation to the relative importance ratings of the group; the Spearman rank-order correlation between the two is -.16. The two items rated lowest and next to lowest by the group receive the third highest and highest estimation weights. In addition, two items that are given high weights by the computation, namely Respect and Variety, have negative coefficients quite contrary to the perception of the group.

If the results of the computation were taken literally they would indicate that the group has a relatively clear perception of the basic ingredients of their sense of well-being, but that they have a very poor perception of the actual relative importance of the qualities. In particular, Dominance and Respect, which the group rates as least important, would be among the most important; and Security and Love and Affection, which the group ranks high, would be comparatively unimportant. Variety and Respect would make negative contributions to quality of life.

5. Additional Analyses

There is no way to reject this interpretation on the basis of the data as it stands, and it certainly furnishes some intriguing hypotheses for further investigation. To cast some additional light on the stability of the qualities identified by *the* cluster technique, several factor analytic studies of the data were carried out. In the most extensive study, nine factors were identified. Only the first six factors were selected for rotation. (These six accounted for .807 of the variance; the next three: increased the variance accounted for to .826.) The six factors and the loadings of the qualities on them is presented *in;* Table 4. The first factor is a mixture of Self-confidence, Respect, and Understanding others -a genera-lized esteem factor. The second factor is almost entirely Dominance. The third might be considered a generalized success factor, involving Sense of achievement, Variety, and Comfort:: The fourth is an interesting and rather puzzling combination of Love and Affection , and Challenge. The fifth is primarily Leisure. The sixth *is* difficult to interpret. The two highest loadings - for Comfort and Being Needed are negative.

Table 3
Linear Estimation Weights for Qualities

Self-confidence	-0.29281
Security	-0.08595
Achievement	0.56667
Variety	-0.49070
Love and Affection	0.11297
Challenge	-0.06746
Comfort	0.51529
Understanding	0.12987
Needed	0.02671
Leisure	0.22044
Respect	-0.79423
Dominance	0.53622

All of the qualities that have high estimation weights in Table 3 have high loadings on no more than one factor; In addition, for all the factors except the rather mysterious factor six, there is a relatively sharp separation between the few qualities that load highly on the factors and all the others. In this respect, the factors are relatively "clean." This consideration lends some weight *to* the assumption- that most of the aggregated qualities defined by the cluster analysis are fairly sharply defined.

The exercise is probably too small to draw any firm conclusions. It is possible that the counterintuitive results of the linear estimation analysis stem from nonlinear trade-offs. The variations within the group in self-ratings on the qualities are large (standard deviations range from 39 to 55) and thus the members of the group are scattered widely in the quality space, and nonlinearities in the trade-offs could make major differences in the contributions of the qualities to the overall score of different subjects. In short, the linear model may not be appropriate.

Another possibility is that despite the high multiple correlations obtained, an omitted variable is producing systematic effects. If the data on overall scores are analyzed by a 2 X 2 breakout on income and age, as in Table 5, a clear indication is, obtained concerning the combined effects of these two. The numbers in the table are the median overall ratings of the members of the designated classes. The numbers in the small boxes refer to the number of cases. The younger, lower-income group tends to rate itself higher than the older, higher-income, group. The relationship between age and income is so strong (the off-diagonal cells have few members) that it is not possible to distinguis h the separate effects of the two variables. It is rather surprising that all the negative self-ratings occur in the high-income, older group, and all the ratings of very happy occur in the younger, lower-income group.

Table 4

Factor Analytic Study

		Factor					
	Variable	1	2	3	4	5	6
1	Security	0.25873	0.008	0.30332	0.19134	0.3348	0.29012
2	Leisure	-0.02679	-0.03369	-0.03813	-0.09306	0.9847	-0.01437
3	Challenge	0.06144	0.20352	0.08204	0.72517	-0.05388	0.17499
4	Self-confidence	0.64276	0.0334	0.17468	0.26906	-0.02419	0.02358
5	Variety	-0.15179	0.10266	0.6355	0.0917	0.0723	-0.05646
6	Understanding	0.68341	0.09297	-0.13627	0.28061	0.18222	0.00691
7	Achievement	0.20224	-6.0824	0.94888	-0.05396	-0.07511	0.07892
8	Comfort	0.10102	0.00348	0.58909	0.11555	0.05377	-0.45189
9	Needed	0.12594	0.03455	0.20275	0.34421	0.28623	-0.42441
10	Love and Affection	0.07256	-0.04199	0.03769	0.87896	0.0277	-0.22396
11	Dominance	0.02457	1.03902	-0.04365	0.05515	-0.03478	0.00276
12	Respect	0.61814	0.33582	0.17243	-0.23114	-0.05515	-0.07596

The relationship between age and income and overall ratings appears to be stronger than those found for the general public, and as far as income is concerned, counter to the relationship found in the general public, where increasing income is correlated with reports of higher happiness [5].[3]

Some of the anomalous features of the linear estimation weights in Table 5 may reflect the combined role of age and income. Thus, one would expect a fairly high correlation between age-income and prestige.

[3] However, the survey results may not be extrapolable to the income range involved here. Most of the survey data have a tap category of $15,000 or over. A recent (March 1972) telephone survey conducted by the advertising firm Batten, Barton, Durkine, and Osborn indicated that individuals with high

Table 5
Analysis of Overall Self-Ratings in Terms of Income and Age

		AGE	
		35 or less	Over 35
INCOME	20 K/year or less	70 [7]	50 [3]
	Over 20 K	50 [4]	30 [13]

The numbers in small boxes refer to the number of cases.

6. Discussion

The issue is left open in the exercise whether the individual, in making a judgment concerning his own quality of life, is estimating some directly perceived internal state (e.g., a "feeling tone") or whether he is, in effect, using his status on the various quality dimensions as a basis for estimating a nonperceived condition. In this respect, the QOL estimation task is formally similar to a number of perceptual' or judgmental tasks - e.g., a psychiatrist making a diagnosis of mental illness on the basis of certain symptoms, or an individual estimating distance using various visual cues.

There is a fairly extensive literature on the type of model most appropriate for such multidimensional judgments, beginning with the work of, Brunswick. For many perceptual judgments, objective measures exist by which the judgements can be evaluated. For nonperceptual tasks --- e.g., the psychiatrist's diagnosis --- the situation is often similar to the QOL judgment in that there is no readily available objective measure.

Robin Dawes has recently reported on a series of studies of multidimensional judgments of both sorts-clinical estimation of degree of psychosis based on the set of test scores furnished by the Minnesota Multiphasic Personality Inventory, prediction of grade-point averages by faculty screening committees, and the like [4]. In these studies, weights for the putative parameters selected at random gave much more accurate results (measured by correlation with the judgments of a criterion team in the case of the clinical diagnosis and against subsequent actual grade-point averages in the case of the performance predictions) than the judgments of the individuals (or in the case of the clinical judgment, of the groups as well). Equal weights did even better. Dawes concludes that for these relatively poorly determined estimation "tasks, human judgment may be relatively good in selecting the relevant dimensions, but relatively poor in determining the weights to be attached to those dimensions.

In this respect, Dawes' conclusions are quite compatible with the results reported earlier in the paper. Since writing the previous sections, the QOL exercise has been replicated with a group of thirty-two graduate students attending classes on Futures Studies at UCLA and UC, Irvine. Once again, the linear estimation of self-rated QOL gave a high R^2 (.75 in this case), but the estimation weights differed significantly from those obtained with the engineers.

There appear to be several issues raised by the present study as well as those of Dawes and others for Delphi methodology. One of the basic tenets of the Delphi approach to decision problems has been that for uncertain questions, where solid data or theories are lacking, informed judgment is the only resource available (leaving aside the option of postponing judgment until better information is available).

Dawes has suggested that for certain kinds of judgments, we may be better off with nominal assumptions than with human judgment. The evidence he presents warrants taking the suggestion seriously:

One possibility here --- which I have to admit is highly tentative-is that an extension of what is often referred to as sensitivity analysis may be called for. For example, in the case of multidimensional estimates, if the correlation with some criterion is to be the figure of merit for the estimates, then the analysis may be highly insensitive to the precise form of the model or the weights attached to a linear estimation.

As a specific example, consider a two-dimensional problem where a quantity Q= Q(x,y) is to be estimated. Suppose Q is a multiplicative function of x and y, i.e., Q= xy. How close an approximation is the linear function L = x + y? To make the computation simpler, assume that x and y have been normalized so that the region of interest is the unit square. Assume that x and y are uniformly distributed over the unit square. This is the least favorable assumption, since any linear correlation between x and y would increase the correlation between Q and L.

We have

$$\rho Q,L = \int_0^1\int \frac{(xy-\overline{xy})(x+y-\overline{x+y})}{\sigma_{xy}\sigma_x+y}\,dx\,dy.$$

Performing the integration and evaluating gives

$$\rho Q,L = \frac{\sqrt{6}}{\sqrt{7}} = .925.$$

As another illustrative 'example, let Q = ax + by and R = cx + dy. Then

$$\rho Q,R = \frac{ac+bd}{(a^2+b^2)^{1/2}(c^2+d^2)^{1/2}},$$

where, if a- 1, *b-2, c-2, d = 1,*

$$P = .8.$$

Here, the error-free correlations are sufficiently high so that it would be questionable if the noisy kind of data met with in practice could distinguish the models.

These examples give some feeling why linear models are so "robust" when correlation with a criterion is the figure of merit. But more to the point, they suggest that ordinary human judgment would probably not do as well as nominal assumptions, if the underlying dimensions have been correctly identified. By a nominal assumption here is meant an assumption based on some quasi-logical rule such as symmetry or "equal ignorance." A good example is the assumption frequently met in statistical estimation that a priori probabilities are equidistributed over the set of hypotheses. Attempts to give a logical foundation to such an assumption (based on "insufficient reason" or equal ignorance arguments) have failed. And yet the assumption often works quite well in practice.

These comments whittle away the region for which human judgment is most appropriate. They do not, of course, directly affect the advantage of group judgment (Delphi) over individual judgment, for those cases where human judgment is the best available (or recognizable!) form of information.

Acknowledgments:

The author was aided in many ways in conducting the study by Daniel Rourke, now at Rockefeller University, New York. The author would also like to express his appreciation for the help of Professor Bonham Campbell of the Engineering Systems Department, UCLA and the members of the Executive Engineering Program for 1971-72. Methodological portions of the research were supported by the Advanced Research Projects Agency.

References

1. Dalkey, Rourke, Lewis, and Snyder, Studies in the Quality *of Life, D.* C. Heath and Co., Lexington, Mass., 1972.
2. S C., Johnson, "Hierarchical Clustering Schemes," Psychometrika 32, No. 3 (1967), pp. 241-54. See also Carroll-Wish, Chapter VI, C, this volume, for a fuller discussion of cluster techniques and their role in Delphi studies.
3. M. Rokeach, "Values as Social Indicators of poverty and Race Relations in America," The Annals *of* the American Academy *of* Political and Social Science, No. 388, *(March* 1970), pp. 97-111.
4. R, M. Dawes, "Objective Optimization under Multiple Subjective factions," presented at the Seminar on Multiple Criteria Decision *Making,* Columbia, South Carolina, October 1972.
5. *Paul* A. David and Melvin W. Reder, eds., *Nations* and *Households* in Economic Growth: Essays in Honor *of* Moses Abramovits (New York: Academic Press, *Inc.,* 1974).