Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing

1988

FOREWORD

These are proceedings of the Second Sawtooth Software Conference, which took place at Sun Valley, Idaho in April, 1988. The conference topics were Computer Interviewing, Conjoint Analysis, and Perceptual Mapping.

The speakers furnished written papers for this proceedings. In most cases the papers were more extensive than the spoken remarks, but some papers, being in outline form, were less so. The papers are published essentially as received, having required only minor editing that the authors have accepted.

Our speakers were conscientious in preparing for their talks, and there is information in the papers that should be useful to many readers. We thank them all for their efforts.

As always, we invite our readers to comment, and are particularly grateful for suggestions about how to make our conferences even more successful.

Richard M. Johnson
June 20, 1988

CONTENTS

CONJOINT ANALYSIS

Solving Practical Problems

PERCEPTUAL MAPPING

Major Research Companies and Perceptual Mapping Practices

SPECIAL SECTION

COMPUTER INTERVIEWING

Special Opportunities and Problems

Disks-By-Mail: A New Survey Modality

CONJOINT ANALYSIS: ITS RELIABILITY, VALIDITY, AND USEFULNESS

Dick R. Wittink and John W. Walsh
Cornell University

Abstract

We review some of the available evidence on the reliability, validity, and usefulness of conjoint analysis. The results tend to be favorable; the procedure in its many variants appears to perform reasonably well, and has been used for many different purposes. Many questions remain, however. It is especially important to know the degree of uncertainty associated with predictions based on a conjoint study. We also suggest opportunities to improve the conduct and evaluation of applications.

INTRODUCTION

Conjoint measurement was introduced to the marketing literature by Green and Rao (1971). Johnson (1974) introduced the technique of tradeoff analysis. Over time, many variations of what is now referred to as conjoint analysis have been suggested. The methodology has become a standard tool in the repertoire of market research techniques offered by commercial outlets. In this paper, we review selected studies that have focused on the reliability, validity, and usefulness of conjoint analysis. In the discussion we indicate what aspects have been investigated, and we provide an interpretation of the results. We suggest that the importance of reliability lies in the basis it provides for quantifying the uncertainty inherent in conjoint analysis. If this uncertainty is quantified for a study, then it is also possible to determine whether the actual result from a decision based on conjoint analysis is within the bounds of uncertainty. Such bounds provide a meaningful basis for assessing the predictive validity of results obtained. Without more precise information on the reliability and validity of conjoint analysis, we can only speculate about its true usefulness.

Our primary focus in this paper is on the reliability and validity of conjoint analysis. We define reliability as consistency or agreement in results between equivalent or comparable conditions. Validity is defined as the extent to which the results are correct or apply to the marketplace. Although we discuss these two aspects separately, it is often difficult to know which one applies in a given study. We identify below some of the many options available to the conjoint analyst for the conduct of a study. Our knowledge about the relative superiority of one option over another is limited. To the extent that we cannot argue on theoretical or empirical grounds why one method for conjoint should be preferred over another, it seems appropriate to consider the various sources of variability in results under the heading of reliability.

RELIABILITY

Campbell and Fiske (1959) define reliability as the agreement between two efforts of measuring the same trait through maximally similar methods. Much of the literature on reliability, especially in psychology, is focused on the importance of employing "reliable" instruments for assessing a variety of characteristics or traits. In conjoint analysis, reliability is desired for the preference judgments respondents provide. Green and Srinivasan (1978) describe two tests of reliability - at the level of input judgments and at the level of estimated parameters. Test-retest reliability is determined by asking a respondent to provide preference judgments for a second set of concepts, some time after a first set is evaluated. A subset of the concepts employed in the first set is used in the second set. The correspondence between the judgments about the common concepts in the two sets is an index of the reliability of the input data for that respondent.

An example of reliability measured at the level of parameter estimates is provided by Parker and Srinivasan (1976). A respondent is asked to complete a second conjoint task, some period of time after the first task has been completed. The second set of stimuli does not contain any concepts used in the first set, although there is no difference in the number and identity of attributes and attribute levels between the two sets. Also, the same number of stimuli is used in the two tasks. A coefficient of equivalence is calculated from the product-moment correlations[1] of the two parameter vectors estimated from the two stimuli sets' preference judgments provided by a respondent.

Reliability of the input judgments can only be assessed across a common set of concepts. Thus, this measure is limited to a quantification of variability in the input data, perhaps involving instability over time. By focusing on the estimated parameters, additional sources of unreliability can be considered. For example, it is common to use a fractional factorial design to identify a subset of all possible objects for respondents to evaluate. If we are indifferent between two or more alternative fractions, we believe that each fraction is equally valid for the collection of preference judgments. In that case it is of interest to determine the consistency in the parameter estimates obtained for alternative fractions of objects. Note, however, that we require a preference model and an estimation procedure for such a reliability assessment. Thus, our choices on those aspects may affect the reliability results.

Apart from time and the set of stimuli, there are many other sources of systematic or random error. These sources include: method of data collection (e.g. full profiles, tradeoff matrices; Jain et al. 1979, Wittink et al. 1982), definition of preference (e.g. intention to buy, overall liking), measurement scale for preference judgments (e.g. ratings, rank order; Jain et al. 1979, Leigh et al. 1984), number and definitions of attributes (e.g. Scott and Wright 1976, Malhotra 1982), number and identity of attribute levels (e.g. Reibstein et al. 1987, Wittink et al. 1982), stimulus design characteristics (e.g. uncorrelated or correlated attributes), method of presenting options to respondents (e.g. descriptions, pictures), means of collecting preference judgments (e.g.

pencil and paper, computer interactive; Finkbeiner and Platz), order of presenting attributes (e.g. Acito 1977, Johnson 1981), preference model (e.g. main effects only, with interaction effects), estimation procedure (e.g. Wittink and Cattin 1981, Srinivasan et al. 1983, Green 1984), and the number and types of respondents.

Considering the large number of sources of measurement error, it is easy to argue that we do not yet know much about the degree of consistency in results across the many alternatives. Of course, we can sometimes argue on a priori grounds in favor of one method or another. For example, if we believe that there are important attribute interaction effects, we may exclude the tradeoff matrix approach from consideration. Thus, both reliability (do we obtain consistent results across options?) and validity (do the results pertain to the phenomenon we are interested in?) issues tend to apply when we evaluate the sources of measurement error listed above.

Studies have been completed on many of the sources of error listed above. Unfortunately, the conclusions usually do not point in one direction. For example, several studies have examined differences in instability between methods of data collection (Jain et al. 1979, Leigh et al. 1984, Segal 1982, Reibstein et al. 1987). Conflicting results exist regarding the difference in reliability between data collection procedures. Jain et al. (1979) and Leigh et al. (1984) obtain no significant differences between alternative data collection methods. Segal (1982) finds a superior result for the full-profile method, and Reibstein et al. (1987) show both higher and lower reliability for the tradeoff matrix approach (dependent upon the manipulation, as discussed below). The actual results obtained may well depend on the product category, the number and characteristics of respondents, as well as the choices made on other elements of the studies.

The most comprehensive study of reliability is by Reibstein et al. (1987). They investigate three different sources of unreliability: over stimulus sets, over attribute sets, and over data collection procedures. Their results focus on the differences in reliability between three data collection methods, by varying the stimulus set and the attribute set. In addition, the number of levels for one attribute is varied. Importantly, they use five highly different consumer products, thereby enhancing the generalizability of their results.

In a comprehensive study such as the one by Reibstein et al. many issues have to be addressed. To evaluate their results, it is necessary to identify the components of their study. For example, in many conjoint applications respondents have an opportunity to gain experience with the conjoint task prior to the collection of preference judgments. This opportunity was not provided to the respondents in the study by Reibstein et al. Consequently, it is possible that respondents may learn about their own preference structures in a first task and that learning influences subsequent preference judgments. In addition, respondents gain familiarity with the task itself. Such systematic differences in respondents' knowledge and experience between two tasks may, of course, affect the reliability results. To some extent, these possible systematic differences can be examined. For example, respondents' learning about their preference structures may show up as a higher goodness of fit for the second task, for a given product category. Increased familiarity with the task may result in a higher goodness of fit for all tasks subsequent to the first (each respondent provided judgments for four sets of stimuli across two product categories).

We now turn our attention to the specific elements manipulated in the reliability investigations by Reibstein et al. In discussing the characteristics of their study, we focus on the comparison of results across the data collection approaches.

Reliability Over Stimulus Set. To estimate this type of reliability, two different fractional factorial designs, two different sets of tradeoff matrices, and two different sets of paired comparison profiles are used. It is easy to show that the degree of similarity between two fractional factorial designs for the profiles is high. One way to examine this similarity is to assume at least one two-level attribute among the attributes used. In that case it follows that two mutually exclusive and collectively exhaustive designs differ only in the levels of this one two-level attribute (see Figure 1). If this hypothetical attribute is also the least important, respondents will not notice a great deal of difference between the two sets of profiles. Of course, what this suggests is that the choice of a particular fractional design may be relatively inconsequential. In other words, it is an attractive feature of this type of design that two alternative designs are potentially quite similar. We show in Figure 1 that two subsets of eight profiles each are identical on the first three attributes (a, b and c), while the levels of the fourth attribute (d) are reversed.

FIGURE 1

Full and Fractional Factorial Designs for a Study
With Four Attributes, Two Levels Each,
Using Concept Evaluations*

Fraction A

a	b	c	d
1	1	1	1
1	1	2	2
1	2	1	1
1	2	2	2
2	1	1	2
2	1	2	1
2	2	1	2
2	2	2	1

Fraction B

1	1	1	2
1	1	2	1
1	2	1	2
1	2	2	1
2	1	1	1
2	1	2	2
2	2	1	1
2	2	2	2

*The full factorial design is the set of all 16 objects

For matrices, alternative designs differ in the attribute pairs used. If one design contains half of all possible matrices, it follows that the other design must have the other matrices. In this case, there is no overlap at all between the two sets. For example, if attribute A is paired with B in the first set, it would <u>not</u> be paired with B in the second set. Given that the reliability measure is based on the similarity of parameter estimates, and that the parameter estimates express the influence of one attribute relative to other attributes, it is clear that the specific pairs used may have a systematic influence on the parameter estimates. In practice, there are of course more than two sets of matrices that can be used. We show in Figure 2 three alternative designs, each one containing

FIGURE 2

Full and Fractional Designs for a Study
With Four Attributes, Two Levels Each
Using Tradeoff Matrices

The pairs are ab, ac, bd, cd (fraction A)

ab, ad, bc, cd (fraction B)

ac, ad, bc, bd (fraction C)

The full design consists of all six matrices ab, ac, ad, bc, bd, cd

four matrices. However, by using two mutually exclusive and collectively exhaustive sets of matrices the reliability calculated is likely to be the lowest possible. We discuss in the Appendix, assuming lexicographic processing by a respondent, the effect of fractionation on the reliability of results for the data collection methods. Of particular interest is the substantial difference in results between full profile and tradeoff data.

In addition, if the attributes differ in the number of levels, the matrix sizes differ between the two sets as well. Thus, the preference rank orders for the cells are not directly comparable across matrices (and across the two designs). Standardization of the preference ranks may be appropriate, but this is unlikely to eliminate the noncomparability of parameter estimates both within and between the two sets of matrices.

Similarly, for paired comparison profiles, the percentage of times an attribute level is used in the two sets may still not be equal. Standardization should probably be used here as well, although again the parameter estimates may still not be comparable.

These considerations suggest that the full profile data collection approach should show a higher degree of reliability than the other data collection methods. Indeed, Reibstein et al. show an average alpha[2] level of .42 for profiles, .37 for paired comparisons, and .22 for tradeoff matrices, as reproduced in Table 1 (the first three columns). Obviously, other factors not considered here are likely to influence differences across product categories. But the pattern is fairly consistent across the ten observations.

TABLE 1

Average Type I Error (Alpha) for the Comparison of Individual-Level Parameter Estimates

	Reliability Over Stimulus Set			Reliability Over Attribute Set		
	Full Profiles	Tradeoff Matrices	Paired Comparisons	Full Profiles	Tradeoff Matrices	Paired Comparisons
	.32[1]	.32	.25	.43	.94	.55
	.41	.17	.42	.38	.91	.88
	.32	.12	.42	.44	.34	.83
	.37	.03	.43	.49	.84	.90
	.47	.14	.23	.62	.86	.71
	.37	.20	.26	.91	.80	.71
	.33	.23	.34	.61	.95	.97
	.50	.14	.42	.64	.95	.88
	.44	.40	.48	.72	.91	.74
	.64	.47	.40	.67	.94	.91
Average	.42	.22	.37	.59	.84	.81

---

[1] The ten observations in each column represent the average alpha values (averaged across respondents) for five product categories times two price level variations.

<u>Reliability Over Attribute Set</u>. To examine the influence of attribute substitutions on reliability, Reibstein et al. delete the attribute judged to be least important by management, and substitute another unimportant attribute in the second design. To simplify this substitution, the two attributes have the same number of levels. The degree of similarity between the two fractional factorial designs should again be high, especially if the substitute attributes are the two least important to respondents.

In the tradeoff matrix approach, the attribute pairs used do not differ between the two designs, except for the substitution of one attribute for another. These two substitute attributes have the same number of levels. Therefore, the attributes that are used in both sets are used in matrices that do not differ in the dimensions between the two sets. If a is paired with b in the first set, and neither a nor b is one of the least important attributes, a is also paired with b in the second set. Thus, in the matrix approach, if there are five matrices in each set, only two matrices differ at all between the two designs. And, if the substitute attributes are indeed both least important, the relative results obtained for the common attributes should be very stable. It is also important to note that while the comparability of the preference ranks across matrices of different sizes is still an issue, this issue does not affect the *comparison* of common parameter estimates across the two designs.

For paired comparison profiles, the issue of comparability across attributes exists as well. However, the comparison of parameter estimates for a *given* attribute across two designs is not affected if the paired comparisons differ only in the attribute substitution.

These design issues suggest that the full profile data collection approach may show similar results for attribute set reliability and stimulus set reliability. However, the tradeoff matrices (and by implication, paired comparisons) should show greater attribute set reliability. This is true, especially, if the standardization issue (which does not affect *attribute* set reliability in the study by Reibstein et al.) is severe. The results obtained by Reibstein et al., as reproduced in Table 1, include an average alpha level of .59 for profiles (somewhat higher than for stimulus set reliability), an average alpha of .84 for matrices (much higher than for the stimulus sets), and .81 for paired comparisons (also much higher than for stimulus sets).

We suspect that both the design issue (different attribute pairs in the stimulus set reliability determination) and the noncomparable parameter estimates (different matrix dimensions across the two designs) contribute to the lower reliability for matrices than for profiles in the stimulus set comparison. Obviously, the specific pairs selected may influence the parameter estimates since they represent *relative* values. This is, therefore, an issue worth serious consideration.

Results such as those provided by Reibstein et al. provide useful insights into the reliability of conjoint results. However, a great deal is still required. For example, we need to know whether reliability scores for each of the sources of measurement error listed above are independent. The Reibstein et al. results suggest that they are not (for example, stimulus set reliability depends on the data collection method). We also need to identify the appropriate focus of reliability for a particular study. For example, if we want to make conclusions about individual consumers, we desire acceptable reliability levels for the individual-level parameter estimates. On the other hand, if the primary purpose of a study is to make market share predictions (for example, for a modified product), we may only be concerned about aggregate-level reliability. Reliability at the aggregate level may be assessed by contrasting the *distribution* of parameter vectors for all respondents across two measurements. Thus, we do not necessarily want to know how consistent the vectors of parameter estimates are for each individual respondent. It turns out that the measure of reliability used by Reibstein et al., while computed at the individual level (and then averaged), is closer to an aggregate-level measure (Wittink and Walsh, 1988).

Finally it is important to keep in mind that higher reliability does not necessarily mean higher validity. For example, a particular data collection method may allow respondents to simplify the task, or to structure the task of providing preference judgments in such a way that the results are consistent. However, the simplification or the structure used may be unrepresentative of market behavior. Thus, it is better to use a structure for the task that incorporates important elements of market behavior, even if this structure can be demonstrated to have lower consistency than some other structure that is not representative of the market environment.

VALIDITY

It is obvious from the previous discussion that knowledge about the reliability of conjoint analysis is, at best, incomplete. That is, there is still a great deal more to be learned about it. But also, no matter how much we know about reliability, we can only favor the use of conjoint if it provides valid information. By validity we mean the extent to which the results can be generalized to an external environment, such as marketplace choices. Often the focus is on predictive validity (also referred to as external validity, see Green and Srinivasan, 1978).

Even though conjoint analysis has received a great deal of attention and use in both the academic and market research communities since the early 1970's, there is still little formal evidence regarding its validity. Of course, its intensive use as an aid to management decisions (Cattin and Wittink 1982, 1986) is one kind of testimony of its validity. However, few formal investigations have been completed regarding the absolute and relative validity of the methodology. Measures of absolute validity are important for the determination of conjoint analysis' applicability to specific problems. In addition, various types of relative validity can be distinguished. For example, the technique is likely to be more valid for new products that represent modifications of existing ones than for new-to-the-world types of new products. We also should concern ourselves with the question of the relative superiority (inferiority) of conjoint analysis for certain kinds of problems, compared with alternative methods. And, of course, there are opportunities to distinguish the relative validities for different variants of conjoint analysis.

It would seem unnecessary to discuss why validity measures are important. However, the fact that so few validity results are known may suggest that having evidence of validity is not especially critical nor useful. Perhaps in many cases it is easy to justify an application of conjoint on its theoretical merits. However, there are strong reasons for insisting on more precise guidance. For example, if management decisions are based on conjoint results it seems appropriate for management to insist on a quantification of the uncertainty underlying the information obtained and of the risk inherent in any decisions made. The danger of over-representation or optimism is perhaps best illustrated by the recent lawsuit of Beecham against Yankelovich (involving predictions based on a simulated test-market model). Some predictive validity results exist in the area of simulated or pretest-market models (e.g. Silk and Urban 1978). Typically, an estimated standard error (which would be based on both reliability and validity issues) is computed from a comparison of actual and predicted results across a number of different applications of a given model. However, precise knowledge of the uncertainty associated with a particular prediction, based on unique characteristics of the specific study, would have been required to avoid problems like those encountered by Yankelovich in the Beecham case.

A Hierarchy of Validity Measures. The validity of conjoint analysis can be determined at a number of levels. We discuss several different measures, in the order of strength or severity of the validation effort. For example, at the lowest level of severity we can examine the interpretability of substantive conjoint results. Such interpretation may take place by examining parameter estimates averaged across all respondents. This type of validation is often referred to as face validity. At a somewhat more severe level, we can examine whether the parameter estimates differ across respondents in a manner that is consistent with certain demographic or socioeconomic characteristics (e.g. Currim et al. 1981). To the extent that hypotheses are postulated prior to data collection, such examinations may also fall under the heading of theory testing. For example, Krishnamurthi and Wittink (1987) compared the conjoint results for automobile mileage with expectations based on economic dictates. They found a reasonable degree of validity in the conjoint results at the individual level. However, a higher degree of validity was obtained when respondents were asked to make a direct assessment of the value of improvements in mileage (a form of self-explicated part worths).

Often the primary reason for a conjoint analysis study is to have the ability to make market share predictions. Thus, we may compare predicted with actual market shares for products available in the marketplace at the time of data collection. A necessary condition for the interpretability (validity) of parameter estimates is that such market share predictions are accurate. An application is provided by Clarke (1987). Given the subjectivity involved in the assessment of face validity for parameter estimates, we argue that the predictive accuracy of market share estimates is a more severe test. More demanding still is a comparison of actual and predicted choices at the individual respondent level (e.g. Parker and Srinivasan 1976). Comparisons at the individual level are more severe, because at the individual level the accuracy is also a function of the uncertainty about individual-level parameter estimates. This uncertainty goes to zero when the focus is on market share predictions and the number of respondents becomes large. We note, however, that the market share predictions typically require a weighting scheme. The weights may be based on the frequency or likelihood with which respondents purchase items in the product category under study. Thus, the validity of market share predictions is also influenced by the accuracy of the weighting scheme, as well as by the extent to which the respondents can be assumed to represent the population of all consumers of the product category.

The most severe test of validity involves a comparison of actual and predicted results, subsequent to the implementation of a decision based on the conjoint analysis study. This type of test is more severe than the previously discussed ones for two reasons. One, the marketplace results pertain to a different environment. For example, management may have introduced a new product. The greater the uniqueness of this new product relative to existing items, the more severe this test. In addition, the possible reaction of competitors[3] would further increase the test's severity. Two, the results are obtained some time period after the data collection. Obviously, the longer this time period, the more difficult the test. Difficulties involve both the fact that consumer preferences are not stable, and that advertising and other marketing programs not explicitly considered in conjoint analysis may alter consumer choice.

It is often sufficient to examine the predictive validity at the level of an aggregate market. Thus, the test may be based on a comparison of actual and predicted market shares, especially if we can assume that total product category demand is exogenous. However, if the conjoint results also are the basis for a determination of the best prospects (e.g. market segment) for, and perhaps the components of, a marketing campaign, then the predictive validity should also be determined at appropriate disaggregate market levels. Individual-level predictive validity is useful to know if one of the purposes is to identify the most promising individual targets for a specific marketing program.

The distinction between individual- and market-level predictions is important, as shown by Hagerty (1986). For example, we can employ conventional statistical tests, at the level of the individual respondent, to distinguish between simple (e.g. main effects only) and complex (e.g. with interaction effects) models. These tests often favor the simple models. However, Hagerty shows that for market share predictions these test results are influenced by the amount of error variance associated with the individual-level parameter estimates. At the market share level, this error variance tends toward zero. Indeed, Hagerty argues that the error of market share predictions is not much greater if nonexistent interaction effects are included. On the other hand, if interaction effects are ignored at the individual level, when they do exist, the increase in prediction error of market shares is substantial. These results are most interesting, because we tend to make comparisons between simple and complex models at the individual level. At the individual level, simple models are often favored even if they are incomplete, because the tradeoff between bias and variance favors bias at that level.[4]

Only a few studies exist that document the predictive validity of conjoint analysis at the individual level for a marketplace environment that was relatively unknown to respondents at the time of data collection. Wittink and Montgomery (1979) determined the extent to which individualized preference functions, using tradeoff matrices on eight attributes, predicted job choices made by MBA students. Given the fact that job offers do not have the same characteristics across individuals, predictive validity could only be determined at the individual level. For 48 respondents who provided preferences based on the tradeoff matrix data collection approach, the proportion of individual choices correctly predicted was 63 percent. This compares to 26 percent in case of random choice (p. 70).

For a different group of MBA students, a comparison was made between individuals who had accepted a job offer prior to the day on which preference judgments were obtained and individuals who had not yet accepted a job. This study used two data collection methods. Interestingly, the group of individuals who had not yet accepted a job turned out to have higher predictive validity (62 versus 45 percent for the tradeoff matrix approach, 49 versus 25 percent for the full profile approach). Although there are many possible explanations for this difference, there is some reason to believe that a motivational difference between the groups is responsible. For example, the individuals who had not yet accepted a job reported a greater amount of time used to provide responses (27 versus 23 minutes for tradeoff matrices, 41 versus 31 minutes for full profiles). And this group also showed a higher average confidence[5] in the responses provided (5.5 versus 5.0 for tradeoff matrices, 4.8 versus 4.5 for full profiles). On other response scales, such as the adequacy of job characteristics, there was no average difference between the two groups. These results suggest that individuals faced with a choice problem may have greater motivation to provide valid preference judgments (especially if they believe that the task is helpful to them).

The existence of systematic differences in predictive validity, as detailed above, suggests that it is useful to delete respondents who lack the motivation to participate in the task. A common test is to delete respondents who show a high degree of inconsistency in their preference judgments. Another implication is that it may be wise to avoid respondents who have just completed the purchase decision of a large item[6]. In addition, it may be possible to enhance the validity of preference judgments by asking respondents to imagine that they have to make a choice at the time of the study. Wright and Kriewall (1980) use a "state of mind" manipulation in a study of college choice. They expect higher predictive validity when subjects are told to imagine that the choice among alternatives is imminent, as opposed to sometime in the future.

Although the predictive validity is far from perfect, it is useful to keep the following points in mind. One, in the MBA job choice studies, eight attributes is too small a number for complete and detailed job preference model specifications (the average adequacy of the job characteristics in the study was only 4.5, for both groups of respondents, on a seven-point scale, seven being fully adequate). Two, eight attributes is too many for the traditional full profile data collection method. Procedures are available to accommodate larger numbers of attributes (e.g. Green 1974). Three, some of the prediction errors at the individual level may cancel out at the aggregate level. Four, even though job offers may be largely unknown to the respondents, at the time preference judgments about hypothetical jobs are collected, the prediction problem may not be similar to marketplace options. For example, job offers are made consecutively, while product options are available simultaneously in most product categories. In addition, job offers may include an option that dominates or is dominated by one or more other options. In the marketplace, the set of products available to consumers tends to be characterized by negatively correlated attributes (measured across the set of products). For example, one brand may have the best level on quality, but is worse ("higher cost") on price, while another brand has a lower quality and is better ("lower cost") on price. There is some evidence that a compensatory (e.g. part worth) model estimated from choices made from an orthogonal set of profiles may perform poorly when the validation sample consists of a non-dominated set of profiles and the actual choices are dictated by a non-compensatory model (Johnson et al. 1987).

Conjoint versus Self-Explicated Weights. Although it is useful to know about the validity of conjoint analysis results, it is also important that the method is preferable over alternative procedures. For example, we may require that conjoint provides higher predictive validity than any form of self-explicated weights. Several studies suggest that the predictive validity (typically evaluated at the individual level) is higher for conjoint (e.g. Green et al. 1981, Neslin 1981, Akaah and Korgaonkar 1983). Other studies suggest the opposite result (e.g. Wright and Kriewall 1980, Leigh et al. 1984). Due to the large number of differences in the way studies are conducted, it is difficult to explain these inconsistent results. Obviously, the superiority (inferiority) of conjoint should depend on such factors as the number of attributes, the operationalization of self-explicated weights, and the characteristics of the subjects.

Recently, Srinivasan (1987) proposed to model choice as a two-stage process. The first stage consists of the elimination of options (objects) with attribute levels that are unacceptable to a particular respondent. In the second stage, a compensatory model is applied to distinguish between the remaining (acceptable) options. For the second stage, self-explicated weights are obtained based on measures of attribute importances and the desirability of attribute levels. The importances are defined in terms of the difference between the best level and the worst of the acceptable levels. Srinivasan obtained slightly higher relative accuracy (compared with a random choice model) for the self-explicated approach, based on 1982 MBA job choice data than was obtained with the tradeoff approach by Wittink and Montgomery on 1979 job choice data.

Other Considerations. To the extent that commercial applications of the method include predictions of market results, it is relevant to ask for what time period these predictions are made. Generally, the implicit assumption appears to be that the predictions apply instantaneously, subject to the complete availability (full distribution) of a new product, and awareness (complete brand awareness and knowledge of features) on the part of the target market. Given the likelihood of incomplete availability and awareness, it is helpful to make this dependency explicit as is done in pretest market models (e.g. Silk and Urban, 1978). By allowing availability and/or awareness to change over time, the market predictions become time dependent.

A related question involves the specification of conditions outside those included in the conjoint analysis. For example, the results may be conditional upon the advertising budgets for each of the competitors, or upon the attention the press may give to recent research findings on aspects related to the values respondents associate with product features.

We suggest that in commercial applications claims about the reliability and validity of the results are made for specified time periods and market conditions. In addition, such claims should reflect the primary purpose of the study. For example, for a given study, the validity of aggregate market predictions may be acceptable, while the validity of market segment differences may not be. Unfortunately, much of the validity knowledge that exists today is likely to be proprietary, and may be specific to a particular firm due to unique data collection, analysis, or other aspects. The commercial community at large would be greatly aided by ongoing studies that provide a quantification of the magnitude of invalidity resulting from each of the choices that have to be made for a particular study.

USEFULNESS

Although there is curiously little formal evidence about the validity of conjoint analysis, its heavy use for commercial applications suggests there is a great deal of confidence in the method. We assume, therefore, that both absolutely and relatively the validity is acceptable. Obviously, validity is required before we can ascertain the method's usefulness.

Cattin and Wittink (1982) have identified the frequency with which conjoint analysis involved specified purposes, such as new product identification and pricing, during its first decade of commercial use. In an update, Cattin and Wittink (1986) found that new product identification is still the most frequently cited specified purpose. However, competitive analysis is the second most frequently cited purpose. On a set of rating scales, respondents to the update survey on the commercial use of conjoint analysis turned out to have the highest degree of usefulness of the method for product modifications. Insight into the impact of competitors' actions was accorded a relatively low degree of usefulness (below prices or price changes, new product opportunities, and segment differences). Yet, it is conceivable that respondents' judgments of the method's usefulness for competitors' actions is influenced by the difficulty of anticipating what competitors might do. Thus, the frequency of use may indicate that the insights are valuable, while the degree of usefulness rating may reflect other aspects. For an application of the method for evaluating alternative competitive scenarios, see the Clark Material Handling case (Clarke, 1987).

It is also clear that the method has limited applicability to predicting the outcome of new product introductions. For example, it is difficult to imagine that the method is directly applicable for the assessment of market potential for a new-to-the-world type of new product. There are several reasons for this. One, members of the target market for whom the new product is intended would lack experience with the product category. Most respondents have a hard time providing valid preference judgments about disposable diapers if they have not had occasion to use the item on a child. Two, there is uncertainty about the extent to which respondents will purchase from a product category that is still under development. Three, respondents' preferences are likely to be influenced by opinion leaders and early adopters of new products, as well as by the marketing communication campaigns.

The aforementioned difficulties associated with the use of conjoint analysis for new-to-the-world types of new products can perhaps be overcome. For example, the selection of respondents can be concentrated among what are expected to be the early adopters. And these individuals can first be provided with an opportunity to experience the product. This is a common part of test marketing. In one application, AT&T used conjoint analysis to obtain preference judgments of customers who had used their picture phone meeting service, when it was still in a test market phase. The information obtained was meant to guide the further development as well as the pricing of the service.

A high degree of usefulness of the method appears to be associated with the estimation of the market value provided by specific features, if added to existing products. Procter and Gamble makes frequent use of the method to improve its understanding of the market for certain product categories. The results obtained in one recent application were used to estimate the damages associated with the alleged copyright infringement by a competitor. At a minimum the method can help separate the sales or profit effects of multiple product features introduced simultaneously. Thus, its usefulness is especially promising when actual market data are incapable of providing the insights management desires. Currim et al. (1981) used the method to estimate the relative values respondents associated with distinct features of a performing arts series. Of particular interest to management was the value of these features relative to ticket prices and discounts for season tickets. Historical market data were of little use because the discount percentage had been constant for many years.

A different application of conjoint analysis to the recreational area involved the value of a prominent professional basketball player to the Houston Rockets[7]. By the inclusion of attributes involving attendance, season ticket prices, and the win-loss percentage, estimates were obtained of the marginal value of the player to the team's owners. This information could then be compared with the salary demands of the player. Conjoint analysis is a logical procedure for this type of problem, given the inability to use market data.

Even when the product category is not new to respondents, there may still be considerable uncertainty about the quality (validity) of their preference judgments. For example, one or more profiles may be perceived to be unrealistic or impossible to provide. Automobile manufacturers have experienced some difficulty with the use of car specifications that were technically feasible but perceived by respondents to be infeasible. At a minimum, it seems appropriate to allow respondents to disregard profiles they consider unrealistic. However, it may also be possible to educate respondents about new developments prior to eliciting their preference judgments.

Closely related to this idea of informing respondents about new technological developments is the notion of showing commercials. Commercials are often tested against each other prior to the selection of a new advertising campaign. Often these tests are limited to aspects such as brand awareness, respondent knowledge of key ideas, etc. There is, however, considerable potential for the integration of advertising campaigns with the use of conjoint analysis. Specifically, if conjoint analysis is used to identify product modifications that will suit the target market, an advertising campaign must also be developed to communicate the benefits provided. Currently, an advertising campaign tends to be selected subsequent to and independent of the selection of product modifications. Conceivably, the simultaneous determination of product features and advertising campaigns results in a different outcome. But unless an advertising campaign is purely informative (e.g. communication of brand name and key features), it is useful to estimate the influence of an advertising campaign on product preferences.

The most common conjoint application still appears to consist of the elicitation of preference judgments that form the basis for conditional choice model specifications. That is, conditional upon a purchase from the product category, we predict which of the items included in a specific competitive scenario each respondent should choose. This prediction is then combined with other information about the frequency with which the respondent uses the product category to obtain market share and sales predictions. However, for a subset of product categories this frequency of usage can be made a function of the product features offered. In addition, it may be important to assess multiple brand use, either involving different members of a household unit or organization, or based on the notion of variety-seeking behavior for frequently-purchased items.

CONCLUSION

The reliability of conjoint analysis results seems to be acceptably high. Even so, we do not know enough about the variability in results introduced by each of the design choices that has to be made in an application. Much of this information has to be gathered by manipulating design features in empirical studies. It is, however, possible to identify some instability in results by comparing specific design options, and making assumptions about respondent evaluations (see, for example, table A1 in the Appendix). Unfortunately, the magnitude of unreliability for a given option (e.g. stimulus set), has been shown to depend on other characteristics (e.g. data collection method). Thus, it is impossible to treat each of the sources of variability as being independent of each other. In addition, the conclusions about reliability depend on the purpose of the study. If the primary purpose is to predict the market share for a new or modified product, the focus should be on the reliability of results at the aggregate level. We should also keep in mind that one design option may produce higher reliability than another, and yet the second option may provide information with higher external (predictive) validity.

To broaden our knowledge about validity, we must insist on systematic comparisons between the predictions, based on conjoint analysis, and the actual results obtained. Most of the validity results apply to the individual-respondent level, even though managers are primarily interested in the accuracy of aggregate-level predictions. The difficulty with the individual-level predictions is that inaccuracy may be substantially influenced by the uncertainty associated with individual-level parameter estimates. Much of this uncertainty cancels out at the aggregate level. Indeed, it has been demonstrated that standard statistical tests, applied to a comparison of alternative models at the individual level, are likely to favor simple models. Yet, more complex models estimated at the individual level are expected to be superior when the focus is on predictive validity at the market level.

Although we still do not know enough about the reliability and validity of conjoint analysis in its traditional forms, the field is moving on to new approaches. Perhaps the most promising innovation is the use of computer-interactive data collection. Substantial advantages may be claimed for this approach. For example, instead of using the same set of attributes for each respondent, the set can be adapted to the respondent's orientation. Similarly, the number and definition of objects about which preference judgments are desired can vary among respondents. Both of these components allow a researcher to control the degree of reliability. And, if all other conditions are comparable to the traditional data collection approaches, the reliability of results should be greater for computer-interactive conjoint analysis.

By using stimuli that are specifically designed for a given respondent, the task should be more relevant as well. This characteristic is expected to enhance the (predictive) validity, independent of the positive impact on validity (at least at the individual level) resulting from higher reliability. Thus, we expect improved results when computer-interactive procedures are used. Whether actual comparisons will demonstrate superiority depends on other factors, such as the manner in which the approach is implemented, and the extent to which respondents feel comfortable with a computer terminal.

There is also a tendency to rely on self-explicated attribute weights, often in conjunction with conjoint analysis. Hybrid models have achieved some popularity. Yet, it is conceivable that, at least for some applications, self-explicated weights are preferable to conjoint analysis. The interpretation of self-explicated results requires that precise and well-calibrated questions are asked. Srinivasan (1987) provides a good example of how such an approach can be implemented.

Typical applications still use uncorrelated attributes for the construction of a set of stimuli and a compensatory main-effects model. Although the use of uncorrelated attributes has the advantage of maximum efficiency for the parameter estimates, there is concern about the possible systematic influence due to the inclusion of dominant or dominated objects.

This concern is especially pertinent if the set of stimuli available in the marketplace is characterized by (negatively) correlated attributes. In addition, there is doubt about the robustness of both the compensatory nature of the model and the exclusion of interaction effects.

To enhance the opportunity for all commercial users of conjoint analysis to learn about the validity of conjoint analysis, we suggest that a clearinghouse be established for the collection of predictions from conjoint analysis, and the actual results obtained. Individuals interested in getting access to the data base, consisting of studies conducted by many different parties, would be able to learn about the difference between actual and predicted results across many studies. In addition, with information on design and other characteristics of the studies, bias and variance in prediction errors can be obtained for a subset of the results available. Similarly, it would be possible to estimate the influence of design characteristics on the predictive validity of a study. It is now time to set this process in motion so that ten years from now, we are not still wondering how accurate our results really are.

Footnotes

1. The product-moment correlation as a measure of reliability is subject to certain limitations. For example, the larger the within-vector variation in parameters, the greater the correlation between the values across the two vectors, ceteris paribus. Similarly, the measurement of agreement for the input judgments, based on correlation, may be influenced by the measurement scale.

2. Alpha quantifies the similarity between parameter estimates across two data sets. Specifically, alpha is the probability of making a type I error if the null hypothesis of equal parameter vectors across two data sets in rejected. Thus, the higher alpha, the greater the reliability.

3. The predicted results are, of course, adjusted for competitive reactions that involve the attributes included in the conjoint study. Thus, if both the firm conducting the study, and one or more competitors make changes in their product or marketing mixes, the environment changes more dramatically than if changes are limited to one firm.

4. It is interesting to note that Hagerty (1985) shows how the predictive accuracy of conjoint analysis can be improved by using a Q-type factor analysis. This method is shown to be superior to the grouping of respondents based on cluster analysis. Hagerty also shows that the use of Q-type factor analysis tends to be superior to maintaining entirely separate parameter estimates for each respondent. However, this superiority is evaluated at the individual level. Based on Hagerty (1986), we expect such superiority to disappear at the aggregate (e.g. market share) level. Indeed, to the extent that the "optimal weighting" scheme recommended provides biased estimates (e.g. when each respondent has unique parameter estimates), the bias can outweigh the reduction in variance obtained with factor analysis.

5. Confidence is measured on a seven-point scale. Respondents were asked to indicate how confident they were of their responses ("how sure are you that the choices you indicate reflect your true feelings..."). The end points were labelled as follows: one: not sure at all, seven: very sure.

6. An automobile company might not want to use lists of recent purchasers for the selection of respondents for an automobile preference study. Apart from the fact that the list is not representative of the broader target market, the individuals no longer face the resolution of an automobile choice problem. The lack of an imminent choice problem is expected to reduce the respondents' motivation to provide valid preference judgments.

7. Richard Batsell, personal communication.

Appendix

Reliability

To consider design issues (data collection methods) and their influence on results, suppose that a study involves 4 attributes with 2 levels each. If concept evaluations are desired, a full factorial design consists of 24 = 16 objects, while each of two fractional designs involves 8 objects (see Figure 1).

For tradeoff matrices, a full design consists of 6 matrices (each attribute is used in three matrices), while each of three fractional designs consists of 4 matrices (each attribute is used in two matrices), as shown in Figure 2.

To compare the possible results, as well as the variability in results (unreliability), we assume that a respondent uses lexicographic processing of the objects. Specifically, let a > b > c > d, and let 1 > 2 for each attribute, without loss of generality. We assume also no error in the rank order preferences.

Under these conditions, the relative importances of the attributes obtained under both full and fractional data sets, for both data collection procedures are shown in Table A1. The results, based on least-squares

TABLE A1

Importance Weights for Two Data Collection Procedures, Using Both Full and Fractional Designs

	Profile		Tradeoff	
Attribute	Full	Fraction	Full	Fraction
a	.53	.57	.33	.33
b	.27	.29	.28	(.25-.33)
c	.13	.14	.22	(.17-.25)
d	.07	0	.17	.17
	1.00	1.00	1.00	1.00

regression analysis, show that there is little difference between the numbers obtained for full and factorial designs in the concept evaluation procedure. Also, there is no variability in the results between the two fractions, based on concept evaluations. There is also little difference in results between the full and fractional sets of tradeoff matrices. However, there is some variability in the weights for attributes b and c, between the three fractional sets. Thus, in the absence of human judgment error, and assuming lexicographic processing, concept evaluations are more consistent (reliable) than tradeoff matrices. In addition, the set of relative importances (not the order) is extremely different between the two data collection procedures.

References

Acito, Franklin (1977), "An Investigation of Some Data Collection Issues in Conjoint Measurement," AMA Educators' Conference Proceedings, 82-5.

Akaah, Ishmael P. and Pradeep K. Korgaonkar (1983), "An Empirical Comparison of the Predictive Validity of Self-Explicated, HuberHybrid, Traditional Conjoint, and Hybrid Conjoint Models," Journal of Marketing Research, 20 (May), 187-97.

Campbell, Donald R. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, 56, 81-105.

Cattin, Philippe and Dick R. Wittink (1982), "Commercial Use of Conjoint Analysis: A Survey," Journal of Marketing, 46 (Summer), 44-53.

Cattin, Philippe and Dick R. Wittink (1986), "Commercial Use of Conjoint Analysis: An Update," unpublished working paper, July.

Clarke, Darral G. (1987), Marketing Analysis and Decision Making, Redwood City CA: The Scientific Press (Case: Clark Equipment I, II, and III), 180-211.

Currim, Imran S., Charles B. Weinberg and Dick R. Wittink (1981), "The Design of Subscription Programs for a Performing Arts Series," Journal of Consumer Research, 9 (June), 67-75.

Finkbeiner, Carl T. and P.J. Platz, "Computerized versus Paper and Pencil Conjoint Methods: A Comparison Study," working paper, no date.

Green, Paul E. (1984), "Hybrid Models of Conjoint Analysis: An Expository Review," Journal of Marketing Research, 21 (May), 155-69.

Green, Paul E. (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," Journal of Consumer Research, 1 (September), 61-8.

Green, Paul E., Stephen M. Goldberg, and Mila Montemayor (1981), "A Hybrid Utility Estimation Model for Conjoint Analysis," Journal of Marketing, 45 (Winter), 33-41.

Green, Paul E. and Vithala R. Rao (1971), "Conjoint Measurement for Quantifying Judgmental Data," Journal of Marketing Research, 8 (August), 355-63.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, 5 (September), 103-23.

Hagerty, Michael R. (1985), "Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis," Journal of Marketing Research, 22 (May), 168-84.

Hagerty, Michael R. (1986), "The Cost of Simplifying Preference Models," Marketing Science, 5 (Fall), 298-319.

Jain, Arun K., Franklin Acito, Naresh K. Malhotra, and Vijay Mahajan (1979), "A Comparison of Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models," Journal of Marketing Research, 16 (August), 313-22.

Johnson, Eric J., Robert J. Meyer, and Sanjoy Ghose (1987), "When Choice Models Fail: Compensatory Representations in Correlated Environments," unpublished working paper, January.

Johnson, Richard M. (1974), "Trade-Off Analysis of Consumer Values," Journal of Marketing Research, 11 (May), 121-7.

Johnson, Richard M. (1981), "Problems in Applying Conjoint Analysis," in: Analytic Approaches to Product and Market Planning: The Second Conference, Rajindis K. Srivastava and Allan D. Shocker (eds.), Marketing Science Institute, 154-64.

Krishnamurthi, Lakshman and Dick R. Wittink (1987), "The Influence of Attribute Descriptions on Conjoint Results," unpublished working paper, March.

Leigh, T.W., David B. Mackay, and John O. Summers (1984), "Reliability and Validity of Conjoint analysis and Self-Explicated Weights: A Comparison," Journal of Marketing Research, 21 (November), 456-62.

Malhotra, Naresh K. (1982), "Structural Reliability and Stability of Nonmetric Conjoint Analysis," Journal of Marketing Research, 19 (May), 199-207.

Neslin, Scott A. (1981), "Linking Product Features to Perceptions: Self Stated versus Statistically Revealed Importance Weights," Journal of Marketing Research, 18 (February), 80-6.

Parker, Barnett R. and V. Srinivasan (1976), "A Consumer Preference Approach to the Planning of Rural Primary Health Care Facilities," Operations Research, 24 (September-October), 991-1025.

Reibstein, David, John E.G. Bateson, and William Boulding (1987), "Conjoint Analysis Reliability: Empirical Findings," unpublished working paper, September.

Segal, Madhav N. (1982) "Reliability of Conjoint Analysis: Contrasting Data Collection Procedures," Journal of Marketing Research, 19 (February), 139-43.

Silk, Alvin J. and Glen L. Urban (1978), "Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," Journal of Marketing Research, 15 (May), 171-91.

Scott, J.E. and Peter Wright (1976), "Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations," Journal of Marketing Research, 13 (August), 211-24.

Srinivasan, V. (1987), "A Conjunctive-Compensatory Approach to the Self-Explication of Multiattributed Preferences," unpublished working paper, March.

Srinivasan, V., Arun K. Jain, and Naresh K. Malhotra (1983), "Improving Predictive Power of Conjoint Analysis by Constrained Estimation," Journal of Marketing Research, 20 (November), 433-8.

Wittink, Dick R. and Philippe Cattin (1981), "Alternative Estimation

Methods for Conjoint Analysis: A Monte Carlo Study," Journal of Marketing Research, 18 (February), 101-6.

Wittink, Dick R., Lakshman Krishnamurthi, and Julia B. Nutter (1982), "Comparing Derived Importance Weights Across Attributes," Journal of Consumer Research, 8 (March), 471-4.

Wittink, Dick R. and David B. Montgomery (1979), "Predictive Validity of Trade-Off Analysis for Alternative Segmentation Schemes," AMA Educators' Conference Proceedings, 69-72.

Wittink, Dick R. and John Walsh (1988), "An Interpretation of Conjoint Reliability Findings," unpublished working paper, March.

Wright, Peter and MaryAnn Kriewall (1980), "State of Mind Effects on the Accuracy with which Utility Functions Predict Marketplace Choice," Journal of Marketing Research, 19 (August), 277-93.

CONJOINT PREDICTIONS: 15 YEARS LATER

John A. Fiedler
POPULUS, Inc.

Introduction

In November 1972, a series of papers on conjoint measurement was presented to the annual conference of the Association for Consumer Research. Among these papers was one (Fiedler, 1972) which discussed the application of conjoint measurement to the pricing and design of a pair of condominium towers being constructed on the New Jersey Palisades directly across the Hudson River from Manhattan.

The builder, Centex Homes, had begun selling units in 1971 for occupancy in late 1972.

One building was close to completion, the second would be started soon.

Each of the two 31 story towers comprised a variety of apartments varying in size from three bedrooms - three bathrooms down to one bedroom - one bathroom. The initial pricing reflected the builder's experience in other projects: the larger the unit and the higher the floor, the more expensive the purchase price. Accordingly, units were priced in a range from $35,000 to $78,700 depending on size and floor. After a few weeks of sales, Centex realized that it had made serious mistakes in either the complex's design, its pricing, or both. Table 1 shows the results of these early sales.

Table 1
Sales in Building One (200 Winston Drive): 1/31/72

		Sold	
Unit	Available	#	%
Plan A: 3BR Corner	124	13	10
Plan B: 2BR Corner	122	19	16
Plan C: 2BR Deluxe	122	25	20
Plan D: 2BR Regular	123	12	10
Plan E: 2BR Small	62	19	31
Plan F: 1BR	61	55	90
River View	338	121	36
No River View	276	22	8

Confounding the situation was the fact that the middle range of floors was selling much faster than the lower or upper floor units. Despite adjustments in pricing, the building continued to sell out unevenly.

Survey research in the construction industry in 1972 was as uncommon as it is today. But because the selling experience in the development was so contrary to Centex's experience, Market Facts, Inc. was commissioned to investigate what factors were producing the uneven sellout and what might be done to correct the current situation. Centex also wanted to prevent a comparable occurrence in the developer's second tower, soon to be constructed.

Market Facts 1972 Research Study

Market Facts recommended a study employing conjoint measurement, a technique which, until then, had been employed on a more or less experimental basis for Xerox Corporation. Aside from Market Facts' proprietary experience with Xerox, evidence as to the applicability of conjoint measurement to marketing research had not been substantiated. Centex was understandably skeptical when the method was proposed to them. But a deal was quickly struck: Market Facts would conduct the study for a nominal fee if Centex would permit publication of a paper demonstrating the application of conjoint measurement to market research.

The research was fielded in the summer of 1972. A sample of 188 prospects was interviewed at the model apartments. The survey situation was ideal. Respondents qualified themselves by responding to sales advertisements for the development, and the model apartments' main features yielded comprehensive and distinct questionnaire stimuli. Only four attributes were necessary to describe all the apartments available:

Table 2
Design of Market Facts 1972 Research

Layout - 6 levels: Plans A through F
Price - 10 levels: $46,000 to $82,000
Floor - 4 levels: 28th, 20th, 12th, 4th
View - 2 levels: River view, no view

The measurement task was a simple one. A series of questionnaire grids was constructed, each showing pairings of all possible combinations. Respondents rank ordered their choices from most to least desirable. Given the simple design of the study, all possible pairings of attributes were included.

Findings of the 1972 Research

Neither the report prepared by Market Facts nor the data files remain, but the principal findings are well remembered by this author who has presented the case history many times since 1972. In brief, they are:

- Height premiums were not justified above the lowest tier of floors. The benefits of height are usually a better view and the absence of street noise. In the case of Winston Towers, the quality of the view was not improved as a result of height and there was not a great deal of street noise. On the contrary, higher floors resulted in longer elevator rides and little else.

o Utility values for larger apartments were not proportionate to either purchase price or square footage. The additional rooms and amenities of the larger units were not worth the prices charged.

o View was significantly underpriced. This was no surprise given the sales results. What was surprising was that the spectacular view of Manhattan had, in the survey, an even greater value than might have been deduced from the pattern of sales.

The research recommended that the pricing of the first building be further modified by increasing the premium for the few remaining apartments with a view, by eliminating the floor premiums above the 12th floor, and by lowering the base prices of the larger apartments.

To assure a more even sellout of the second building, a pricing model was developed. Using Monte Carlo simulations, Market Facts was able to produce a pricing schedule which predicted an even sellout in the second building.

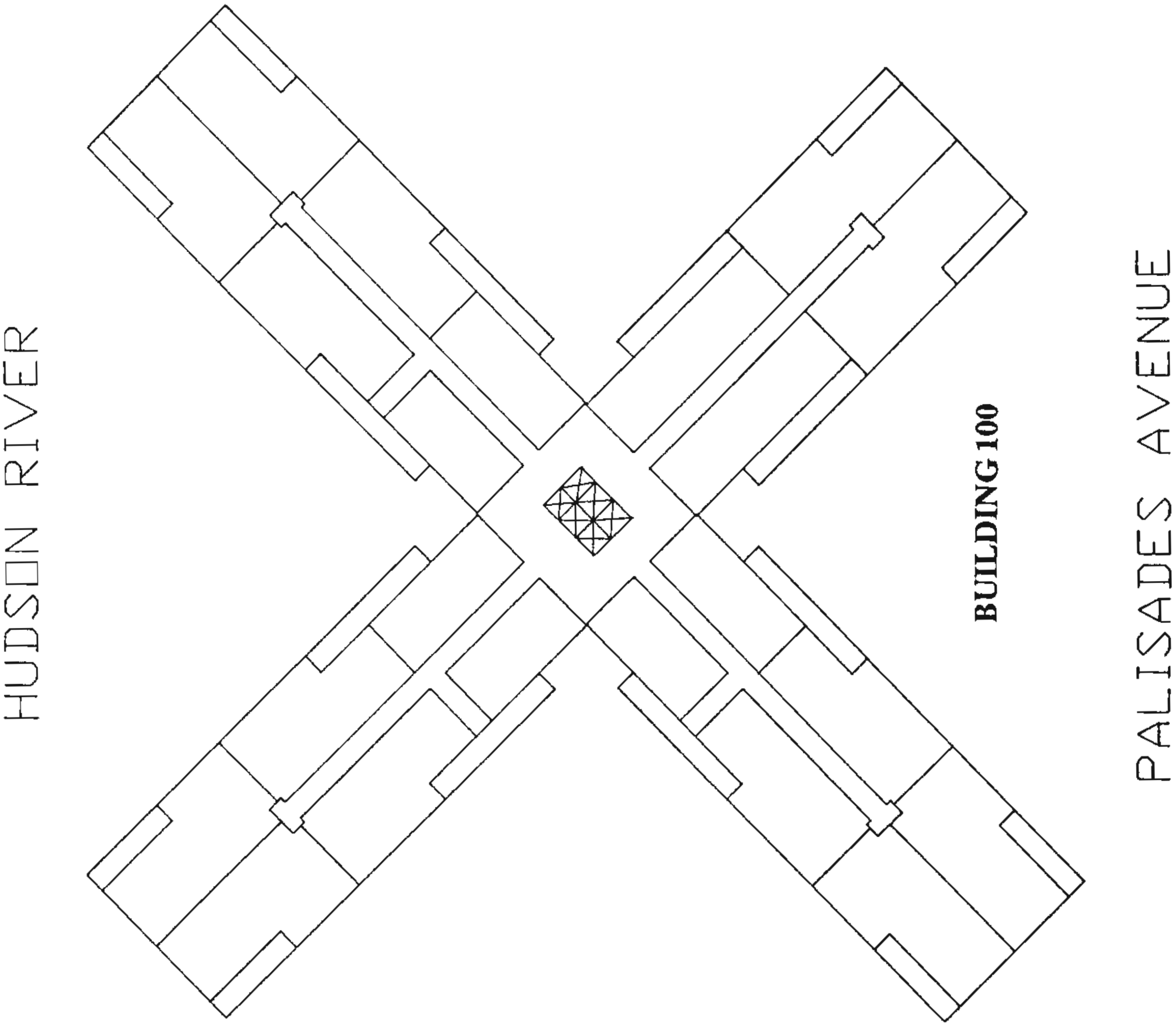
What Happened on the Palisades?

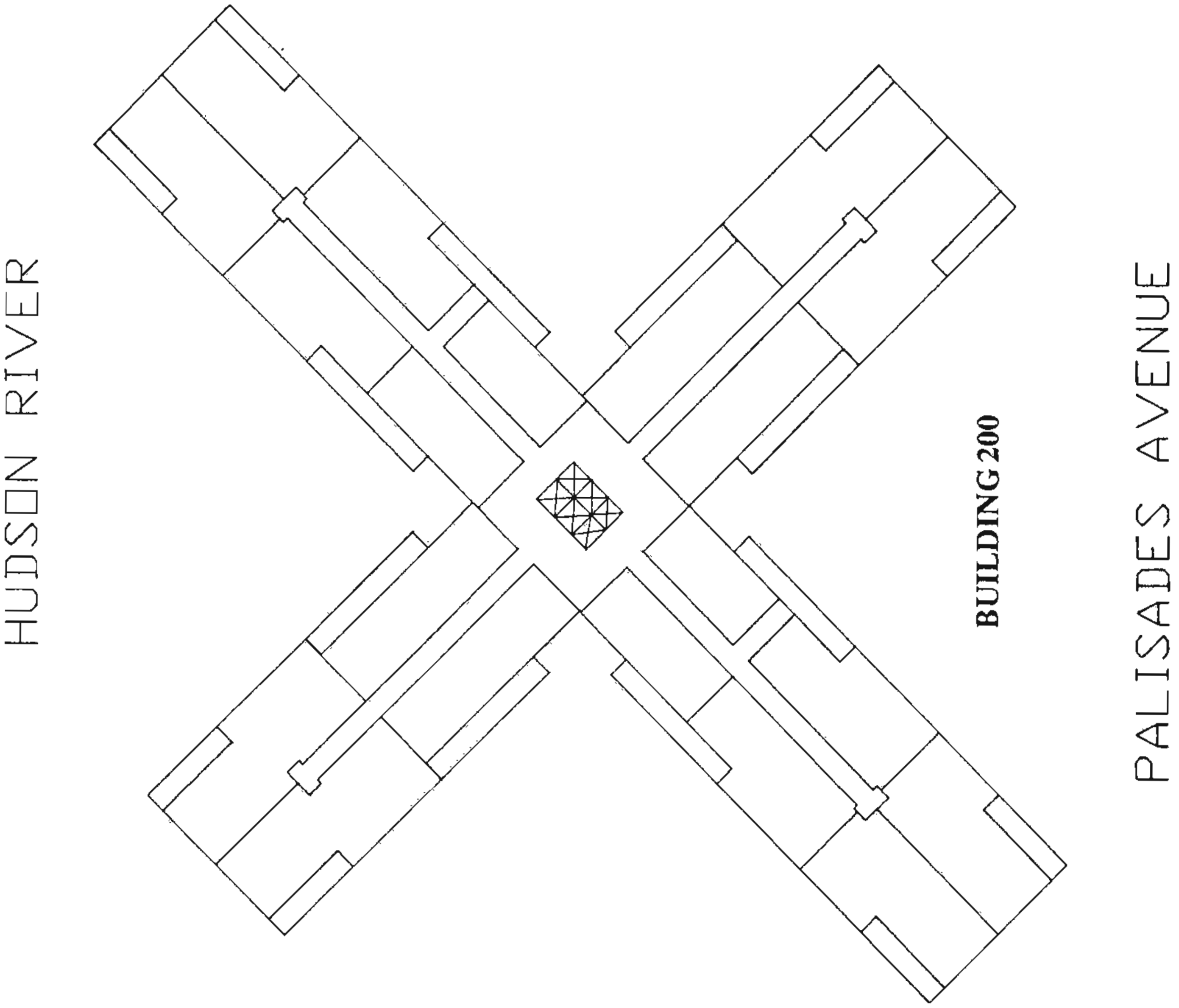
The design of Building Two was modified. While appearing quite similar to the first building, several key changes were incorporated. The east wing of the building was extended approximately 20 feet. Although the resulting design was less symmetrical than the first building, there were two important benefits. First, this design allowed more apartments with a view of the river. Second, the modified design permitted six, rather than four, apartments on each floor of the east wing. The new layout reflected what was selling: more corner and deluxe two bedroom apartments in place of the corner three bedroom apartments, and two more one bedroom units on each floor.

Table 3
Typical Floor Plans: Winston Towers

Unit	Building 1	Building 2
Plan A: 3BR Corner	4	2
Plan B: 2BR Corner	4	6
Plan C: 2BR Deluxe	4	6
Plan D: 2BR Regular	4	3
Plan E: 2BR Small	2	1
Plan F: 1BR	2	4

The floor plan of each building is shown:


HUDSON RIVER
BUILDING 100
PALISADES AVENUE


HUDSON RIVER
BUILDING 200
PALISADES AVENUE

The author recalls that the developer did not charge as substantial a premium for the view as the conjoint pricing model suggested.

While both buildings are fully sold today, it is not known whether or not the second building sold out more evenly than the first. On the basis of Building 2's floor plan alone, it is not unreasonable to surmise that it did.

1988 POPULUS Replication of 1988 Condominium Research

In early 1988, POPULUS undertook a research study to re-examine the Winston Towers condominiums from a conjoint measurement perspective. It was hoped to obtain the cooperation of the tenants' association and the condominium management to permit on site interviewing, but no organization connected with Winston Towers was willing to cooperate with the research effort.

A limited program of research was designed to compare the Sawtooth Software ACA conjoint measurement approach with the earlier method. In addition, tax records for the two buildings were obtained showing the assessed value of each unit as well as the selling prices for 57 units sold from November, 1985 through December, 1986.

There were several key differences between the study conducted in 1972 and the current research:

- o Interviewing Method
    - 1972: Self-administered questionnaires, interviewer supervised in the sales office;
    - 1988: Computer assisted telephone interviewing via WATS;
- o Conjoint Measurement Model
    - 1972: Market Facts proprietary application of Johnson's nonmetric factor analysis;
    - 1988: Sawtooth Software Adaptive Conjoint Analysis (ACA);
- o Sample
    - 1972: 188 Prospective buyers;
    - 1988: 100 Owner residents.

Questionnaire Design

Attribute descriptions for unit types, floor level, and view were identical in both studies. Because of inflation, prices had to be adjusted. In the 1972 study, there were 10 price levels, in $3,000 increments, ranging from $46,000 to $82,000. In the 1988 study, there were 8 price levels, in $25,000 increments, ranging from $150,000 to $325,000. In the interview, the number of price levels any individual respondent saw was reduced to five using a "most likely" question. Three calibrating questions were utilized. After the ACA portion of the interview, respondents were asked to identify the unit number of their apartment, the building in which they lived, the year they purchased their apartment, and the price paid. Interviewing was conducted from March 4 to March 18, 1988.

Findings from 1988 Research

One of the most striking findings of the current study is the ease with which a conjoint measurement study can be fielded and the data analyzed. A process which took months fifteen years ago can be accomplished in weeks; a research procedure that was extremely costly then is extremely cost effective today. The respondent task is far easier and the computational procedures more effective.

Conjoint measurement permits a "goodness of fit" measure to be computed for each respondent's utility values. In 1972, the Market Facts procedure was to compare, on a pairwise basis, the rank orders of computed utilities with the respondent's raw data, using Kendall's tau. Sawtooth ACA reports the correlation between a respondent's purchase likelihoods for a series of concepts and the combined utilities for those concepts. While neither the procedures nor the measures are directly comparable, it is interesting to compare the results.

Table 4
Comparison of "Goodness of Fit" Measures

	Tau (1972)	Correlation (1988)
(Base)	(188)	(100)
	%	%
1.000	13	0
.950 - .999	27	64
.900 - .949	29	12
.850 - .899	14	9
.800 - .849	9	7
< .800	6	8

Mean utility values from the 1988 study are parallel to the author's remembered findings from the 1972 research. Again, units on the lowest range of floors are very undesirable. The value attached to a Hudson River view is very high. The relatively high desirability of the one bedroom apartments (Plan F) which were the first to sell out in 1972 is reflected in the 1988 data: these units, on average, are more highly valued than the small two bedroom units (Plan E).

Table 5
Mean Utility Values: 1988 Condominuium Research

Attribute/Level	Utility
Plan A: 3BR Corner	+.52
Plan B: 2BR Corner	+.10
Plan C: 2BR Deluxe	0.00
Plan D: 2BR Regular	-.37
Plan E: 2BR Small	-.60
Plan F: 1BR	-.49
River View	+.54
No River View	-.58
28th Floor	+.16
20th Floor	+.20
12th Floor	+.07
4th Floor	-.53
$325,000	-.65
$300,000	-.57
$275,000	-.42
$250,000	-.19
$225,000	+.10
$200,000	+.32
$175,000	+.51
$150,000	+.60

Validation

The 1988 research provided an opportunity to validate the application of conjoint measurement to condominium pricing by comparing the utility values owners ascribe to features to the dollar value the market place ascribes to these same features.

To provide data for this comparison, it was first necessary to examine the Bergen County tax records for recent sales of Winston Towers units. From November, 1986 through December, 1987, 57 units were sold. From the unit and building numbers corresponding to each sale, it was possible to determine the view, floor level, and plan for each unit sold. The sales prices were decomposed to determine the part-worth market values of each level of each attribute. This was done through a series of multiple regression analyses using dummy variables reflecting each level (less one) of each attribute.

The utility data were then rescaled to a dollar metric by regressing the mean utility values for each level of plan, floor, and price against the corresponding part-worth sales data. The function resulting from this regression ($UTIL = 64676.17*UTIL + 2664.89) was then applied to the utility values.

To make the resulting data more easily comparable, the value of the least desirable level of each attribute was set to zero and the remaining levels adjusted accordingly.

Table 6
Comparison of Rescaled Utility Values
with Part-worth Sales Data
($000)

Attribute/Level	Rescaled Utilities	Part-worth Sales Data
28th Floor	45	22
20th Floor	47	15
12th Floor	39	20
4th Floor	0	0
Plan A: 3BR Corner	65	134
Plan B: 2BR Corner	38	93
Plan C: 2BR Deluxe	32	64
Plan D: 2BR Regular	8	30
Plan E: 2BR Small	7	19
Plan F: 1BR	0	0
River View	72	16
No River View	0	0
$325,000	69	
$300,000	74	
$275,000	84	
$250,000	99	
$225,000	117	
$200,000	132	
$175,000	144	
$150,000	150	

At first glance, this "best fit" rescaled comparison does not appear to be much of a fit at all. The conjoint model of rescaled utilities has overestimated differences in attribute levels of view and floor and underestimated the effects of plan and price.

The model most accurately reflects the differences in floor level. In each case the 4th floor has considerably less value than any other floor. The range of differences across the top three tiers of floors is $8,000 for the rescaled utilities and $7,000 for the market values. Most of the distortion may be due to survey respondents "over rejecting" the lowest tier of floors.

Comparing the findings for the different plans reveals that the conjoint model consistently underestimates the value of each type of unit. Conjoint utilities fail to discriminate between the corner and deluxe two bedroom apartments and between the regular and the small two bedroom apartments. Further, all conjoint utilities underestimate the value of each unit.

The conjoint model overestimates the value of a river view by a factor of four. This is the greatest error in prediction. One hypothesis is that those residents who have river view apartments place a far greater value on view than do those who do not have such a view. Table 7 shows the utilities of view for each group. This psychological rationalization seems appropriate in light of their purchase choices.

Table 7
Utilities for View By Resident Apartment

	Apartment	
Attribute/Level	River View	No View
River View	.61	.37
No River View	-.64	-.44

While those living in river view apartments do have greater utilities for that view, the conjoint model would overestimate the market value for view even if the predictions were based on only those living in non-view apartments.

The last factor to be investigated was selling price. The utility values for purchase price are monotonically inversely related to price as one would expect. The rescaled dollar utility values may be thought of as the amount of money someone would pay to avoid spending a certain amount for a condominium. If the model were perfect, the dollar utilities would decline dollar-for-dollar as purchase price rises; they do not. The value of money spent is substantially underestimated.

Conclusions and Implications

It is too easy to review these analyses and conclude that the conjoint model is a weak predictor. While the model's errors are substantial, they are also understandable.

They appear to replicate the findings of the 1972 research. Centex was justified in its decision to implement price premiums significantly less "steep" than those suggested by the Market Facts model.

More importantly, the conjoint model overestimates the effects of those attributes which may be more emotionally laden, such as the benefits of a glorious view or the consequences of living on the lowest tier of floors. Correspondingly, the model underpredicts the more concrete attributes such as price and floor plan.

These methodological findings suggest that in the design of conjoint measurement research, there is great risk in attempting to measure across attributes, some of which are concrete descriptors and others of which are more benefit oriented. (See Reynolds, Fiedler & Gutman, 1984). It further suggests that there is still substantial work to be done to fine tune calibration procedures.

J.A. Fiedler, "Condominium Pricing and Design: A Case Study in Consumer Trade-off Analysis," Proceedings of the ACR (1972), 279-293.

T.J. Reynolds, J.A. Fiedler, J. Gutman, "Understanding Consumers' Cognitive Structures: The Relationship of Levels of Abstraction to Judgments of Psychological Distance and Preference," Psychological Processes and Advertising Effects (1985), 261-272.

NOTE: The author wishes to acknowledge the assistance of Richard Miller of Consumer Pulse in conducting the field work for this research and Sheri Nadel of Marketing Perceptions for obtaining floor plans and tax and sales records for the Winston Towers condominiums.

ATTRIBUTE SELECTION AND REPRESENTATION IN CONJOINT ANALYSIS: RELIABILITY AND VALIDITY ISSUES

Douglas L. MacLachlan
University of Washington

Michael G. Mulhern
Mulhern & Associates

Allan D. Shocker
University of Washington

INTRODUCTION

In order to examine the reliability/validity issues of attribute selection, we must recognize what conjoint analysis is all about. There are essentially three broad goals or reasons for doing conjoint analysis. It is done: (1) as a way of helping management think about how customers make buying decisions, (2) as a way of obtaining substantive measures of the value to customers of the factors that enter into those decisions, and (3) as a way of predicting the outcomes (i.e., preferences or choices) implied by those decisions. The first two of these reasons are critically concerned with the issue of what attributes and attribute levels are selected to describe the product or service under study. Wrong choices of attributes by the researcher will mean incorrect understanding of customer decision-making and, most certainly, invalid measurement of the values customers attach to the decision factors. Although the third reason for doing conjoint analysis provides a way of assessing the correctness of the researcher's attribute selection process, it does not bear directly on that process.

Conjoint analysis provides a way of attempting to get inside the customer's mind. As we can never really do this, we must rely on external reflections of that internal reality. We must get people to tell us what we are unable to observe directly. The issue of validity refers to whether or not what we observe (by listening to what they tell us) is actually what goes on in the heads of customers. Reliability, on the other hand, involves the consistency with which we observe on each observation occasion. In fact, we may observe something with great precision, hence have great reliability in our measurement process, yet be observing the wrong thing, and therefore be using invalid measures.

The conjoint design and measurement process can be conceptualized in four stages. Initially, the researcher must make some decisions regarding attributes of the product-concepts under investigation. Specifically, the investigator must identify, reduce, select and, in some fashion, represent a relevant and meaningful attribute set to the respondent in the form of attribute bundles. This, in turn, determines the complexity and "realism" of the choice or evaluation task the respondent must perform. Upon obtaining data from respondents, the researcher derives attribute importances (from attribute-level part-worths) and simulates product-concept choices for each respondent and/or market segment of interest. Finally, the researcher assesses the validity and reliability of the attribute importances and predicted choices.

To the extent that respondent judgments are made among "described" product-concepts, the validity of the data collected (and thus the researcher's inferences) will depend, at a minimum, on the selection and representation of the attributes used to describe the products. If the researcher omits relevant attributes, misspecifies the appropriate attribute ranges, or presents attribute-level combinations in a confusing or incomprehensible manner, respondent judgments are not likely to reflect marketplace reality.

Regarding the attribute representation issue, one must realize that validity of the conjoint measurement task is a matter of communication between researcher and respondent. The researcher (sender) provides stimuli (encoded message) to respondents (receivers) with a specific understanding as to what the stimuli mean. The respondent, on the other hand, observes (or decodes) the stimuli with all the limitations inherent in human beings' cognitive processing capabilities. That is, respondents have limited capacity for handling complexity and they may selectively attend to and/or distort the information that is presented to them (Nisbett and Ross, 1980). No matter how accurately the researcher selects the attributes that are important in choice among stimuli, the ways they are <u>presented</u> to the respondent will influence the validity of their responses.

It seems logical to assume that the less "real" the stimuli presented to respondents, the less valid will be their responses. If respondents are allowed to feel, smell, taste and otherwise experience the product-concepts they are asked to compare, one would expect their judgments to be more predictively valid than if the stimuli were represented in more artificial ways. In most conjoint tasks, researchers do not have the luxury of presenting respondents with real products. The question then becomes: How much validity is lost by presenting the product-concepts verbally, pictorially, or in other artificial ways?

Obviously, conjoint analysis is not perfect; nor is it a panacea for the problems of product designers or marketers. But when used carefully and properly the method has substantial capabilities, especially when judged in comparison with other feasible approaches to providing managers with customer (i.e., market) perspectives. It is our purpose in this paper to alert researchers to the problems that arise and solutions to consider when selecting and representing bundled attributes (i.e., stimuli) for a conjoint analysis study.

ATTRIBUTE IDENTIFICATION, REDUCTION, AND SELECTION: WHICH ATTRIBUTES TO INCLUDE?

A product-concept is typically characterized as a bundle of attributes (encompassing physical properties, benefits, costs, and other characteristics to be discussed at length later). It is possible to uncover a substantial number of these as one begins preparation for the conjoint task. In fact, the result may be a list of attributes that contains too many for the average respondent to react to and they may be of the wrong character. It is incumbent on the researcher to generate the "right" set and limit its size.

Issues that pertain to attribute inclusion can be viewed from either a conceptual or measurement perspective. Conceptually, attribute choices for conjoint analysis follow from customers' product-concept choices within the market context. Although this simple truth may seem obvious, there are reasons that it is not attended to in practice. Company managements often think of their products in terms of their objective properties, the characteristics that need to be modified to change the products. However, customers may not perceive the products in terms of such properties. When a woman buys a lipstick it is unlikely that she is concerned about its chemical composition, but rather likely that she is interested in how it will make her look when wearing a particular color of dress. Similarly, a man who buys a package of cigarettes may have no direct interest in the amount of tar in the smoke, but be influenced by his perception of the type of person who smokes the particular brand. Customers might attempt to use a different set of attributes when making marketplace decisions than management considers when making product/service design decisions. Furthermore, the essential character of the attributes might be different.

Characterizing Attributes

We have noted that the nature of attributes depends fundamentally on the product-concept under investigation. However, the researcher often must make decisions regarding the character of the attributes chosen. These decisions have implications for the reliability and validity of conjoint results.

With respect to the characterization of attributes, the Myers-Shocker (1981) typology of attributes provides a helpful classification. In their terms, attributes may be classified as physical or pseudo-physical properties, benefits, or imagery attributes. Physical properties (which take a product referent) are such attributes as temperature, weight, height, chemical composition, or price in dollars. Pseudo-physical properties (also taking a product referent) are attributes such as fragrance, spiciness, or style. Benefits or negative benefits, i.e., costs (which are task or outcome referent) are attributes such as durability, handling, performance, or comfort (negative benefit examples are tendency to promote cancer, inconvenience, or inaccessibility). Finally, imagery attributes (user referent) are exemplified by association with relevant others such as movie actors or athletes (e.g., this type of shoes is worn by Michael Jordan).

Identifying Attributes

A place to begin the quest for customer-based attribute selections is to observe the products that customers "naturally" consider as alternatives. If the product under consideration is a symphony performance, what other leisure-time activities are viewed as competing with it? Such product choices can, in turn, be affected by factors such as the knowledge and awareness of those making that choice, their goals and purposes, the product alternatives considered available, and even the research task and purposes. Let us examine these ideas in greater detail.

To understand the complexity of the researcher's attribute selection problem, it is useful to consider the concept of an affordance (Gibson,1977). This may be viewed as a potential benefit of a product-concept. For example, a book may be evaluated in terms of its content, color, weight, and size for some purposes and on the basis of the size of its print or the level of readability for others. The set of attributes that are relevant depend both on the person making the judgments and the purpose for which the choice is being made. If the person doing the evaluation has poor eyesight, the size of the print becomes relevant. If one wishes to use the book for a doorstop, its weight becomes relevant. These issues simply suggest that the researcher must try, to the extent possible, to imagine the ways product-concept choices will be evaluated by prospective users in the marketplace. If the person-situations can be anticipated, this will indicate both how to constrain the conjoint decision task and how to select respondents (i.e., how to define the target population for the study). If several types of persons are deemed important for the researcher's purpose, separate conjoint tasks may need to be designed. Furthermore, it will typically be necessary for the researcher to specify clearly to the respondent the type of usage situation to be considered while making stimuli evaluations.

Reducing the Attribute Set

The researcher must identify what have been called determinant attributes (Alpert, 1971). These are attributes that relate to preference and choice and distinguish the choice alternatives in meaningful ways. It is not enough that a given attribute is deemed by customers to be "important" in the sense of having intrinsic value; the relevant set of alternative product-concepts must also differ significantly on that attribute. Color was not a determinant attribute for automobiles when the only color available was black. It still might not be a determinant attribute for automobile choice except for some small subset of the population. But now that cars are available in many colors, color has the potential of becoming a determinant attribute.

Attributes must be meaningful to respondents. As illustrated by the lipstick and cigarette examples before, attributes that are meaningful to management are not always most meaningful to respondents. Furthermore, ranges of attribute levels and attribute-level combinations must be both plausible and complete. Although it may be appropriate in conjoint studies to introduce attributes and attribute levels that respondents have had no experience with in order to measure their reactions to potential product-concepts, such attributes or attribute levels must be viewed as being within the realm of possibility by respondents. A luxury car described as being offered at a subcompact price would not be taken seriously; nor would any car be taken seriously if it were described as being available at a $100 price.

Regarding completeness of the attribute set, the stimuli descriptions must convey all the information that respondents feel they need to make their decisions. If the set of attributes is not considered complete, there is the danger that either the task will lack credibility or respondents will assume missing attribute information to be correlated with certain included attributes. An example of the latter case is where the respondent who is not given information about certain quality features of a table wine simply infers such features on the basis of price. In either case, the researcher can be misled by the responses.

One solution to the problem of reducing the set of attributes to a short list is to consider in the conjoint task only those attributes that are both alterable by management and meaningful to customers. Such actionable attributes may be the consequence of management's own initiative or arise in response to competitors' actions. Keeping in mind the caveats above, this may be a way to reduce the attributes employed in the conjoint task to a relatively small set. The task can be made meaningful by placing it in a well-defined usage context and selecting respondents from a properly designated target population. If an electronic instrument manufacturer is considering producing a product with much higher resolution than those currently available on the market, the researcher investigating the value of that attribute should include it as part of the stimulus descriptions. But care must be taken to describe the situations where a high degree of resolution would be useful and to select respondents from among those who would find themselves in such situations.

Measurement Issues

With respect to measurement issues in attribute selection, published studies (e.g., those listed in Green and Srinivasan's (1987) bibliography) have used as few as three attributes and up to forty or more. An example of the latter is a study of the pricing of hotel amenities by Goldberg, Green and Wind (1984). The so-called QFD or "quality function deployment" studies (cf. Hauser and Clausing, 1988), while not technically conjoint-analytic studies, employ thousands of attributes to describe product-concepts (e.g., up to 100 reportedly just to describe subcomponents of automobiles such as a car door). Conceptually, at least, the number of attributes that could conceivably describe virtually any product-concept is almost infinite. There are, of course, practical limits to the numbers that can or ought to be used in actual studies.

The basic research compromise is between "realistic" description of the product concepts on the one hand and respondent task complexity on the other. Both sides of the problem will affect the validity and reliability of conjoint task results. The lower the realism of the task, the less likely it is that a valid evaluation of the product-concepts will be made. Also, because of respondents' limited cognitive capacity, limited attention span, and likelihood of fatigue and boredom, the more complex the task, the less likely it is also that the results will be valid (at least beyond some threshold).

An important subissue is the number of attribute levels. Again a basic compromise lies between realism and task complexity. Some guidance is afforded by the customer's familiarity with the product-concept. The greater the familiarity, the better able the respondent is to make fine distinctions and hence the greater number of categories that are appropriate (Park and Lessig, 1981). Reliability will be enhanced the fewer levels that are used (because respondents will be less likely to misclassify product-concepts into wrong ordinal categories). A case can be made, however, that the lesser realism associated with fewer levels could result in lower validity.

Generally speaking, attributes should be selected that have low correlation across familiar product-concepts. Not only will there be a degree of redundancy if attributes are correlated, but it will be difficult to untangle the respondents' evaluations of the correlated attributes. This is particularly a problem with the price attribute, since price is often associated in many people's minds with levels of quality. Yet the attribute combinations should be plausible, and respondents will suspend belief if they are asked to evaluate product-concepts that seem unreasonable or unlikely to occur in the marketplace.

In summary, the best advice for the researcher who wishes to obtain valid conjoint measures is to try to imagine the most realistic tasks possible for the respondent. Having in mind a usage context for the product-concept will help assure that the right attributes are selected and the appropriate respondents are asked to respond to the conjoint task.

ATTRIBUTE REPRESENTATION: HOW SHOULD ATTRIBUTES BE REPRESENTED?

As described before, it is essential to present attributes and attribute levels to respondents in terms they can understand. All conjoint tasks require judgments about product-concepts that are abstract representations of real marketplace offerings. Therefore, the farther the task is removed from the actual choice situation, the less valid is the task.

The stimuli typically provided respondents in conjoint tasks are abstractions of potential or actual product-concepts. Much as models simplify a complex reality in order to concentrate attention on the important features of phenomena under study, conjoint stimuli offer respondents choices uncluttered by details that seem unimportant for their evaluations. On the other hand, stimuli must be realistic enough to be meaningful to the respondents. A delicate balance must be struck between drowning respondents in detail and describing the product-concepts at a meaninglessly abstract level.

Attribute Representational Forms

One can visualize a spectrum of representational forms, ranging over the following levels: (1) actual products that people can touch, taste, or otherwise directly experience (e.g., Fiedler (1972) described a conjoint task following a tour of model condominiums); (2) mock-ups or prototypes of products (e.g., General Motors and Boeing use physical models of their products that people can sit in and physically inspect before responding to conjoint questionnaires); (3) movies, cartoons, or CAD with changing perspectives; (4) photographs (e.g., Louviere, et al., 1987) or recorded sounds; (5) drawings, sketches, or diagrams (e.g., Domzal and Unger, 1984); (6) graphs or scale positions; and (7) written words or numerical values. Additionally, the last several forms can be presented on paper or cards for sorting (or other evaluations) or, as is becoming increasingly popular, displayed on video screens such as CRTs. It is only a matter of time, if it is not being done already, before video disk technology is coupled with computers to provide increasingly realistic product-concept descriptions (possibly with motion) for respondent evaluation.

Measurement Properties of Attribute Labels

Reliability and validity of the conjoint task will also depend on the measurement properties of the attribute labels chosen. Some of the relevant distinctions uncovered in a review of conjoint studies are (1) features versus dimensions; (2) numerically-valued versus verbally-described attributes; (3) concrete versus abstract attributes; (4) the type of anchoring used, if any; and (5) briefly- versus extensively-described attribute levels. These are neither mutually exclusive nor exhaustive categories of attribute label choices, but do provide some means of considering the many decisions evidenced by the published literature.

Features versus Dimensions. This distinction is described in detail by Garner (1978). Features are characteristics that are simply present or absent in a product concept, such as an automobile's sunroof, whereas dimensions are characteristics that may vary over a continuous scale range, such as weight or age. The latter might be considered dimensions even if they are described in dichotomous form such as heavy-light or young-old.

Numerically Valued versus Verbally Described. Attribute-level labels such as 5 lbs., 20%, or $10 are examples of the former type, whereas low, heavy, or bright are examples of the latter type.

Concrete versus Abstract. This distinction is addressed by Johnson and Fornell (1987). To some extent the distinction is related to the numerical-valued versus verbally-described attributes. However, it involves the contrast between relatively specific attributes (e.g., TV screen size) and more general ones (e.g., entertainment). The latter, superordinate attributes, are necessarily described in more abstract terms.

Type of Anchoring. Attributes may be globally or objectively anchored in such a way that everyone will use the same reference points for evaluation of the levels (e.g., $10, zero probability, 9 a.m., or 25%). Alternatively, the attributes might be respondent anchored in subjective terms (e.g., lower than average, 5% higher than usual). Or, they might be unanchored (e.g., low, moderate, or heavy).

Brief versus Extensive Descriptions. In many cases the attribute levels can be listed in abbreviated fashion, in others greater detail is provided. The latter, while obviously providing greater realism to the task, also requires greater cognitive processing by the respondent. In one of the few studies to address this issue, Armstrong and Overton (1971) found no significant differences in respondents' intentions to purchase given brief versus comprehensive product descriptions.

An issue of interest is the impact of mixing together combinations of attributes having different character and measurement properties. For example, do people attend more to attributes that are concrete physical properties rather than to ones that are abstract benefits? If so, this would affect the validity and reliability of the derived part-worths. No research seems to have been directed to this issue.

VALIDITY AND RELIABILITY ENHANCEMENT: SUGGESTIONS FOR IMPROVEMENT OF CONJOINT STUDY QUALITY

On the basis of our experience and a review of published conjoint studies, we offer the following recommendations to minimize the design and measurement problems previously discussed.

Attribute Identification

Initial identification of attributes should be done broadly to assure that all important attributes are considered (thereby helping to increase the validity of the exercise). Common methods mentioned in the published studies include in-depth interviews with industry experts and representative customers as well as exhaustive search through all relevant published and unpublished secondary sources. The concept of affordances may be used to suggest attributes that emerge as important in different usage scenarios or for different types of respondents. Attributes uncovered through such methods are typically augmented with the researcher's and manager's judgment.

Attribute Reduction and Selection

Reduction of the attribute set is typically done in a variety of formal and informal ways. Focus groups and in-depth interviews are again useful, as are the researcher's own judgments. It is necessary to be cautious, however, in that directly assessed attribute importances may not correspond to those that will turn out important in actual conjoint tasks. The former may involve stereotypic judgments that are not present in the customer's actual marketplace choice situation. As indicated earlier, the set of attributes can be reduced by employing the notions of determinant attributes (customer perspective) and actionable attributes (manager perspective). The attribute set can be narrowed by constraining the situational context of the choice task and/or the characteristics of the customer presumed to make the choice. Also, the problem may be simplified by recognizing the hierarchical nature of a customer's decision problem. First, the analyst can select an evoked or choice set on the basis of a few attributes (noncompensatory rules), then make the final decisions on the basis of a relatively few additional attributes (compensatory rules). Finally, one might collect judgments from a pretest sample of respondents using a large set of attributes, then reduce the data set via factor analysis in order to select a representative set of relatively uncorrelated attributes for the final study.

There are, of course, various ways to reduce the task complexity for any given number of attributes and attribute levels. Examples include fractional factorials and blocking experimental designs (Green,1974), sampling of stimuli (Kienast, et al., 1983), hybrid designs (Green,1984), and adaptive methods (Johnson,1987). However, most researchers try to reduce the number of attributes to an "essential" set of relatively uncorrelated attributes that both completely describe the product-concepts in terms of customer choice criteria and include the "actionable" attributes under managerial control.

Attribute Representation

For some types of product-concepts, "complete realism" is desirable. In such cases, actual products or prototypes are probably best used. Examples might be where evaluations are to be made of the seat-comfort or handling characteristics of an automobile. There may simply be no adequate words to describe what must be evaluated.

In other cases, pictures might be worth the proverbial "thousand words". An example might be places of scenic beauty for travel destinations (cf. Louviere, et al., 1987). Mere verbal descriptions would not do justice or are likely to be misinterpreted by respondents and it would be too costly to have respondents actually experience the alternatives. For some types of product-concepts, sketches or diagrams might be appropriate representations. For example, automobile stylings might be well represented by CAD (computer-aided design) drawings which, with appropriate software and display might enable respondents to visualize different perspectives and colors of design.

In the majority of conjoint studies to date, conjoint stimuli have been composed of verbal descriptions of attribute-level combinations written on paper, cards, or CRT screens. Cattin and Wittink (1986) determined that 70 percent of their sample of commercial uses of conjoint analysis employed verbal or paragraph descriptions of the stimuli. The validity issue becomes, in such studies, whether the verbal descriptions capture enough of the reality of the marketplace to make the conjoint tradeoffs meaningful. It seems reasonable to suggest some rules of thumb. To the extent possible, for example, verbal descriptions should use words and semantics intelligible to customers in the target market. Do not assume that attribute descriptors that are well-understood by the researcher or managers will be understood the same way by the respondent. As in survey research, it is important to pretest product-concept "descriptions" in order to ascertain their meanings to the target population. The stimuli themselves or the conjoint task explanation should convey non-product information (e.g., decision purpose or task, purchase or consumption role) if such is useful to aid respondent understanding. For example, articles of apparel will be evaluated differently if they are to be worn at work than if worn in other contexts. Similarly, the apparel might have different meanings for respondents if told that they are to be worn by people in high versus low status occupations. Also, the researcher must include for evaluation, or keep constant in the conjoint task, certain customer-relevant costs (e.g., prices or inconvenience associated with the product-concepts). To omit them would invite the loss of task credibility or the unintended addition of respondent-initiated associations discussed earlier.

In many cases it will turn out that customers derive more meaning out of benefits and costs than physical characteristics of product-concepts. When this is true, management and R&D people may have difficulty translating respondent evaluations into appropriate levels of actionable physical attributes. One reason is that benefits, such as riding comfort, might be comprised of or summarize a number of physical characteristics such as shock-absorber rating, weight of car, noise levels, and firmness and shape of the car seats. If physical characteristics are employed in the conjoint tasks, respondents may make inferences about the benefits they imply that are unknown or misunderstood by the researcher. For example, a cereal's fiber content might suggest to some respondents something about its healthfulness and to others something about its crunchiness.

Whatever decisions are made by researchers about how to represent stimuli to respondents, the fundamental question is how well respondents understand the comparisons they are asked to make.

CONCLUSION

Conjoint analysis is replete with problems for the researcher, not the least of which are the selection and representation of attributes for respondent evaluation. We have described the key issues involved and have suggested guidelines for resolving them. Some research has been done to investigate the reliability and validity of attribute selection and representation choices (Bateson, Reibstein, and Boulding, 1988). However, much remains to be done.

Our primary advice to the researcher is to project oneself into the role of the customer and determine how he or she would react to attribute and attribute-combination descriptions in a real buying or choice situation.

REFERENCES

Alpert, Mark L. (1971) "Identification of Determinant Attributes: A Comparison of Methods," Journal of Marketing Research, Vol. VII (May), 184-91.

Armstrong, J. Scott, and Terry Overton (1971) "Brief vs. Comprehensive Descriptions in Measuring Intentions to Purchase," Journal of Marketing Research, Vol. VIII (February), 114-7.

Bateson, John E. G., David Reibstein, and William Boulding (1988) "Conjoint Analysis Reliability and Validity: A Framework for Future Research," Review of Marketing, JAI Press, forthcoming.

Cattin, Philippe, and Dick R. Wittink (1986), "Commercial Use of Conjoint Analysis," Working Paper (July).

Domzal, Teresa J., and Lynette S. Unger (1984) "Judgments of Verbal Versus Pictorial Presentations of a Product with Functional and Aesthetic Features," Advances in Consumer Research, Vol. 12, eds. E. Hirschman and M. Holbrook, 268-72.

Fiedler, John A., (1972) "Condominium Design and Pricing: A Case Study in Consumer Trade-off Analysis," Proceedings of the Association for Consumer Research Annual Conference, 279-93.

Garner, W. (1978), "Aspects of a Stimulus: Features, Dimensions, and Configurations," in Cognition and Categorization,eds. E. Rosch and B. Lloyd, New York: Wiley.

Gibson, J. J. (1977), "The Theory of Affordances" in R. Shaw and J. D. Bransford, eds. Perceiving, Acting, and Knowing, Hillsdale, IL: Erlbaum.

Goldberg, Stephen M., Paul E. Green, and Yoram Wind (1984) "Conjoint Analysis of Price Premiums for Hotel Amenities," Journal of Business, Vol. 57, S111-32.

Green, Paul E. (1974) "On the Design of Choice Experiments Involving Multifactor Alternatives," Journal of Consumer Research, 1 (September), 61-8.

________ (1984) "Hybrid Models for Conjoint Analysis: An Expository Review," Journal of Marketing Research, 21 (May), 155-9.

________,and V. Srinivasan (1987) "A Bibliography on Conjoint Analysis and Related Methodology in Marketing Research," Working Paper, (July).

Hauser, John R., and Don Clausing (1988) "Quality Function Deployment: Competitive Advantage by Listening to the Voice of the Customer," Working Paper 88-035, Harvard Business School (February).

Johnson, Michael D., and Claes Fornell (1987), "The Nature and Methodological Implications of the Cognitive Representation of Products," Journal of Consumer Research, Vol. 14 (September), 214-28.

Johnson, Richard M., (1987) "Adaptive Conjoint Analysis," Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, Ketchum, ID: Sawtooth Software.

Kienast, Philip, Douglas L. MacLachlan, Leigh McAlister, and David Sampson (1983), "Employing Conjoint Analysis in Making Compensation Decisions," Personnel Psychology, Vol. 36, 301-13.

Louviere, Jordan J., Herb Schroeder, Cathy H. Louviere, and George G. Woodworth (1987), "Do the Parameters of Choice Models Depend on Differences in Stimulus Presentation: Visual Versus Verbal Presentation?", in M. Wallendorf and P. Anderson, eds., Advances in Consumer Research, Vol. 14, Provo, UT: Association for Consumer Research, 79-82.

Myers, James H., and Allan D. Shocker (1981) "The Nature of Product-Related Attributes," Research in Marketing, Vol. 5, JAI Press, 211-36.

Nisbett, Richard, and Lee Ross (1980), Human Inference: Strategies and Shortcomings of Social Judgment, Englewood Cliffs, NJ: Prentice-Hall.

Park, C. Whan, and V. Parker Lessig (1981), "Familiarity and Its Impact on Consumer Biases and Heuristics," Journal of Consumer Research, Vol. 8 (September), 223-30.

COMPARISON OF CONJOINT METHODS

Manoj K. Agarwal
State University of New York

Introduction

In this paper we will compare the common methods of doing conjoint analysis. These include full profile, paired comparison, Adaptive Conjoint Analysis (ACA), etc.

The following issues are of concern when conducting a typical conjoint study:

A. Design of the study

1. number of attributes and their selection
2. the number of levels and their description for each attribute
3. the attribute combinations to use
4. the form of presentation
5. the form of judgments to be obtained
6. the analysis technique

B. Field Administration

C. Estimation of the utilities or partworths

D. Validity and Reliability of the results

E. Simulations to project market shares etc.

Rather than discussing each of the steps in detail, this paper discusses the differences between the traditional methods of doing conjoint with paper-and-pencil data collection procedures, and the ACA method of computer interactive interviewing.

DESIGN OF THE STUDY

In the design phase, the first two steps, i.e., specification of the attributes and their levels are the same irrespective of the method used.

In the third step, i.e., specification of the attribute combinations, there is a difference. In traditional conjoint, one has a choice of using either a pairwise comparison approach or the full profile approach. For a small number of attributes, the pairwise approach is feasible. As the number of attributes increase, the number of pairs to be shown to the respondent can increase dramatically (for 8 attributes, one could have 28 tradeoff tables) although partial designs frequently are used. The full profile method appears to be more common in the industry.

In determining the combinations to be shown to the respondents, the pairwise approach is simple, as one only makes tables with the levels of each attribute shown on the sides of the tables. The full profile methods generally use fractional factorial designs. These are available in published form (Addelman 1962). Some computer programs, like Conjoint Designer (Carmone 1986), are also available to help generate the profiles to be used. These profile specifications then have to be printed on cards and the card sets have to be created for respondents. The cards should not be too large or too small; the respondents may find it difficult to handle them. All the respondents get the same set of profile cards.

In ACA, this task is incredibly simplified. You just input the attributes and their levels in the program, and it decides what levels will be shown and when. Additionally, respondents are shown attributes and levels that they find important and relevant in their decision making. This is customized for each respondent. This keeps up the interest. You can also control how many attributes are shown on the screen in each profile (up to a maximum of 5). In the traditional methods, all the attributes are shown at one time. This can create information overload, and the respondents may resort to just looking at the more important attributes. This overload will not happen in ACA.

The form of presentation in ACA is predetermined. The pairwise tradeoff method is used. In traditional conjoint, one can choose between the tradeoff table approach or the full profile method, as already mentioned.

The analysis technique to be used will determine the form of the judgments to be obtained from the respondents. Nonmetric methods work with ranking data, while metric methods are more often used with rating type data.

FIELD ADMINISTRATION

In traditional conjoint, one needs multiple sets of the profile cards, questionnaires, and response forms. In ACA, the need for all of these is obviated.

One of the concerns in field administration is the amount of time it takes to do a traditional conjoint versus an ACA administration.

TABLE 1

COMPARISON OF TIMES BETWEEN ACA AND TRADITIONAL CONJOINT

		ACA	TRADITIONAL CONJOINT
FINKBEINER AND PLATZ (1986)			
3^4 x 4^2 x 2^2 DESIGN	ACTUAL TIME	14.2	9.4 *
8 ATTRIBUTES			
16 CARDS	PERCEIVED TIME	12.0	7.6 *
BANKING			
N=173			
AGARWAL (1988)			
5^1 x 4^6 x 2^3 DESIGN	ACTUAL TIME	17.7	16.2 *
10 ATTRIBUTES			
32 CARDS	PERCEIVED TIME	19.3	19.4
JOBS			
N=205			

* Significant at the .05 level

Table 1 shows some results from Agarwal (1988). Although Finkbeiner and Platz (1986) found a difference of about 5 minutes in the two tasks, Agarwal (1988) found the difference to be only about a minute and a half, and this was for a much larger design than the one used by Finkbeiner and Platz. One reason for this may be that in the Agarwal (1988) study, the respondents were students who may have been much more familiar with computers than the normal respondent.

In ACA, some other issues, primarily relating to doing computer interviewing, become of concern:

1. The easy access of computers. In central location interviewing, this should not be a problem. In doing studies at the respondent locations, this may be a problem. Laptops are now available, but sometimes renting enough of them in all kinds of geographical areas can be difficult.

2. The availability of interviewers comfortable with computer software and hardware. If any unforseen problems arise, someone should be available locally to fix them. This includes things like bad diskettes, incompatible hardware, etc.

ESTIMATION OF THE PARTWORTHS

In traditional conjoint, one can use nonmetric methods, i.e., LINMAP, MONANOVA etc., or metric ones, primarily ordinary least squares regression (OLS). There is enough evidence available to show that OLS does as well as nonmetric methods in terms of validity (Carmone, Green, and Jain 1978; Jain et. al. 1979).

If OLS is used, then individual regressions have to be run for each respondent. The data has to be first put in a form appropriate for individual regressions. This involves considerable preprocessing. After the regressions are run, the estimated partworths have to be saved in a data file. Surprisingly, even in a package like SPSS, there is no option to write the results of the regressions to a file. SAS does have this option. After this, the estimated partworths have to be combined with the original data file for each respondent. Again this involves a nontrivial amount of post processing.

In ACA, all the above steps are bypassed. The partworths are estimated and ready to use immediately after the interview is over. This can save a tremendous amount of time.

VALIDITY / RELIABILITY OF RESULTS

Traditional conjoint methods have been shown to be fairly robust. Even in the presence of error, they are able to recover the true partworths. The cross validity of traditional conjoint is also respectable, in that the correlations between actual and predicted choices using holdout profiles range from 0.37 (Akaah and Korgaonkar 1983) to 0.82 (Moore 1980).

The validity and reliability of ACA has been recently investigated in two studies by Agarwal (1987; 1988)[1]. The first of these was a simulation study where ACA was compared with LINMAP III and MONANOVA. Different levels of error were added to partworth configurations, then these were used as input to the three algorithms. The estimated partworths were then compared with the original partworths. The results indicated that there was a significant difference between ACA and LINMAP or MONANOVA. ACA was found to be better than the other two, for situations where the amount of input data used by them was equal to or less than that used by ACA. This was true for errors ranging from 0% to 30%.

The second study was a field study, where ACA was compared with full profile conjoint using OLS. Students participated in a three-week long study where they chose jobs. All the respondents did the ACA task and the full profile task, separated by a week. Holdout profiles were also used. The results indicate that there is no significant difference between the cross validity (i.e., the correlations between the estimated and actual

---

[1] The two working papers available from the author.

choices for the holdout profiles) between ACA and full profile method. The results from both these studies are very encouraging for ACA. Since the traditional conjoint methods have a long history, they have been pretty thoroughly investigated in terms of the validity and reliability. ACA performs at least as well as they, and may do even better for especially large designs with large numbers of attributes and levels.

CONJOINT SIMULATORS TO PROJECT MARKET SHARES

The last step in the process is the actual usage of the result. Conjoint simulators can be used as decision support methods for:

1. Product development
2. Competitive strategy analysis
3. Market share forecasting

In the traditional conjoint approach, considerable post processing has to be done to perform any of the above market simulations. If one is doing a large number of ongoing studies, then it is a one time cost to write a choice simulator.

In ACA the choice simulator is built into the package. In addition data from the Ci2 System for computer interviewing can be combined with the ACA results to do segment level analysis. Segments can be weighted to reflect actual distributions of the respondents in the market place. External factors like distribution, advertising support, etc., also can be incorporated into the simulations.

In choice simulators, the predictions are done using either the maximum utility model, the Bradley-Terry-Luce model, or the Multinomial logit models. ACA has the maximum utility and the logit model available and also has a correction for the Independence of Irrelevant Alternatives (IIA) assumption in the logit models. Another nice feature in ACA is that the utilities are calibrated for each individual from the data obtained in the calibration section. This should make the predictions more robust.

DISCUSSION AND CONCLUSIONS

ACA provides an elegant comprehensive package to help design studies, collect data, estimate partworths, and use the results of a conjoint study. ACA appears, from the limited studies done so far, to be as reliable and valid as traditional conjoint. Thus it can be used without too many reservations.

Some other results from Agarwal (1988) are also worth mentioning. Two aspects of ACA were investigated: the impact of the number of attributes (either 2 or 3) shown in the tradeoff pairs and the maximum number of pairs (either 15 or 30) shown in that part of the task.

In terms of how many attributes to show in the pairs in ACA, two appears to be adequate. This results in an easier task and also takes less time, without any loss in predictive validity.

The impact of the number of pairs on predictive validity is mixed. Although the maximum number of pairs was limited to either 15 or 30, the actual number of pairs was different as the respondents were allowed to eliminate unacceptable attribute levels. A correlation between the validity and the number of pairs actually shown was performed, for both the 2- and the 3- attribute case. For the 3 attributes, no significant correlation was found between the number of pairs shown and the validity. For the 2 attributes, there was a significant but weak correlation between the validity (measured by Spearman's rank order correlations) and the number of pairs. So it appears that if 2 attributes are shown in the pairs, it might make a slight difference if a higher number of pairs are shown. This result is tentative and not very strong.

The feedback from the respondents indicated that ACA was more interesting and taken more seriously. The ACA task was found to be less difficult.

Some other issues should also be considered in the choice between traditional conjoint and ACA:

1. An advantage of ACA is that if there are a large number of factors in the design, it obtains partworths by asking questions only on the factors most important to that respondent. That set of factors can be different for different respondents. If traditional paper-and-pencil tasks are to be used, the only way to deal with a very large number of factors is to reduce the factors judgmentally, estimate utilities at group rather than individual level, or use hybrid designs (Green 1984).

2. ACA can only handle main effects designs. The only way to include two attributes that interact is to combine them into one composite attribute with the number of levels equal to the product of the levels in the two attributes. Thus ACA will not be very useful in situations with a large number of interactions.

REFERENCES

Addelman, Sidney (1962), "Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments," Technometrics, 4 (February), 21-46.

Agarwal, Manoj K. (1987)," A Monte Carlo Study Investigating Configuration Recovery in Adaptive Conjoint Analysis," Working Paper # 87-139, School of Management, State University of New York at Binghamton.

Agarwal, Manoj K. (1988)," An Empirical Comparison of Traditional Conjoint and Adaptive Conjoint Analysis," Working Paper # 88140, School of Management, State University of New York at Binghamton.

Akaah, Ishmael P. and Pradeep K. Korgaonkar (1983)," An Empirical Comparison of the Predictive Validity of Self-Explicated, HuberHybrid, Traditional Conjoint, and Hybrid Conjoint Models," Journal of Marketing Research, 20 (May), 187-97.

Carmone, Frank J., Paul E. Green and Arun K.Jain (1978), "The Robustness of Conjoint Analysis: Some Monte Carlo Results," Journal of Marketing Research, 15, 300-3.

Green, Paul E. (1984),"Hybrid Models for Conjoint Analysis: An Expository Review," Journal of Marketing Research, 21 (May), 155-69.

Jain, Arun K., Franklin Acito, Naresh K. Malhotra, and Vijay Mahajan (1979), "A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models," Journal of Marketing Research, 16 (August), 313-22.

Moore, William L. (1980),"Levels of Aggregation in Conjoint Analysis: An Empirical Comparison," Journal of Marketing Research, 17 (November), 516-23.

Wittink, Dick R. and Philippe Cattin (1981), "Alternative Estimation Methods for Conjoint Analysis: A Monte Carlo Study," Journal of Marketing Research, 18, 101-106.

A COMPARISON OF RATING AND CHOICE RESPONSES IN CONJOINT TASKS

Jordan J. Louviere
University of Alberta

and

Gary J. Gaeth
University of Iowa

ABSTRACT

This paper compares rating and choice formats in conjoint tasks. Choice tasks have the advantage of directly generating choice data rather than simulating choice data, choices are the response of primary interest in most conjoint studies, choice tasks accommodate purchase delays or non-purchase options, choice tasks can be designed to examine considerably more profiles per respondent than rating or ranking tasks and choice data can be analyzed by means of widely available software. On the downside, choice responses require considerably more data per individual respondent to develop statistically efficient individual-level models; hence, one normally develops models for aggregates of individuals.

Evidence is presented that parallel rating and choice tasks produce statistically similar aggregate utility functions, and therefore, one can expect to obtain similar results from either response/task format. Because of the advantages of choice tasks, we suggest that conjoint researchers may wish to consider using choice response formats for some studies.

INTRODUCTION

It is now well-known that ranking and rating data produce very similar individual and aggregate results in full-profile conjoint tasks (See, e.g., Green and Srinivasan, 1978). As well, Cattin and Wittink (1985) report that most recent applications of full-profile conjoint tasks used rating responses. Unfortunately, neither rating nor ranking of conjoint profiles is the behavior of primary interest to most conjoint researchers. Rather, the behavior of interest is the choice that consumers or buyers are likely to make among competing profiles. Thus, although useful, and apparently externally valid (See, e.g., Levin, Louviere, Schepanski and Norman, 1983), rating and ranking tasks have some major disadvantages:

1. Respondents rate or rank profiles one-at-a-time, rather than choosing among competing options.

2. Tasks become difficult and/or tedious as the number of profiles increases.

3. It is difficult to get respondents to rate or rank more than 32 total profiles in rating or ranking tasks, hence, to obtain individuallevel results, tasks are often artificially restricted to a small number of attributes and/or a limited number of levels.

4. Although existing product offerings, purchase delay and nonpurchase options can be accommodated in rating or ranking tasks, one rarely sees such accommodation.

5. A choice simulator must be developed to forecast aggregate market shares from the individual-level conjoint results. Such choice simulators use stochastic equations based upon certain error assumptions in a deterministic manner to forecast choices, a process which is at best ad hoc, and at worse, a methodological contradiction.

Discrete choice response modes are a relatively recent development in conjoint technology (e.g., Louviere and Woodworth, 1983). Few comparisons have been made of discrete choices with other response modes. A review of the marketing and psychology literature indicates that there have been some comparisons of binary responses with ratings or rankings (Huber and Czajka, 1981; Green and Srinivasan, 1978, p. 18), but we were unable to find comparisons of multinomial choice responses with other responses. Green and Srinivasan (1978) suggest that such responses should produce results similar to ratings or rankings, but the Huber and Czajka (1981) results contradict this expectation for binary responses. We conclude that there is little theory or empirical evidence about this issue, and as we have access to several sets of data that permit us to make comparisons, we believe that we can shed some light on this issue.

Thus, the purpose of this paper is to review the advantages of choice tasks which obviate most of the aforementioned disadvantages of rating and ranking tasks. Empirical evidence is presented to support the contention that choice and rating tasks produce statistically equivalent aggregate results. If replicable and generalizeable, this means that researchers can feel secure about using choice response formats; hence, one can weigh the advantages of a choice response format versus rating or ranking formats for particular conjoint problems. Having anticipated our conclusions, the paper is organized as follows to reach them: First, we review choice tasks and their design because choice formats require a different design strategy than full-profile rating or ranking tasks. Next, we present evidence of the comparability of results from rating and choice tasks, reviewing three studies involving airline ticket options, new state and local park options, and alternative career options. Finally, we review the major conclusions of the paper.

WHY CHOICES?

Why do marketing researchers need another response mode for conjoint studies when current response formats appear to produce satisfactory results, design and analysis software are now available to implement these traditional formats, and simulators can easily be developed? The primary reason is that current conjoint tasks are not realistic; i.e., they do not study the process of primary interest, which is choice behavior. Not only are traditional full-profile conjoint tasks ill-suited for observing choice behavior, but one must make strong, and often questionable assumptions to translate the results of rating and ranking tasks into choices. For example, it has been well-known for over 15 years that strong assumptions like the constant ratio or "independence of irrelevant alternatives" (IIA) property (see, e.g., Tversky, 1972) (see also, Tversky, 1972) usually do not hold for real choice data.

Thus, assuming that respondents will choose the alternative with the highest predicted utility value in a choice simulator is tantamount to believing in IIA, a belief that is often ill-founded. One cannot avoid the consequences of this assumption by using ranking or rating information to explode a single choice set, using the expanded observations to estimate individual-level or more aggregate logit models. Unfortunately, one must make very strong assumptions about ranking behavior to derive stochastic choice models from such data, and such assumptions normally involve IIA. Hence, unless one can observe the choices that respondents make directly, one is normally forced to make strong and often untenable assumptions to translate predicted utilities into choices. That is not to say, however, that models produced from making such assumptions cannot forecast choice behavior reasonably well for certain problems. Rather, it is simply to warn that, in general, such assumptions will rarely be satisfied in practice, and therefore models produced using such assumptions will often be wrong, however well they might predict to holdout samples of rankings or choices (which, of course, are not the same as real choices).

More importantly, violations of IIA can make a very big difference in the advice one gives to managers based on conjoint tasks. For example, if respondents actually use nested or hierarchical elimination processes, advice based on assumptions of IIA not only will be wrong, but can be very misleading. Choice response formats have the major advantage, therefore, of allowing one to observe choice behavior directly. Furthermore, one can design choice tasks in ways that permit one to detect violations of assumptions like IIA. Once one knows that IIA is violated and can make inferences about the nature of the violations, one can apply a variety of alternative choice models to estimate the effects of attributes and alternatives on choice. This means that choice tasks have the advantage of high face validity and flexibility -- they can be designed to accommodate a wide range of possible choice processes and statistical choice models.

Thus, choice tasks have a number of major advantages over rating and ranking tasks that suggests that practitioners should give them serious consideration when deciding on a particular research approach. Some of the major advantages are as follows:

1. It is important to understand and study actual choices, not simulated choices. Choice tasks make this possible.

2. Choice tasks are a natural task for respondents and closely simulate the behaviors of primary interest in many marketing research projects involving conjoint analysis techniques.

3. Choice tasks make it easy and natural to accommodate current product offerings, as well as delay of purchase and non-purchase options.

4. Choice tasks make it easy to examine much more of the statistical response surface than is usually possible with traditional fullprofile conjoint tasks.

5. Choice tasks can be designed to accommodate a wide variety of choice models and utility specifications.

6. Choice tasks permit one to develop statistical choice models directly from choice data, thus obviating the need to develop conjoint simulators that may require questionable assumptions.

Choice tasks have the following disadvantages:

1. It is very difficult to develop individual-level choice models because discrete choice responses do not contain as much statistical information as rating and/or ranking responses.

2. The statistical properties of discrete multivariate statistical models are only asymptotic; thus, even if one can estimate the utility parameters of interest, one cannot test them with any statistical confidence.

3. Choice tasks are more difficult to design than full-profile conjoint tasks because one must design both the profiles and the choice sets into which to place them.

Our experience with well over 100 applications of choice experiments since we first developed the design theory that permitted us to implement them in practical applications (See, e.g., Louviere, 1981), leads us to suggest that these disadvantages most often are offset by the advantages. For example, the lack of individual-level modeling capability can be offset by using a common set of choice sets to which all individuals respond to cluster individuals into segments exhibiting similar choice behavior. Obviously, one also could use other measures as a basis for clustering; however, similarity in choice behavior is often a logical and managerially important basis for deriving segments that behave differently with respect to the attributes or alternatives.

Rather than continuing to discuss the advantages of choice experiments, let us turn our attention to some simple examples of choice experiments that illustrate the advantages cited above.

Designing Choice Experiments To Illustrate Some Of Their Advantages

Louviere and Woodworth (1983) discuss a wide variety of possible ways to design and implement discrete choice or allocation experiments. They point out that choice experiments require one to develop two experimental designs: i) an alternative generating design to create full profiles, and ii) a choice set generating design into which to place the profiles. Thus, design of choice tasks requires the additional complexity of designing choice sets. Choice experiments are not new; one form of choice experiment -- the paired comparison experiment -- has been around for many years (see, e.g., Thurstone, 1927), and is sometimes used in conjoint tasks. However, paired comparison tasks require a large number of comparisons if the number of profiles exceeds 10, and are not very efficient statistically (Louviere and Woodworth 1983).

Although paired comparison experiments constitute one way to design choice experiments, one can design paired comparison experiments without using all possible pairs. Consider, for example, a problem involving seven two-level attributes. The minimum main effects plan to create the full-profiles for this problem requires eight treatments. This design is saturated, and all of the main effects are perfectly confounded with many unobserved interactions. Thus, to avoid the potential bias in the part-worth utility estimates that can result if one fails to control for any unobserved interactions, one might well choose to create a 16 treatment design in which some two-way interactions can be estimated. Alternatively, one might opt for a 32 treatment design that allows one to estimate all main effects and two-way interactions. Of course, there are other design possibilities, but these represent common alternatives. Clearly, asking respondents to compare all possible pairs is impractical in field applications for the 16 or 32 treatment designs.

This design problem represents a good case to illustrate some of the advantages of choice experiments because there are a variety of design possibilities that one might consider if one wants to develop a choice experiment instead of a rating or ranking experiment:

1. One could create a 32 treatment design that permits estimation of all main effects and two-way interactions, and fractionate this design into two 16 treatment main effects plus selected interactions plans. Random pairing of the profiles from the first fraction with those from the second creates 16 different choice sets. If one accepts the IIA assumption, one can estimate all main effects and two-way interactions from respondents' choices among the pairs. This design requires only 16 choices to be made, a simple task for most respondents (see Louviere, 1988 for a discussion of this design strategy). Table 1 contains an example of this type of design strategy (see also Louviere, 1984).

The design illustrated in Table 1 requires respondents to compare 16 pairs of profiles, and to choose one of the two. A third alternative -- the option to delay choice or to choose neither -- could also be added (Louviere and Woodworth, 1983 cite the advantages of adding a constant option). The column labelled RANDOM# is the random number assigned to each profile, and is used to create the pairs: the two profiles whose RANDOM# equals "1," for example, are paired, then the two with RANDOM# equal "2," etc. Respondents' discrete choices are aggregated into frequency counts, and the aggregate counts are analyzed using binary or multinomial logit regression (or some other choice model if one thinks it appropriate).

TABLE 1: A PAIRED COMPARISON CHOICE EXPERIMENT USING TWO HALVES OF A 32 TREATMENT, 2-TO-THE-7, ORTHOGONAL, FRACTIONAL FACTORIAL

	TREATMENT SET 1								TREATMENT SET 2							
TRMT #	A	B	C	D	E	F	G	RANDOM#	A	B	C	D	E	F	G	RANDOM#
1	1	1	1	1	1	1	1	16	1	1	1	1	2	2	2	10
2	1	1	1	2	2	1	1	13	1	1	1	2	1	2	2	9
3	1	1	2	1	1	2	2	4	1	1	2	1	2	1	1	7
4	1	1	2	2	2	2	2	14	1	1	2	2	1	1	1	2
5	1	2	1	1	2	1	2	6	1	2	1	1	1	2	1	1
6	1	2	1	2	1	1	2	5	1	2	1	2	2	2	1	15
7	1	2	2	1	2	2	1	15	1	2	2	1	1	1	2	6
8	1	2	2	2	1	2	1	8	1	2	2	2	2	1	2	14
9	2	1	1	1	1	2	1	1	2	1	1	1	2	1	2	12
10	2	1	1	2	2	2	1	12	2	1	1	2	1	1	2	11
11	2	1	2	1	1	1	2	10	2	1	2	1	2	2	1	5
12	2	1	2	2	2	1	2	11	2	1	2	2	1	2	1	16
13	2	2	1	1	2	2	2	7	2	2	1	1	1	1	1	3
14	2	2	1	2	1	2	2	2	2	2	1	2	2	1	1	8
15	2	2	2	1	2	1	1	3	2	2	2	1	1	2	2	13
16	2	2	2	2	1	1	1	9	2	2	2	2	2	2	2	4

2. One can examine additional treatments by creating additional 16 treatment designs to have the statistical properties one desires, assigning these at random to additional columns (e.g., a 3rd or 4th column) to create choice sets of size three, four, or whatever one thinks is appropriate. We have used this technique to examine between two and nine alternatives (see, e.g., Louviere, Schroeder, Louviere and Woodworth, 1987; Louviere, 1986) in a number of studies. By increasing the number of profiles among which respondents choose, one can greatly increase the number of profiles that are observed. Hence, such choice experiments are by design more efficient statistically than full-profile conjoint experiments for estimating generic utility functions.

3. Design suggestions 1 and 2 normally require one to assume that IIA holds. If one is unwilling to make that assumption, there are other design possibilities (see, e.g., Louviere and Woodworth 1983). In particular, one could create pairs of profiles by treating the seven two-level attribute example as a two-to-the-14 design (seven two-level attributes for each of two alternatives). Sixteen choice sets can be designed by selecting a 16 treatment main effects plan to vary the 14 attributes. Alternatively, one can create larger designs and generate more choice sets if one thinks that respondents can handle the task, or one can create different designs to assign to different groups of subjects.

Table 2 illustrates the use of a 2-to-the-14 main effects plan to create 16 choice sets for two options, each of which have seven two-level attributes.

The design illustrated in Table 2 would not be appropriate if all attributes were quantitative because some profiles would dominate others. However, it is usually a simple matter to interchange columns or reverse levels to eliminate dominance and retain the independence properties. As with the design illustrated in Table 2, one can add the delay or no purchase option to this design. Respondents choose one of the options (or to delay or not purchase) in each choice set. Aggregate frequencies of choices are analyzed using binary or multinomial logit regression (or some other choice model specification).

The main effects plan used in Table 2 to create the choice sets and the profiles in the choice sets has the property that all attributes of both alternatives are independent of one another. This permits one to test whether the attributes of one alternative significantly affect the second, which constitutes a test of the IIA assumption (see, e.g., McFadden, 1986; Louviere and Woodworth, 1983). Cross-effects multinomial logit models allow one to capture violations of IIA and are versatile tools for developing choice models that capture violations of IIA.

TABLE 2: USING A 2^{14} MAIN EFFECTS PLAN TO CREATE 16 CHOICE SETS IN WHICH TWO ALTERNATIVES COMPETE

	OPTION A ATTRIBUTES							OPTION B ATTRIBUTES						
CHOICE SET #	A	B	C	D	E	F	G	A	B	C	D	E	F	G
1	1	1	1	1	1	1	1	2	2	2	2	2	2	2
2	1	1	1	2	1	1	2	1	1	2	1	1	1	2
3	1	1	2	1	1	2	1	1	1	1	2	2	1	1
4	1	1	2	2	1	2	2	2	2	1	1	1	2	1
5	1	2	1	1	2	1	2	1	2	1	1	2	1	1
6	1	2	1	2	2	1	1	2	1	1	2	1	2	1
7	1	2	2	1	2	2	2	2	1	2	1	2	2	2
8	1	2	2	2	2	2	1	1	2	2	2	1	1	2
9	2	1	1	1	2	2	1	1	1	2	1	1	2	1
10	2	1	1	2	2	2	2	2	2	2	2	2	1	1
11	2	1	2	1	2	1	1	2	2	1	1	1	1	2
12	2	1	2	2	2	1	2	1	1	1	2	2	2	2
13	2	2	1	1	1	2	2	2	1	1	2	1	1	2
14	2	2	1	2	1	2	1	1	2	1	1	2	2	2
15	2	2	2	1	1	1	2	1	2	2	2	1	2	1
16	2	2	2	2	1	1	1	2	1	2	1	2	1	1

One can create as many independent sets of profiles as one wishes by creating successively larger main effects plans to place the profiles into independent sets. For the present example of seven two-level attributes, three alternatives could be generated by means of a 2-to-the-21 main effects plan, which can be designed in as few as 24 treatments. Thus, one can use these designs to test violations of IIA assumptions and to observe choices directly without the need to develop a conjoint choice simulator. Levels can be interchanged as needed in each attribute column in order to eliminate dominance problems in the competing profiles.

4. Choice designs can easily accommodate brand names or, more generally, alternative names. One can simply "brand" the alternatives in design examples 1, 2, or 3 by making each set of 7 attributes pertain to a particular brand. This has the advantage of explicitly nesting attributes under brand, thereby permitting one to estimate brand-specific effects for each attribute. This allows each brand to have differential effects for price, features, etc. If one uses design options 1 or 2, one can obtain brand-specific effects, but only in special cases can one obtain cross-effects.

Design option 3 permits one to develop both brand-specific effects and cross-effects. That is, one can estimate the effects of the attributes of brand A on the choice of brand B, and vice-versa. If any of these cross-effects are statistically significant, this signals a violation of IIA. Such violations can be captured by the addition of the cross-effects terms in the utility specifications.

5. Choice designs can easily accommodate existing brands, delay of purchase or no purchase options. In the case of existing brands, design illustration 4 provides one way of handling variations in the attributes of such brands. However, if one wishes to assume that particular brands currently available will remain constant, one can offer the respondents a choice between a new product concept described by attribute profiles and the current brands. As well, one can ask subjects in design illustrations 1-4 to say whether, if faced with a particular choice set, they would delay purchase or not purchase. Choice tasks make these decisions natural, and their analysis is also appealing because the effects of these options are included in the model analyses directly.

The illustrations discussed above represent only some of the many possibilities for designing choice experiments that have good statistical properties. For example, Louviere and Woodworth (1983) discuss the use of 2-to-the-N (N is the number of alternatives) designs, which can be used to vary choice set size and composition (see also McFadden, 1986). As well, balanced incomplete block designs and other designs can be used to vary choice set size and composition, or to keep the number of alternatives down to a manageable size in each choice set. Thus, unlike traditional full-profile or pairwise tradeoff conjoint experiments, the number of ways to design choice experiments is very large. Furthermore, choice experiments allow one to i) simulate real choice environments, ii) observe many more profiles per respondent and iii) use non-profile options such as current brands, purchase delays, or non-purchases.

Finally, choice experiments can be merged with data from real choice situations because, unlike traditional conjoint experiments, the types of models one employs to analyze choice data can be applied to analyze choices in conjoint choice experiments or choices recorded in real world sources of data such as surveys, scanner panels, inventory withdrawals, etc. Thus, the models one employs to analyze choices in conjoint choice experiments belong to the same family of models that one uses to analyze parallel real world choice data.

DO RATING AND CHOICE RESPONSES PRODUCE SIMILAR RESULTS?

In this section we discuss three studies in which both rating and choice responses were used in such a way that the parameters of models based on these response modes could be compared.

1. Choice Of Airline Ticket

The first study involved 35 middle level and senior managers who attended a marketing research seminar in Sydney, Australia. They were asked to rate the desirability of 16 different tickets from Sydney to the West Coast of the US. These tickets were described by levels of fare in Australian dollars, advance payment requirement in days, number of stopovers en route and cancellation penalty as a percent of the total fare paid. The design was based on a complete factorial of each attribute at "high" or "low" levels. In practice, the values of "high" or "low" were randomly assigned from ranges of "high" and "low" such that four different values of "low" and "high" appeared for each attribute. Although this design is not exactly orthogonal, it is highly efficient. Participants rated each of the 16 tickets on a 13 category scale relative to a standard ticket which was described as being average on all attributes.

Following the rating task, participants were asked to choose the ticket that they thought was the most desirable to fly to the West Coast of the US in each of 15 choice sets. The tickets contained in these choice sets were based on a 12 treatment main effects plan, and were placed into choice sets according to a 2-to-the-12 main effects plan as described in Louviere and Woodworth (1983). OLS regression was used to estimate parameters from the ratings data, while multinomial logit regression was used to estimate parameters from the choice data. Parameters are compared in Table 3.

TABLE 3: COMPARISONS OF CHOICE AND RATING ESTIMATES FOR AIRLINE TICKETS

	Choice Data	Rating Data
Airfare	-.0010	-.0010
Advance	.0007	.0003
Stopovers	.3000	.5480
Cancellation	-.4800	-.3210

The correlation between the two vectors of parameters is .94, and a regression of the rating parameters on the choice parameters yields an intercept of -0.93 and a slope of .84, with a standard error on the slope of .21. The regression equation is significant at the .06 alpha level. Thus, we cannot reject the hypothesis that the parameters are equivalent for the two sets of data. Moreover, we would draw similar inferences regarding the effects from both models.

2. Career Choice

A national sample of 400 US high school and college students evaluated hypothetical careers described by 22 attributes. Subjects were randomly assigned into a rating or choice condition; within each of these two conditions there were two additional conditions: 1) a 2-to-the-22 main effects plan (100 subjects) and its foldover (100 subjects), and 2) a 3-to-the-6 main effects plan combined with a 2-to-the-6 main effects plus selected interactions plan (100 subjects) and the foldover of these two plans (100 subjects).

The 22 attributes were assigned levels of "high" or "low" based on the students' previous ratings of the position of a sample of 26 careers on the 22 attributes. The 6 attributes were based on a factor analysis of the ratings of the 22 attributes for the sample of 26 careers; they represent the composite "factors" uncovered in the factor analysis. The purpose of the two separate conjoint studies is to compare choices among careers based on independent vectors of the 22 attributes with choices based on composite factors consisting of sets of correlated attributes.

The results of the four different conditions are listed in Tables 4 and 5. Table 4 contains the results for the 22 independent attributes, while Table 5 contains results for the six composite attributes.

The estimates in Table 4 have a correlation of 0.73; the slope and intercept of the regression of the rating estimates on the choice estimates are respectively, 0.4 and 0.06. The standard error of the slope is 0.08. These results indicate that the estimates are proportional to one another, but are not identical. Nonetheless, a glance at Table 4 indicates that many of the estimates are quite close.

The two sets of estimates in Table 5 have a correlation of 0.97. The intercept and slope are respectively, -0.019 and 1.072; the standard error of the slope is 0.037. Thus, these estimates appear to be statistically identical.

TABLE 4: RATING AND CHOICE ESTIMATES FOR 22 CAREER ATTRIBUTES

ATTRIBUTES	CHOICES	RATINGS
SALARY	.283	.388
PERKS	.102	.088
PRESTIGE	.124	.103
ADVANCEMENT	.142	.139
JOB SECURITY	.109	.149
STRESS	.005	-.028
WORK ENVIRONMENT	.079	.079
EDUC. REQUIRMENT	.075	.046
SKILL TRANSFER	.066	.057
INTRICATE WORK	-.050	.002
USE INTELLECT	.084	.113
JOB EXPERIENCE	.039	.049
HUMAN INVOLVEMENT	.043	.087
WORK WITH KIDS	.057	.041
HELPING OTHERS	.094	.072
JOB VARIETY	.040	.124
JOB SATISFACTION	.138	.179
AMOUNT OF TRAVEL	-.081	-.011
AMOUNT OF TEDIUM	-.049	-.052
AMOUNT OF CLERICAL	-.062	-.024
AMOUNT OF DANGER	-.137	-.090
AMT. MANUAL LABOR	-.047	-.067
AMT. OF INDEPEND.	-.050	-.056

TABLE 5: RATING AND CHOICE ESTIMATES FOR COMPOSITE CAREER ATTRIBUTES

VARIABLE	CHOICE DATA	RATING DATA
COMPOSITE 1	0.937	0.976
COMPOSITE 2	0.345	0.314
COMPOSITE 3	0.475	0.398
COMPOSITE 4	0.269	0.226
COMPOSITE 5	-0.358	-0.285
COMPOSITE 6	-0.335	-0.243
COMPOSITE 1 SQUARED	-0.088	-0.033
COMPOSITE 2 SQUARED	-0.227	-0.136
COMPOSITE 3 SQUARED	-0.221	-0.147
COMPOSITE 4 SQUARED	-0.079	-0.004
COMPOSITE 5 SQUARED	-0.120	-0.065
COMPOSITE 6 SQUARED	-0.175	-0.164
COM 1 X COM 2	0.134	0.127
COM 1 X COM 3	0.071	0.051
COM 1 X COM 4	-0.063	-0.028
COM 1 X COM 5	-0.112	-0.077
COM 1 X COM 6	-0.045	-0.009
COM 2 X COM 3	0.015	0.010
COM 2 X COM 4	-0.016	-0.019
COM 2 X COM 5	-0.011	-0.011
COM 2 X COM 6	-0.078	-0.070
COM 3 X COM 4	0.001	0.038
COM 3 X COM 5	-0.045	-0.036
COM 3 X COM 6	0.013	-0.017
COM 4 X COM 5	0.084	0.085
COM 4 X COM 6	-0.047	-0.009
COM 5 X COM 6	0.094	0.004

3. City Parks

A sample of 512 Iowa residents were asked to evaluate various city parks that were described by the attributes access, social use, facilities, maintenance, and natural environment. Each attribute was given two levels and an eight treatment main effects plan was chosen to create different parks, which were placed into 15 different choice sets for the choice condition. The rating conditions were based on either a one-half fraction of a 2x2x2x2x2 design or a 24 treatment fraction of a 3x2x2x2x2 design. All respondents were interviewed in person, and all completed the choice task. In the case of the rating tasks, however, there were a number of different conditions used to test a variety of academic hypotheses. The rating results are averaged over these various conditions. The results are contained in Table 5.

TABLE 5: COMPARISON OF CHOICE AND RATING ESTIMATES FOR CITY PARKS

VARIABLE	CHOICE DATA	RATING DATA
ACCESSIBILITY	0.020	0.256
SOCIAL USE	0.924	6.911
FACILITIES	0.037	0.414
MAINTENANCE	0.025	0.287
NATURAL ENVIRONMENT	0.420	1.813

The correlation between the choice and rating estimates is 0.98. The slope and intercept are respectively, 0.135 and 0.023. Thus, the two sets of estimates are proportional to one another. Indeed, one can closely approximate the choice estimates by normalizing the ratings estimates (summing and dividing each by the total). The conclusions one would draw from examination of the two sets of effects is therefore approximately the same for either response mode.

The findings from these three studies indicates that choice and rating results are proportional to one another, and that one can draw similar inferences about the magnitude of the effects of the attributes from both. Of course, these are three isolated studies, and we require more studies before one can be confident of the comparability of the two response modes. Yet, if a choice experiment provides evidence of the failure of the IIA assumption, one could expect less correspondence between rating and choice based estimates. All of the examples cited above require one to assume that IIA holds to estimate parameters from the choice tasks. Future research should examine the generality of the present finding for tasks in which IIA can be tested.

DISCUSSION AND CONCLUSIONS

This paper reviewed discrete choice and ratings responses for conjoint experiments. A number of advantages and disadvantages of discrete choice experiments were noted: Choice experiments have the advantage of being closer approximations to tasks that consumers actually do, easier than rating or ranking tasks, capable of including many more treatments, better at providing evidence of violations of assumptions of naive choice models or choice rules, and better able to accommodate brand name and brand-specific effects. Choice tasks have the disadvantage of not being able to provide individual-level models in most cases and of requiring additional levels of experimental design skills to develop statistically efficient designs.

We also described, discussed, and illustrated several different approaches to the design of choice experiments. Basically, if one accepts common assumptions made in conjoint rating or ranking tasks, one can easily design experiments that allow examination of two to four times as many profiles in a single task as can be examined in more traditional rating or ranking approaches. As well, because choice responses require minimal measurement assumptions and because the results are constrained to add to a fixed sum in each choice set, one does not have to be concerned about scale usage differences among respondents, and one can aggregate the choices of different samples of respondents over different sets of choice sets. Although these advantages come at the cost of sacrificing individual-level analyses, most managers are interested in aggregate-level results for segments or the market as a whole. Hence, such tradeoffs are usually not problematic.

Several data sets in which rating and choice response modes could be compared were analyzed, and it was concluded that similar inferences would be drawn about the relative effects of the attributes in the two different tasks. If these findings can be generalized and replicated, it would suggest that ratings and choice tasks can produce comparable results.

LIST OF REFERENCES

Cattin, P. and D. Wittink (1985) "Commercial Use of Conjoint Analysis: An Update." Paper presented to the ORSA/TIMS Marketing Science Meetings, Richardson, Tx (March).

Green, P.E. and V. Srinivasan (1978) "Conjoint Analysis in Consumer Research: Issues and Outlook." Journal of Consumer Research, 5, 103-123.

Huber, J. and A.L. Czajka (1981) "Modeling Buy/No Buy Decisions: A Comparison of Two Methods." Advances In Consumer Research, 9, 357-362.

Levin, I.P., Louviere, J.J., Schepanski, A.A. and K.L. Norman (1983) "Validity Tests and Applications of Laboratory Studies of Information Integration," Organization Behavior and Human Performance, 31, 173-193.

Louviere, J.J. (1981) "Specification of the Form of the Utility Expression for Tradeoffs in Choice Behavior," In Hensher, D.A. and L.W. Johnson, Applied Discrete Choice Modelling, London: Croom-Helm.

Louviere, J.J. (1984) "Using Discrete Choice Experiments and Multinomial Logit Choice Models To Forecast Trial In A Competitive Retail Environment: A Fast Food Restaurant Illustration," Journal of Retailing, 60(4), 81-107.

Louviere, J.J. (1986) "A Conjoint Model For Analyzing New Product Positions In A Differentiated Market With Price Competition," Advances in Consumer Research, 13, 375-380.

Louviere, J.J. (1988) Modeling Individual Decisions: Metric Conjoint Analysis, Theory, Methods and Applications, Sage University Papers Series, Newbury Park, Ca: Sage Publications, Inc.

Louviere, J.J. and G.G. Woodworth (1983) "Design and Analysis of Simulated Choice or Allocation Experiments: An Approach Based on Aggregate Data" Journal of Marketing Research, 20, 350-367.

Louviere, J.J., Schroeder, H., Louviere, C. and G.G. Woodworth (1987) "Do The Parameters of Choice Models Depend On Differences In Stimulus Presentation: Visual Versus Verbal Presentation?" Advances in Consumer Research, 14, 79-82.

McFadden, D. (1986) "The Choice Theory Approach To Market Research," Marketing Science, 5(4), 275-297.

Thurstone, L.L. (1927) "A Law of Comparative Judgment," Psychological Review, 34, 273-286.

Tversky, A. (1972) "Elimination By Aspects: A Theory of Choice," Psychological Review, 79, 281-289.

COMPARISON OF CONJOINT CHOICE SIMULATORS

Carl T. Finkbeiner
National Analysts Division of Booz·Allen & Hamilton Inc.

So you want to use your conjoint results for something besides plotting attribute utilities? You recognize that the conjoint task yields a type of preference data and you would like to obtain estimates of the demand implied by these preferences. Some of you may know about all the available technology for estimating demand from conjoint results, but you are unsure about which, if any, offer any real advantages in practice. Others of you may have heard of choice simulators -- you might even have the use of the Bretton-Clark CONJOINT ANALYZER or of the Sawtooth Software Adaptive Conjoint Analysis (ACA) programs -- but phrases like maximum utility or first choice, Bradley-Terry-Luce, multinomial logit, or multinomial probit leave you with the cold sinking feeling that you may have missed something in your research methods or statistics courses. Which techniques should you use? If two approaches yield different estimates, which one is more likely to be nearer the "truth"?

This paper offers some theory, data, and a little advice on the topic of choice simulators. A summary of several simulators will be provided, as well as some empirical results bearing on the validity of the demand estimates.

Our attention will be focused on conjoint studies in which partworths are estimated for individual respondents, such as is the case with ACA or with the relatively standard fullprofile, orthogonal design approach as described in a number of textbooks (e.g., Green & Wind, 1973; Urban & Hauser, 1980). The conjoint design consists of a set of attributes, each with its own set of levels, which can be used to "construct" a hypothetical product as a combination of attribute levels. In most conjoint studies of the type with which we are concerned, the primary data are obtained by presenting respondents with combinations of the attributes and asking them to rate the desirability of the combinations relative to other possible attribute combinations.

The output of a conjoint analysis of a respondent's data is an equation which allows us to estimate the respondent's "total utility" for any possible "product" (attribute combination), even those not actually rated by the respondent. The coefficients in this equation are the respondent's partworths (or attribute utilities). "Total utility" is a theoretical measure of the relative attractiveness of a "product" to the respondent. We will not concern ourselves here with the process of designing a conjoint study or of

---

* Acknowledgement: A portion of the data used in the analyses are from the Customer Preference and Behavior Project (RP-2671) funded by the Electric Power Research Institute.

conducting the analysis of the conjoint data, but turn instead to the subject of conjoint choice simulation.

Once a conjoint analysis is completed, it is common to use a choice model with the equations for estimating total utilities to obtain share estimates (Huber & Moore, 1979, Wiley & Low, 1983). These share estimates are often referred to as preference shares, to indicate that they are shares of choice as predicted by preference (conjoint) data only. The process of estimating shares from the conjoint total utilities is referred to as "choice simulation."

In the remainder of this paper, we consider several conjoint choice simulators. Empirical validation studies are described and the results are used to draw some conclusions. Before presenting the choice simulators, we describe one of the validation studies in detail to make the discussion more concrete.

The Checking Account Study

A conjoint study of checking accounts was undertaken for the purposes of comparing two methodologies:

- ACA -- Adaptive Conjoint Analysis
- A standard full-profile, orthogonal design conjoint in which respondents' ratings of attribute combinations were recorded using paper and pencil (referred to hereafter as the P&P approach)

Complete, valid data were obtained from a split sample of 75 and 77 respondents for the ACA and P&P methods, respectively. The sample was a convenience sample.

The attributes and their levels are shown below.

1. Average monthly balance required to avoid a service fee
    a. $500
    b. $1,000
    c. $2,000
2. Service fee per month if balance below average required
    a. $3
    b. $6
    c. $9

3. Time to get to the closest branch office

   a. 5 minutes
   b. 10 minutes
   c. 20 minutes
   d. 30 minutes

4. Availability of loan service at the branch you would use

   a. Yes
   b. No

5. Time to get to nearest automatic teller machine location

   a. 5 minutes
   b. 10 minutes
   c. 20 minutes
   d. 30 minutes

6. Type of institution offering the account

   a. Regular commercial bank
   b. Savings and loan

The P&P method required 16 attribute combinations to be sorted in order to estimate the partworths for an additive conjoint model.

Since the ACA method does not show a full-profile of the attributes to respondents, each attribute must make sense on its own. This is not the case for the first two attributes: if the service fee did not appear, the balance would make no sense, and vice versa. Thus, the first two attributes had to be combined into a new nine-level attribute:

1a. Balance and service fees

   a. $3 fee if average monthly balance falls below $500
   b. $3 fee if average monthly balance falls below $1000
   c. $3 fee if average monthly balance falls below $2000
   d. $6 fee if average monthly balance falls below $500
   e. $6 fee if average monthly balance falls below $1000
   f. $6 fee if average monthly balance falls below $2000
   g. $9 fee if average monthly balance falls below $500
   h. $9 fee if average monthly balance falls below $1000
   i. $9 fee if average monthly balance falls below $2000

We note that with attribute 1a. we are incorporating into our conjoint model an interaction term for the two attributes: service fee and balance. Thus, the ACA model actually has more parameters than the additive model used in the P&P approach.

This study has been described in detail elsewhere (Finkbeiner & Platz, 1986). The relevant conclusions drawn from the analyses performed were as follows:

- Both methods yield conjoint models which account for their corresponding conjoint data reasonably well.

- Aside from the interaction effect built into the ACA method, the methods produce partworths and importance scores that are interpretationally similar.

- Both methods allow roughly equivalent prediction of desirability ratings on hold-out "products," at both individual and aggregate levels of analysis.

A critical aspect of this study was the inclusion of a choice validation task. After the conjoint task was completed, a respondent was given four "products" and told to assume that he/she was going to select a new checking account and that these four represented the only available alternatives. The task was to choose one of the four "products." The respondent was then given the same four "products" plus two more and asked to repeat the same choice task with these six "products." These "products" had the following characteristics.

- An analysis of order-of-presentation effects showed that there were no such effects, justifying subsequent analyses without regard for order-of-presentation.

- The set-of-four "products" was created to span the range from undesirable to very desirable.

- The two additional "products" in the set-of-six were chosen to be like two of the original set-of-four, each varying only on one relatively minor attribute (a different attribute for each additional "product").

- Both the ACA respondents and the P&P respondents received the same sets-of-four and -six "products."

- None of the "products" had been seen previously in the interview by the respondent.

This choice task affords us an opportunity to validate conjoint choice simulators. Since actual choices were made by the same respondents who completed the conjoint, we can estimate preference shares using a simulator and compare these shares against the actual choice shares.

It must be emphasized that this validation task, by itself, is only a test of a model's ability to estimate choice shares in the artificial context of the conjoint interview. The validation task we are using does not allow us to evaluate a model's performance as an estimator of shares in the real-world market, where there are many sources of influence and uncertainty. This can be considered either a positive or a negative: on the one hand, we are not trying to evaluate our estimates in situations in which factors not included in the model are operative; on the other hand, the model which performs best here isn't necessarily the one which is most predictive of real-world shares.

Nonetheless, this is a common validation exercise and, for that reason alone, it is necessary to consider its results. Furthermore, although it may be difficult to generalize from this task, these results may still provide some useful insight into the different choice simulators if we consider the performance of a number of them.

Before examining the validation results, we describe several choice simulators in some detail.

Conjoint Choice Simulators

Many different choice models have been used to simulate choices in the conjoint context, ranging from the very simple to the very complex. We will restrict our attention here to utility maximization models, so-called because these models all assume that consumers tend to choose alternatives for which they have the highest total utility. Virtually all conjoint simulators fall into this general class.

As alluded to previously, we will also focus on models which operate on "product" total utilities calculated for individual respondents. We will not follow the practice of calculating shares based on utilities estimated at an aggregate level.

The following sections discuss the models we will compare empirically. The discussion is somewhat technical and the reader who is only interested in the outcome of the empirical comparison can safely skip these sections.

o Maximum Utility (First Choice)

This method is a relatively straightforward application of the utility maximization principle (Huber & Moore, 1979; Thurstone, 1945). A respondent is assumed to choose that "product" for which his/her total utility is highest.

In a conjoint simulation, we calculate total utilities for each of a set of competing "products," constructed to represent a market. (The set of competing "products" in a hypothetical, constructed market is referred to collectively as a "scenario.") In the checking account study, we would calculate the total utility for each respondent for each hypothetical checking account in a scenario: the checking account with the highest total utility is taken to be the one that respondent would choose.

The share estimate for a "product" is just the proportion of the sample for which that "product" has the highest total utility out of the total utilities for all "products" in the scenario. If two or more "products" are tied for the highest total utility, the respondent is assumed to choose randomly from among them, with equal probability. In mathematical terms:

$$s_j = \underset{i}{\mathrm{Sum}}(p_{ij})/N$$

where

s_j = estimated share for "product" j
p_{ij} = 1/n if $u_{ij} = \max(u_{i1}, u_{i2}, \ldots, u_{iJ})$
= 0 otherwise
J = the number of "products" in the scenario
n = the number of "products" tied for the highest total utility
u_{ij} = respondent i's total utility for "product" j
N = sample size
$\underset{i}{\mathrm{Sum}}$ = sum over i (i.e., over the N respondents)

This model makes no assumptions other than that of utility maximization. Therefore, unlike most other models, no assumptions are made about the, statistical distributions of the total utilities for the competing "products" in a scenario. However, Thurstone (1945) did suggest that this approach could be used as a good approximation to the prediction of choice from a generalization of the Case V Thurstone scaling model (now popularly referred to as the multinomial probit model, see below) which assumes a multivariate normal distribution for the total utilities of the competing "products." In that sense, there is a relationship between the maximum utility model and multinomial probit.

The maximum utility model has two important properties:

- It is the most scale-free of the models we will present here. Specifically, any monotonic increasing transformation of the total utilities within an individual respondent's data produces the same share estimates. This property is very important because the total utilities produced by most conjoint analysis models are, at best, only intervally scaled -that is, linear (or often monotonic increasing) transformations may be applied to them without changing their measurement properties. Consequently, the maximum utility model is the only one we will consider which is invariant under interval or ordinal rescaling of individual respondent data.

- It is computationally very simple and, hence, very efficient, making it quite feasible for implementation on microcomputers. It is safe to say that this is the fastest model of those we will consider.

Anecdotally, many users report that they feel this model is too unstable, producing share estimates that are too extreme. In fact, Wiley & Low (1983) report results from a Monte Carlo study in which the maximum utility model produces estimates that have worse sampling variance than another method. This instability is usually attributed to the fact that the model puts all of a respondent's vote on one "product" regardless of how close that "product's" total utility is to that of any other "product."

Maximum utility can be referred to as a deterministic model to reflect the fact that once total utilities are calculated, choice is exactly determined. However, real choices in the real-world are made with some degree of uncertainty. There are many sources for this uncertainty, including product availability, customer awareness, conflicting marketing efforts, variability in the perceptions of products, and the sheer overwhelming quantity of information available in competitive markets. A model which treats choice probabilistically -- rather than deterministically -- will not only produce stabler, less volatile share estimates (as Wiley & Low found), but must be better able to represent the uncertainties of real market choices.

- <u>ACA Logit</u>

In the ACA method of conjoint (Sawtooth Software, 1986), after the conjoint tradeoff tasks are completed, respondents are asked to rate on a 1-to-9 scale their probability of selecting each of 2 to 9 "products" referred to as "calibration products." These ratings are taken to be ratio scaled data. They are divided by a constant (10) so that the resulting values are in the range of probabilities (0-to-1) and are treated as such.

The respondent's partworths are then rescaled by a two parameter logistic function which is determined so as to best estimate the respondent's probability ratings after conversion to the 0 to 1 range. The variables thus produced can be referred to as "logits." Logits are converted to choice probabilities by dividing the logit for a "product" by the sum of the logits for all other "products" in the scenario. These individual choice probabilities are then averaged across respondents to obtain share estimates. The ACA simulator refers to this model as the "share of preference model, no correction for similarity." Mathematically,

$$s_j = \underset{i}{\mathrm{Sum}}(p_{ij})/N$$

where

$p_{ij} = y_{ij}/\underset{k}{\mathrm{Sum}}(y_{ik})$

$z_{ij} = a_i + b_i\, u_{ij}$

$y_{ij} = \exp(z_{ij})$

a_i , b_i = regression parameters from a logistic regression of respondent i's "calibration product" ratings on the total utilities calculated for those "products"

Note that the parameter a_1 is immaterial since it cancels out of the numerator and denominator in the equation for p_{ij}. Furthermore, note that if b_i is set large enough, this model approximates the maximum utility model. Whether the two models yield similar share estimates in practice is an empirical matter.

To derive the logit model, Domencich & McFadden (1975) assume that the z_{ij} are independently and identically distributed in the Gumbel distribution (a positively skewed, unimodal distribution often justified by its similarity to the normal distribution). The independence assumption is notable -- it states that z_{ij} is independent of z_{ik} for all "products" j and k. This assumption is often called the Independence of Irrelevant Alternatives (IIA) assumption.

The implication of IIA is that choices are affected only by the relative utility of "products" and not by "product" similarity. Intuitively, two similar "products" ought to compete with each other for share to a greater extent than would be the case with dissimilar "products," even if their total utilities happened to be the same. For instance, in the checking account study, suppose that two checking accounts with different fees, accessibility, etc. constitute a scenario and that the choice probability for checking account A is 60% and that the probability for checking account B is 40%. If a third checking account, C, is introduced which is identical to checking account B, then, intuitively, B and C ought to split B's original 40% choice probability. However, the logit model predicts that the combined probabilities of B and C will be 57%, implying that the product with the higher total utility (A) will be less likely to be chosen than a product with a lower total utility (B or C). Such a counter-intuitive outcome is an obvious disadvantage for the logit model.

It should be noted that the maximum utility model makes no such assumption as IIA, and, in fact, would exactly split the 40% probability for B between B and C, just as we would expect.

As regards the scale-freeness and computational efficiency properties of logit:

- The ACA logit model has two components for which scale freeness must be considered: the total utilities from the conjoint model and the direct probability ratings of the "calibration products." The final share estimates are invariant under linear transformations of individual respondent total utilities, but not under non-linear transformations. Different respondents' utilities may be subjected to different linear transformations without affecting share estimates. Furthermore, to the extent that the direct probability ratings do not have the ratio scale measurement properties assumed by the model, then the share estimates will be arbitrary and indeterminate. In the present author's opinion, the assumptions of interval scale properties for the total utilities and of ratio scale properties for the direct probability ratings are probably nearly enough true as to be acceptable for practical purposes. However, these scaling properties ought to be kept in mind, particularly in view of the less stringent measurement properties required by the maximum utility model.

- This model is computationally simple, although it will require more calculations for each respondent than does the maximum utility model. The difference in computation time, however, should not be enough to cause one to reject the logit model on efficiency grounds.

The IIA assumption was troublesome enough for ACA that it was necessary to find some means of adjusting the logit shares to account for "product" similarity.

o <u>ACA Adjusted Logit</u>

This approach is referred to by ACA as the "share of preference model, with correction for similarity." The basic idea for this adjustment is attributed to Richard Smallwood of Applied Decision Analysis, and is akin to an approach due to Lakshmi-Ratan, et al., (1984). The rationale appears to be that if two "products" are the same, they should divide the share that one of them alone would have; if three "products" are the same, they should divide the share of one alone, etc. This is accomplished in an ad hoc fashion, as follows:

A J by J matrix of similarities between pairs of "products" is constructed according to rules (described by Johnson, 1987) which take into account the differences between attribute specifications for the "products." The entries in this matrix (x_{jk}) range from 0 (completely dissimilar) to 1 (identical). The columns are summed to produce c_j, a measure for each "product" of its "total similarity" to the other products.

$$c_j = \underset{k}{\mathrm{Sum}}(x_{jk})$$

The "total similarity" for a "product" can be roughly interpreted as the number of "products" that "product" is like, ranging from a low of 1 (identical only to itself), through 2 (identical to one other "product"), and on up through a maximum possible value of J (identical to all other "products").

For a given respondent, i, the logit probabilities (p_{ij}) are then divided by c_j and the results are renormalized to sum to one, producing the adjusted logit probabilities, q_{ij}.

$$q_{ij} = (p_{ij}/c_j)/t_{ij}$$

where

$$t_{ij} = \underset{j}{\text{Sum}}(p_{ij}/c_j)$$

The final share estimates are then calculated as for the previous methods:

$$s_j = \underset{i}{\text{Sum}}(p_{ij})/N$$

To better understand the properties of this adjustment, consider the checking account example used to illustrate the IIA assumption of the logit method. Checking accounts A and B had logit probabilities of 60% and 40%, respectively. If checking account C, identical to B, is introduced, the (counter-intuitive) choice probabilities are:

	Logit
A	43%
B	29%
C	29%

Suppose that B and C are completely unlike A, so that the "total similarities" are 1, 2, and 2, respectively. Then, the adjusted logit probabilities are:

	Adjusted Logit	
A	60%	(A completely dissimilar)
B	20%	
C	20%	

The adjustment seemingly works as we would intuitively want it to. However, suppose B and C are somewhat similar to A, so that the "total similarities" are 1.5, 2, and 2, respectively. Now the adjusted logit probabilities are:

	Adjusted Logit	
A	50%	(A somewhat similar)
B	25%	
C	25%	

Such a result is still counterintuitive: the "product" with the higher total utility (A) is no more likely to be selected than one of the "products" with lower total utility. While this ad hoc adjustment has improved logit estimates, it has clearly not dealt with the IIA problem in a completely satisfactory manner.

The properties of scale-freeness and computational efficiency for the ACA adjusted logit model are as follows:

- Since the present model is a multiplicative rescaling of the unadjusted ACA logit estimates, their scale-freeness properties are exactly the same.

- The adjustment adds some arithmetic to the unadjusted logit calculations for each respondent, so that the computation time will be longer. However, even on microcomputers, the total elapsed time for most studies is still not a significant consideration.

Given the theoretical drawbacks of the logit models, it is desirable also to consider some other approach to probabilistic choice modeling as an alternative to maximum utility. We will retain the logit models for our empirical validation comparisons, however, since they are so often applied by ACA users and since theoretically incorrect models often work reasonably well in practice.

o <u>Simplified Probit</u>

As mentioned earlier, the probit model is a generalization of the Case V Thurstone scaling model. As a probabilistic choice model, probit is very attractive and has been shown to be superior to other popular choice models (see, for example, Currim, 1982). However, as typically formulated, the probit model cannot be used as a conjoint simulator because it cannot easily represent "product" changes. Consequently, we will follow the formulation of Finkbeiner (1986b) in order to obtain an applicable version of the probit model.

Drawing on the utility maximization principle, probit assumes that the "product" total utilities are jointly normally distributed in the respondent population. (Mathematically: the vector of total utilities for respondent i, $\underline{u}_i$, is drawn from a normal distribution with mean vector, $\underline{m}$, and covariance matrix, $\underline{C}$.) If total utilities were sums of independent variables (attribute level utilities), a general form of the Central Limit Theorem (Kendall & Stuart, 1977, p. 206) exists which tends to justify the normality assumption. This theorem states that the distribution of a sum of variables converges on normality as the number of variables increases, regardless of the form of the variables' original distributions. However, in conjoint, attribute level utilities are not statistically independent (although Central Limit Theorems have been developed for the dependent case) and the number of variables is never "large." Nonetheless, it should be expected that in many conjoint applications, the normality assumption is at least approximately true. When evidence in a particular study indicates that normality does not apply (e.g., via tests for normality), the simplified probit model can still be applied in segments for which normality more nearly holds.

The statistically best estimates of $\underline{m}$ and $\underline{C}$ are the means, variances, and covariances calculated on the total utilities across respondents. In matrix notation, this is:

$$\underline{m} = \underset{i}{\text{Sum}}(\underline{u}_i)/N$$

$$\underline{C} = \underset{i}{\text{Sum}}[(\underline{ui} - \underline{m})(\underline{u}_i - \underline{m})']/N$$

We now have an estimate of the population distribution from which our sample of total utilities was drawn. We can estimate the percentage of this distribution for whom any particular "product's" total utility is highest -- this will be the probit share estimate.

For "product" j, this percentage is obtained mathematically by multiple integration over the region of the normal distribution where $\underline{u}_j > \underline{u}_k$, for all k not equal to j. There are a number of approaches to solving this problem (Daganzo, 1979; McFadden, 1986), the most attractive of which is an approximation due to Clark (1961). This approximation is described in detail in Daganzo (1979). It is a fairly complicated procedure involving the use of higher order moments of the normal distribution and, although it is very efficient computationally, its output share estimates do not necessarily sum exactly to 100%. However, the deviation is slight and so the approximation should be modified slightly by normalizing the output shares so that they do sum to 100%.

From a purist's point of view, the Clark approximation is not completely satisfactory because it is not a very precise computation of the normal integrals. However, in the present author's pragmatic view, the use of the Clark approximation is justifiable on two counts:

- It is very fast.
- While it may not be a very good approximation to the normal integrals, it provides an exact value for some distribution and, furthermore, this distribution has many of the same properties as the normal: e.g., it is unimodal and it has the same scale-freeness properties as the normal distribution.

One obvious difference between the logit and probit approaches occurs in the context of aggregation:

- The logit model computes probabilities for each individual and then averages the probabilities across respondents to get shares. To justify this, logit requires only that its probability estimates conform to some commonly accepted axioms from probability theory.
- Probit aggregates first (computing $\underline{m}$ and $\underline{C}$) to get an estimate of the population distribution of total utilities and then calculates shares directly from that distribution. This practice requires that the total utilities be on a comparable scale across respondents. When a rating scale is used in the conjoint task, rather than a ranking or some other non-metric procedure, the partworths are calculated so as to produce total utilities which best estimate the original ratings. Therefore, it is as safe to assume that the total utilities can be aggregated as it is to assume that the original ratings can be. (Note that partworths in an ACA study are calibrated against the ratings from the calibration task, so they have the desired property.)

Note that probit explicitly does not assume IIA. The fact that a covariance matrix is used in the formulation means that there may be dependencies between the total utilities. Consequently, "product" similarities can affect the share estimates. Consider again the checking account example used to illustrate the IIA assumption of logit. The following population values for the total utilities for checking accounts A and B will yield probit shares of 60% and 40%, respectively, as in the logit example:

- Means: .654 for A and .400 for B
- Variances: 1 for A and B
- Covariance: .5

Now, if we introduce checking account C, identical to B, then it will have a mean of .400, a variance of 1, a covariance with A of .5, and a covariance with B of 1. The shares estimates are:

Probit	
A	59%
B	21%
C	21%

The probit method almost exactly splits B's original 40% share between the two identical checking accounts, as we would expect.

The scale-freeness and computational efficiency properties of probit are as follows:

- Probit is scale-free under certain linear transformations of the total utilities. In particular, total utilities for an individual respondent may be rescaled by an additive constant and total utilities across all respondents may be rescaled by the same multiplicative constant without changing the share estimates (see the theorem proven in the Appendix). This level of scale-freeness is somewhat more restrictive than that of the ACA logit models and is certainly more restrictive than that of the maximum utility model. However, the implied measurement properties of the total utilities (a sort of near interval scaling) are probably true enough in practice.

- The algorithm requires that means and covariances be cumulated across respondents, after which, one pass through Clark's approximation produces the share estimates in a matter of seconds, at worst. Thus, the probit method is at least as efficient computationally as the other methods.

o <u>Multivariate Logistic</u>

Wiley & Low (1983) have suggested a multivariate logistic model which is closely related to the probit model. In philosophy, this approach is the same as probit in that we define share at an aggregate level.

As was the case for probit, we assume that respondent total utilities have a normal distribution, with mean $\underline{m}$ and covariance matrix $\underline{C}$. However, here we restrict the correlations between all pairs of "product" utilities to be equal. It can be seen by reference to Bock (1975, pp.520-522) and Gumbel (1961) that the multivariate logistic model is a good approximation to this restricted version of the probit model. The multivariate logistic model for "product" 1 is:

$$S_1 = 1/(1+ \underline{1}'\exp(h[\mathrm{diag}\ \underline{GCG}']^{-\frac{1}{2}}\ \underline{Gm}))$$

where $h = pi/\sqrt{3}$ and $\underline{G}$ is an indicator matrix defined in the Appendix. Without loss of generality we can reapply this equation to all other "products" conceptually by simply rearranging "products" so that the "product" for which a share is being calculated is always first. Note that in the two-"product" case, the logistic will always be a good approximation for the probit model.

While the multivariate logistic model does not assume IIA (that is, it does allow some restricted level of dependency between total utilities in its constant correlations), it also does not go far enough to fully accommodate IIA, as can be seen in the following illustration.

In the checking account example for probit, we specified means and covariances for "products" A, B, and C which produced the probit results shown in the following table.

	Probit	Logistic		Probit	Logistic
A	60%	61%	A	59%	44%
B	40%	39%	B	21%	28%
			C	21%	28%

The introduction of the third product, C, is still clearly problematic for the logistic model. Nonetheless, as was the case for the ACA logit models, we will retain this model for the empirical comparisons in order to determine its validity in real, practical applications. In fact, it has been shown in other applications that, in certain situations, models with fewer parameters sometimes validate better than more general models (Dawes, & Corrigan, 1974; Ramsay, 1978).

The logistic model can be used in conjunction with the probit model to test whether, in a given application, the full probit model assumptions about non-constant covariances need to be retained. That is, share results from the probit and multivariate logistic models can be compared: if they are essentially the same, then the more restrictive assumptions of the logistic can be adopted; if they are different, then the more general assumptions of probit can be accepted.

The multivariate logistic has essentially the same scalefreeness and computational efficiency properties as probit:

- Because logistic uses the same matrix G as is used in probit, it shares the same somewhat restrictive scalefreeness property. More specifically, total utilities for an individual respondent may be rescaled by an additive constant and total utilities across all respondents may be rescaled by the same multiplicative constant without changing the share estimates.

- The logistic procedure is identical computationally to probit except that the share estimates are computed from the means and covariances of the total utilities using a different function. Consequently, it is also a very efficient algorithm.

o Two Inappropriate Models

In an effort to produce probabilistic models, two inappropriate models have been used. They are discussed here for completeness and to warn the reader against their use. In both models, the shares are estimated by averaging individual probabilities, as in the ACA logit model:

$$s_j = \underset{i}{\mathrm{Sum}}(p_{ij})/N$$

The simplest of the two inappropriate models is sometimes referred to as the Bradley-Terry-Luce model. The individual probabilities are calculated as:

$$p_{ij} = u_{ij}/\underset{k}{\mathrm{Sum}}(u_{ik})$$

The other model is sometimes called a logit model because it performs an exponential transformation on the total utilities:

$$p_{ij} = \exp(u_{ij})/\underset{k}{\mathrm{Sum}}(\exp(u_{ik}))$$

Neither of these models is appropriate because of their scale-freeness properties. Remember that the u_{ij} are (at best) interval scale data: i.e., they can be transformed by a linear transformation without affecting their ability to represent ratings on the original conjoint rating task. Consequently, we can consider the variable v_{ij}:

$$v_{ij} = a + bu_{ij}$$

For all of the previous models, substitution of v_{ij} for u_{ij} will not affect the share estimates. However, in these two inappropriate models, the share estimates depend upon the values of a or b. If we substitute v for u in the Bradley-Terry-Luce model, we obtain:

$$p_{ij} = v_{ij}/\underset{k}{\mathrm{Sum}}(v_{ik})$$
$$= (c + u_{ij})/(Jc + \underset{k}{\mathrm{Sum}}(u_{ik}))$$

where c = a/b. For the simple logit model, we obtain:

$$p_{ij} = \exp(v_{ij})/\underset{k}{\mathrm{Sum}}(\exp(v_{ik}))$$
$$= \exp(bu_{ij})/\underset{k}{\mathrm{Sum}}(\exp(bu_{ik}))$$

In neither case does the probability reduce to the original form so that share estimates depend upon whatever arbitrary choice we have made in scaling the total utilities. For instance, in the Bradley-Terry-Luce model, we can choose c so that, compared to shares from u, the shares are:

- Equalized
- Reversed
- Less than 0% or greater than 100%

Not only can choice of b in the simple logit model equalize or reverse shares, but it also can drive them to be equivalent to the maximum utility shares.

Unless additional data are brought to bear in order to determine the choice of b or c, this kind of arbitrariness in the share estimates is not acceptable. These two models should not be used unaltered in the conjoint simulation context.

We have identified 5 different choice models which can be used in conjoint simulation. They are:

- o Maximum Utility
- o ACA Logit
- o ACA Adjusted Logit
- o Simplified Probit
- o Multivariate Logistic

We turn now to the empirical validation studies as a basis for evaluating these methods. Some key questions suggested by the previous analysis should be borne in mind when reviewing the validation data:

- o Are the generality of its assumptions and its measurement propertiessufficient to overcome the determinism and instability of the maximum utility model?

- o Will the IIA assumption prove too restrictive in practice or will the ACA logit model's fairly general measurement properties allow it to perform well?

- o Does the ACA adjustment on the logit model adequately remedy the IIA assumption?

- o Do its more restrictive assumptions (measurement properties and the normality assumption) allow the probit model to take full advantage of its non-reliance on the IIA assumption?

- o Does the logistic model make assumptions too close to IIA to allow it to validate well, or does its close relationship to the probit model save it?

Illustration Of Validation Analyses

To simplify the discussion, we consider in detail only one validation: that of the maximum utility model on the ACA leg of the checking account study. For the set-of-four validation task, we specify the attribute configuration which defines each of the four "products" and calculate their total utilities for each respondent using respondent partworths from the ACA conjoint. The maximum utility model is then applied to these total utilities to obtain estimated shares. These can be compared to the actual shares shown below.

Set-of-Four Results (N = 75)

Actual	Max.Util.	Error
70.3%	64.0%	6.3
10.8%	13.3%	2.5
16.2%	22.7%	6.5
2.7%	0.0%	2.7

The error in the last column is in absolute value. The mean error is 4.5 share points. However, considering that the sampling error for very small actual shares is smaller than that for shares close to 50%, then a good measure of relative error ought to penalize a method more for producing the same size error when predicting an actual share of 10% compared to when predicting an actual share of 50%. Consequently, we index the absolute error against the 95% confidence interval for the corresponding actual share, to produce an error measure which we shall refer to as "relative error."

The confidence intervals for the above table are: 20.3, 14.2, 16.6, and 8.5, respectively. Thus, the relative errors are as shown below.

Set-of-Four Results (N = 75)

Actual	Max.Util.	Rel. Error
70.3%	64.0%	31.1
10.8%	13.3%	17.6
16.2%	22.7%	39.2
2.7%	0.0%	31.7

The mean relative error is 29.9. This should be interpreted as meaning that the average amount of error for the maximum utility method predictions is about 30% of the sampling error (as measured by the confidence interval). The only way to decide whether this is a good result or not is by contrast with other results.

Now consider the set-of-six validation task from the same study. The error table is shown below.

Set-of-Six Results (N = 75)

Actual	Max.Util.	Rel. Error
38.7%	33.3%	25.1
12.0%	6.7%	35.8
1.3%	1.3%	0.0
0.0%	0.0%	0.0
29.3%	30.7%	6.9
18.7%	28.0%	53.2

The mean relative error is 20.2, which is a little better than for the set-of-four validation. Remember that the first four "products" here are the same as the four in the set-of-four validation, the two additional "products" being different only in small ways from two of the original "products."

To simplify presentation, in the remaining analyses we will present only the mean relative error results. It should be noted that conclusions drawn from this analysis are essentially the same as those we would draw from detailed consideration of relative errors, "product" by "product," or even from consideration of the absolute errors.

Validation Results -- All Five Models

The set-of-four and -six validation results are shown below, rounded off to the nearest integer.

Checking Account Study (N = 75)

	Mean Relative Errors	
Model	Set-of-Four	Set-of-Six
Max. Utility	30	20
ACA Logit	85	70
ACA Adj. Logit	81	74
Probit	46	39
Logistic	31	46

The ACA logit models do considerably worse than the other methods here.

Probit and logistic appear to be similar in their predictive power. Surprisingly, maximum utility is at least as good as, if not better than, all other models. We turn to standard paper and pencil (P&P) results for further investigation of this result. (Note that without the ACA data collection method, we cannot apply the ACA logit models to any of the remaining validation studies.)

Maximum Utility vs. Probit vs. Logistic Validations

In addition to the P&P leg of the checking account study, we also have access to three other P&P studies in which the same kind of validation task was included. In each of these studies, four holdout "products," previously selected to cover a range from good to poor "products," were presented and the respondents were asked to indicate a choice.

The subjects of these three studies and the sizes of the conjoint designs were:

	- Size of Design -
Clothes Dryers	2x4x3x4x2x2x2x2
Water Heaters	4x4x3x3x2x2x2
Central Air Conditioners	4x4x4x3

In all three studies, respondents were sampled using a national area probability sample and were interviewed in person. A thorough description of the project of which these studies were a part can be found in two EPRI publications (EPRI, 1986a, 1986b).

The mean relative errors are presented in the following table. Results from the three models are validated against actual shares first, and then against one another, for the purposes of comparing similarity of model predictions.

Mean Relative Errors

		4-"Product" Validations					6-"Product" Validations	
Model	Compared To...	Chkng. Accts. ACA	Chkng. Accts. P&P	Dryers	Water Heater	AC	Chkng. Accts. ACA	Chkng. Accts. P&P
(N)		(75)	(77)	(268)	(266)	(76)	(75)	(77)
Max.Util.	Actual	30	23	7	43	23	20	27
Probit	Actual	46	29	27	53	28	39	25
Logistic	Actual	31	30	35	43	45	46	25
Max.Util.	Probit	20	14	24	33	15	31	13
Max.Util.	Logistic	15	22	32	20	24	28	8
Probit	Logistic	15	11	17	13	17	13	14

The average of the results across the five set-of-four validations and across the two set-of-six validations are as follows:

Summary of Relative Errors

Model	Compared To...	4-Product	6-Product	Mean
Max.Util.	Actual	25	23	24
Probit	Actual	37	32	35
Logistic	Actual	37	35	36
Max.Util.	Probit	21	22	21
Max.Util.	Logistic	23	18	21
Probit	Logistic	14	14	14

Findings

The following are the major findings obtained in this empirical validation.

o For these studies, measurement and specification error is probably not less than 15 relative error points, and may be more. In one instance, the logistic model out-performed the probit model by 15 points and, since the logistic is essentially a special case of probit, the only way in which that can happen is because of measurement or specification errors.

o The ACA logit and adjusted logit models have substantially worse validation results in the checking account study. On average, the relative errors for the ACA models are about twice that of the other models, with nearly a 40 point spread between the ACA models and the others.

o Although differences between the non-ACA models never exceed the 15 relative error point margin by much, the maximum utility model consistently does as well as or better than the probit or logistic models.

o The average difference in relative errors between maximum utility and probit is 12 points for the set-of-four validations, dropping to 9 points for the set-of-six validations. The difference is probably not statistically significant, although it is suggestive.

o By now, it should be no surprise that share estimates from the probit model and from the logistic model are more like one another than they are like the maximum utility model estimates. Probit shows a very small edge overall compared to the logistic, which is to be expected on theoretical grounds.

Discussion

The 15 point worst-case difference in relative error between the probit and logistic models can be used as an index of the minimum difference in relative error to attend to when viewing the results. It must be pointed out that the difference attributable to measurement or specification error might well be greater than 15 points and probably varies from study to study.

The relative inability of the logit models to validate presumably is largely due to their reliance on the IIA assumption. The correction for "product" similarity in the adjusted model does not overcome the problems of the logit model. While it is unsafe to conclude too much on the basis of one study, it appears as though the ACA models should be modified. As a suggestion, a multiplicative parameter should be included in the exponential term and calibrated using a holdout choice task similar to those used as validation tasks in the present studies. Note that if this altered model is to be validated, a second holdout validation task should be used. In the meantime, users of the ACA simulator would be advised to use the first choice model in that program.

The differences observed among the maximum utility, probit, and logistic models never exceed the 15 point range by much, so that measurement or specification errors may still be accounting for the observed differences. Nonetheless, the fact that the maximum utility model does as well as it does is surprising. It is not clear from these results that maximum utility can be written off as easily as most practitioners in the market research community tend to do.

The present validation data have provided insight into the relative performance of these simulators, as well as some provocative findings. However, the validation exercise should be put into proper perspective. Before making too much of these results, the reader should consider the following points:

The Monte Carlo study undertaken by Wiley & Low (1983) certainly suggests that our present results will not hold generally; there are situations in which the logistic model (and probit as well, since logistic is essentially a special case of probit) should be superior. Obviously, further evidence from other studies on the validity of maximum utility relative to the probit and/or logistic models would be desirable.

These validation studies all used essentially the same validation criterion: respondents chose one out of four or six "products." In fact, the number of "products" in the validation task was observed to have a small effect on the relative error difference between maximum utility and probit. It might well be that choice tasks involving fewer "products" create less uncertainty than is the case with more "products." If so, probit might validate better than maximum utility with, say, 10 or 12 "products" in the validation task, because the increased uncertainty makes a probabilistic model more appropriate. And, of course, 10-12 product scenarios are probably closer to the reality of most markets than are 4-6 product scenarios.

At the individual respondent level, the data from the validation task are not probabilistic in nature -- they are more like the deterministic maximum utility model in which only one "product" is selected with certainty, given "product" utilities. This puts probabilistic models at something of a disadvantage: they are representing an element of reality not present in the validation task.

We must also consider the fact that conjoint does not measure total utility without error. Measurement error implies the need for a probabilistic model; given measures of the "products'" utilities, we cannot infer choice with certainty since utility itself is not measured with certainty.

A point made earlier bears repeating here: real-world choices in real markets involve a great deal of uncertainty and "noise" which is not reflected in either the conjoint or the validation tasks. Factors which cause "noise" arise from some very diverse sources:

- Variation over occasions in a customer's...
  - Information processing rules in the presence of overwhelming quantities of information about the market
  - Perceptions of the products
  - Values (or importances) regarding product attributes, as would be predicted by diffusion theory
- Market factors, including:
  - Product availability
  - Customer awareness of products and of their attributes
  - Conflicting marketing efforts

All of these factors interfere with (or modify) the customer's simple preferences among explicitly defined "products," as captured in a conjoint context.

We have argued from a number of perspectives that choice has a strong probabilistic component:

- Uncertainty caused by information overload (e.g., too many products)
- Measurement errors in the total utilities
- Real-world "noise"

Therefore, to be realistic, the most appropriate validation task would be probabilistic in nature. This will be especially true when a conjoint simulator (with appropriate adjustment) is used to directly predict real-world shares, such as is done by Finkbeiner (1986a).

While it was appropriate and useful in the previous empirical comparisons to use the validation task which we did use, it is strongly recommended that alternative validation criteria be used in future validation studies. For instance, independent repeated choices by each respondent or larger numbers of "products" could be used in the task. However, until the sources of "noise" which moderate customer preferences in real-world markets have been investigated and until conjoint simulators are modified to directly model this "noise," it may be impossible to design validation studies which definitively compare the predictive power of choice models in a realistic context.

REFERENCES

Bock, D. Multivariate Statistical Methods in Behavioral Research. McGraw-Hill, New York, 1975.

Currim, I. "Predictive Testing of Consumer Choice Models Not Subject to Independence of Irrelevant Alternatives." Journal of Marketing Research, 1982, 19, 208-222.

Clark, C.E. "The Greatest of a Finite Set of Random Variables." Operations Research, 1961, 9 , 145-162.

Daganzo, C. Multinomial Probit. Academic Press, New York, 1979.

Dawes, R.M. & R. Corrigan. "Linear Models in Decision Making." Psychological Bulletin, 1974, 81, 95-106.

Domencich, T. & D. McFadden. Urban Travel Demand -- A Behavioral Analysis. North Holland, Amsterdam, 1975.

Electric Power Research Institute. Residential Modeling Framework. EPRI EM-5217, May, 1986a.

Electric Power Research Institute. Customer Preference and Behavior: Project Overview. EPRI RP-2671, November, 1986b.

Finkbeiner, C. "Tool Aids Forecasts for Medical Products." Marketing News, Special Marketing Research Issue, January 3, 1986a.

Finkbeiner, C. "Simplified Multinomial Probit." Paper presented at the ORSA/TIMS Marketing Science Conference, Dallas, March, 1986b.

Finkbeiner, C. & P.J. Platz. "Computerized vs. Paper and Pencil Conjoint Methods: A Comparison Study." Paper presented at the Association for Consumer Research Conference, Toronto, October, 1986.

Green, P.E. & Y. Wind. Multiattribute Decisions in Marketing. Dryden, Hinsdale, IL, 1973.

Gumbel, E.J. "Bivariate Logistic Distributions." Journal of the American Statistical Association, 1961, 56, 335-349.

Huber, J. & W. Moore. A Comparison of Alternative Ways to Aggregate Individual Conjoint Analysis. In Educators' Conference Proceedings, L. Landon, ed., Chicago, American Marketing Association, 1979, 64-68.

Johnson, R.M. "Adaptive Conjoint Analysis." Working paper, Sawtooth Software, Inc., Ketchum, Idaho, March 26, 1987.

Kendall, M. & A. Stuart. The Advanced Theory of Statistics, Vol.1, Distribution Theory. Macmillan, New York, 1977.

Lakshmi-Ratan, R.A., Chaiy, S., & J. May. "Mathematical Modelling of Contextual Effects on Individual Choice Behavior: Axiom and Model of Contextual Choice." Working paper, University of Wisconsin, Graduate School of Business, September, 1984.

McFadden, D. "Discrete Response to Unobserved Variables for Which There are Multiple Indicators." Unpublished working paper, Department of Economics, MIT, 1986.

Ramsay, J.O. "Some Asymptotic Thoughts on Factor Analysis." Paper presented at the Psychometric Society Conference, Hamilton, Ontario, August, 1978.

Sawtooth Software. ACA System for Adaptive Conjoint Analysis. Sawtooth Software, Inc., Ketchum, Idaho, 1986.

Thurstone, L.L. "The Prediction of Choice." Psychometrika, 1945, 10, 237-253.

Urban, G.L. & J.R. Hauser. Design and Marketing of New Products. Prentice Hall, Englewood Cliffs, NJ, 1980.

Wiley, J.B. & J.T. Low. "A Monte Carlo Simulation Study of Two Approaches for Aggregating Conjoint Data. Journal of Marketing Research, 1983, 20, 405-416.

APPENDIX

Probit Scale Invariance Theorem

Theorem: If a J-dimensional vector, $\underline{u}_i$, is an observation drawn from a J-variate normal distribution with mean vector, $\underline{m}$, and covariance matrix, $\underline{C}$, then first choice probabilities are invariant under the following transformation of any $\underline{u}_i$: $a_i\underline{1} + b\cdot\underline{u}_i$.

Proof: A first choice probability is defined for dimension j as the probability obtained by integrating the normal density function over the "first choice" region. A "first choice" region is defined for dimension j as the multiple integral over all regions of the density function in which the value of $\underline{u}$ on dimension j exceeds the values on all other dimensions. More formally, the first choice probability for dimension j is:

$$\int_{u_j>u_1} \int_{u_j>u_2} \cdots \int_{u_j=-\infty}^{+\infty} \cdots \int_{u_j>u_J} \Phi(\underline{u};\underline{m},\underline{C})\; d\underline{u}$$

where Φ is the normal density function. Since the effect of parameter b is to uniformly stretch or shrink all dimensions of the J-variate normal distribution, b can be any value without affecting integrals over Φ. Thus, without loss of generality, we can set b to 1. It only remains to show the effect of a_i.

It can be shown that the first choice probability is equivalent to the joint probability of $u_k - u_j > 0$, for all k not equal to j (see Daganzo, 1979, pp.42-44). This can be accomplished without loss of generality by rearranging dimensions of $\underline{u}$ so that dimension j is the first. Then, construct a matrix G as follows:

$$\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ -1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

The desired probability is obtained by integrating Φ over the region $u_k - u_l > 0$, where the parameters of Φ are now $\underline{Gm}$ and $\underline{GCG}'$. Note that $\underline{G}$ has the property that $\underline{G1} = \underline{0}$.

Now suppose every possible observation $\underline{u}_i$ in the population was rescaled by $a_i\underline{1} + b \cdot \underline{u}_i$, where b is set to 1. The new mean vector will be

$$\underline{t} = \underline{m} + x\underline{1}$$

and the new covariance matrix will be

$$\underline{S} = \underline{C} + \underline{c1}' + \underline{1c}' + v\underline{11}'$$

where x is the mean of the a_i and $\underline{c}$ is the covariance between the $\underline{u}_i$ and the a_i. Now to calculate the required probabilities, we need the parameters $\underline{Gt}$ and $\underline{GSG}'$:

$$\begin{aligned} \underline{Gt} &= \underline{Gm} + \underline{0} \\ &= \underline{Gm} \\ \underline{GSG}' &= \underline{GCG}' + \underline{0} + \underline{0} + \underline{0} \\ &= \underline{GCG}' \end{aligned}$$

Thus, the first choice probabilities using $\underline{m}$ and $\underline{C}$ as the parameters are identical to the first choice probabilities using $\underline{t}$ and $\underline{S}$ as the parameters.

This proves the theorem.

As an aside, we note that the Clark approximation carries out the same transformation involving $\underline{G}$ so that it is also invariant under the same rescaling of $\underline{u}_i$.

The desired probability is obtained by integrating Φ over the region $u_k - u_l > 0$, where the parameters of Φ are now $\underline{Gm}$ and $\underline{GCG}'$. Note that $\underline{G}$ has the property that $\underline{G1} = \underline{0}$.

Now suppose every possible observation $\underline{u}_i$ in the population was rescaled by $a_i\underline{1} + b \cdot \underline{u}_i$, where b is set to 1. The new mean vector will be

$$\underline{t} = \underline{m} + x\underline{1}$$

and the new covariance matrix will be

$$\underline{S} = \underline{C} + \underline{c1}' + \underline{1c}' + v\underline{11}'$$

where x is the mean of the a_i and $\underline{c}$ is the covariance between the $\underline{u}_i$ and the a_i. Now to calculate the required probabilities, we need the parameters $\underline{Gt}$ and $\underline{GSG}'$:

$$\begin{aligned} \underline{Gt} &= \underline{Gm} + \underline{0} \\ &= \underline{Gm} \\ \underline{GSG}' &= \underline{GCG}' + \underline{0} + \underline{0} + \underline{0} \\ &= \underline{GCG}' \end{aligned}$$

Thus, the first choice probabilities using $\underline{m}$ and $\underline{C}$ as the parameters are identical to the first choice probabilities using $\underline{t}$ and $\underline{S}$ as the parameters.

This proves the theorem.

As an aside, we note that the Clark approximation carries out the same transformation involving $\underline{G}$ so that it is also invariant under the same rescaling of $\underline{u}_i$.

COMMENT

Richard M. Johnson
Sawtooth Software

I am grateful to Carl Finkbeiner for his clear explanation of differences among popular conjoint simulation models.

However, there is one point on which I would like to comment: Carl concludes that ACA simulation Models 2 and 3 fail to predict shares of preference in holdout choice tasks, and he recommends that these simulation models not be used until they have been modified. Understandably, several ACA users have asked us to respond.

I would like to examine more fully the data that led Carl to conclude as he did, and to state a point of view about validating conjoint predictions. Finally, I will describe two actions we plan to take in the next year that may shed further light on this question.

ACA Models 2 and 3:

All conjoint methods estimate utilities at the "interval" level of measurement. That is, we get a set of numbers within which relative differences are meaningful, but for which the actual scaling is arbitrary. Another way of saying this is that the utilities reflect the relative values of each attribute level; but a set of utilities could be multiplied by any arbitrary positive constant, and have any constant added to them, without affecting this property.

ACA simulation Models 2 and 3 add up the utilities for each product and then transform those values into "shares of preference" using an exponential transformation. An interesting property of this transformation is that the scaling of the utilites has a dramatic effect on the outcome. Here are two hypothetical share of preference calculations for three products:

	Utility X	Transformed Y = exp(X)	Share of Preference 100 * Y / sum(Y)
Product A	.123	1.13	29.7
Product B	.234	1.26	33.2
Product C	.345	1.41	37.1
		----	----
		3.80	100.0

	Utility X	Transformed Y = exp(X)	Share of Preference 100 * Y / sum(Y)
Product A	1.23	3.42	7.5
Product B	2.34	10.38	22.9
Product C	3.45	31.50	69.5
		-----	----
		45.30	99.9

The utilities in the second computation are identical to those in the first, except each has been multiplied by 10. The effect on the resulting shares is dramatic. The large share has been made much larger, and the small one much smaller. By choosing a multiplier of 100 rather than 10 we could have made the results even more extreme.

It is a useful property that by appropriate choice of multiplier we can tune the results to be anywhere from a completely unresponsive model (where every product always has the same share as every other,) to a very responsive model (where the "best" product for each respondent has a share of 100).

ACA does this tuning with the "calibration section" of the questionnaire, which asks "likelihood of buying" questions. Each individual's utilities are scaled to reflect the differences in stated likelihoods of buying a number of concepts designed to vary from extremely attractive to extremely unattractive. An individual whose answers are the same for all concepts has his utilities scaled by a small constant. An individual who gives some concepts much more favorable responses than others has his utilities scaled by a large constant.

What Carl Observed:

ACA was used with a group of 75 respondents to study five attributes of checking accounts. In addition to the ACA questions, each respondent was shown four cards, each describing a checking account on the five attributes, and asked to indicate his first choice. The proportion of times each concept was selected was used as a criterion for assessing the performance of simulation models.

Each respondent also performed another choice task with six cards. Results were similar for both choice tasks, so I will show results only for the smaller one.

	Actual Choice	Max Util	ACA Mod 2	ACA Mod 3	Probit	Multivariate Logistic
Concept A	70.3	64.0	42.7	43.7	57.3	60.6
Concept B	10.8	13.3	20.0	19.9	15.2	12.5
Concept C	16.2	22.7	29.4	29.3	26.8	25.0
Concept D	2.7	0.0	7.9	7.1	0.7	1.9

The important information is contained in the first row of this table. More than 70% of the respondents actually chose Concept A. Although all simulations predict the four concepts in the right rank order, they differ in the size of the share predicted for Concept A. The closest prediction is by the Maximum Utility model, followed by Multivariate Logit, then Probit, and finally the two ACA models.

The ACA models produce much "flatter" distributions of shares than the other models. However, it is interesting to note what happens to the predictions of the ACA models if the utilities are first scaled by multiplicative factors:

Predictions of ACA Model 2 After Scaling of Utilities

	Scaling Factor			
	2.0	4.0	8.0	16.0
	----	----	----	----
Concept A	51.2	57.6	61.5	63.1
Concept B	16.8	14.6	13.4	13.0
Concept C	28.9	27.1	25.0	23.9
Concept D	3.0	0.7	0.1	0.0

As in the artificial example, larger scaling factors cause the results to converge toward the Maximum Utility Model. (Nearly identical results occur for ACA Model 3.) Since the Maximum Utility Model was found to be the best predictor, the apparent failure of the ACA models to predict these holdout concept choices seems to be due to choice of scaling rather than fundamental inadequacy of the models. There is room for disagreement, however, about whether concepts such as these are an appropriate criterion for validating the predictions of a conjoint simulator.

What Should a Conjoint Study Predict?

Everyone would agree that a buyer's actual purchase decision is affected by factors other than his preference at some previous point in time. It is also affected by out-of-stock conditions, point-of-sale promotions, effects of advertising, appetite for variety, and misperceptions of product attributes, to name a few factors. However, not everyone agrees about the extent to which allowances should be made for these factors in conjoint simulators.

One school of thought holds that the conjoint simulator should be concerned with preference alone, and that the most appropriate test is with holdout concepts.

In ACA, however, we have tried to capture some of the other sources of uncertainty with the calibration section of the questionnaire. An individual who responds with extreme likelihoods will have his utilities scaled so that his simulated shares of preference will be extreme.

On the other hand, a respondent who responds with low likelihoods of buying anything, or equal likelihoods of buying everything, will have his utilities scaled so as to produce relatively flat and unresponsive shares of preference.

It has been my own experience that market share is almost always better predicted by a logit model like ACA Models 2 or 3 than by the Maximum Utility model. However, I am not surprised that when respondents are shown concepts described on a few attributes, most of them are able to select the concepts having highest utility for them. Indeed, if one concept dominates the others so clearly as in this 4-concept example, the choice task may be little more than a reading comprehension test.

Plans for the Future:

We plan two actions that we believe will benefit ACA users.

In the near term, we will act on Carl's suggestion and add a capability to ACA simulation Models 2 and 3. This will permit the user to choose a multiplicative factor to be applied to every respondent's utilities to tune the models to fit external criterion data, such as market shares or holdout concepts. Proper use of this new capability will require that some external criterion data be available, perhaps in the nature of additional holdout concepts. Concepts used for this purpose should be as rich and truly representative of real products as possible, and the researcher should avoid including concepts that dominate or are dominated by others. As a rule of thumb, we have generally tried to produce concepts where the ratio of shares between the most and least popular is about three to one.

Over the year ahead, we hope to conduct or participate in a research project to study preferences of buyers in several product categories, exploring the question of which conjoint simulation methods best predict criteria as close as possible to real market choices.

Meanwhile, I ask users of conjoint analysis, and particularly those who use the new adjustable weighting feature in ACA, to share their conclusions about the relative performance of different conjoint simulation models. We hope to be able to report the results of our project at next year's conference.

STATISTICAL SOFTWARE FOR CONJOINT ANALYSIS:
A BRIEF EXAMINATION OF THREE ALTERNATIVE MODELS

Scott M. Smith
Brigham Young University

Gaining Perspective on Conjoint Analysis

The name "conjoint analysis" implies the study of the joint effects of multiple attribute variables (often product characteristics) on the order of preference or choice for a dependent variable (brand). In application, consumers find the numerical rating of brands to be a difficult task. The alternative of simply ranking brands by preference does not provide sufficient information to identify the set of attribute evaluations that influence the assignment of ranks.

Conjoint analysis provides a solution to this measurement problem. The ranking of pairs of brand configurations provides sufficient information to identify the metric scales representative of the decision maker's cognitive evaluations. Next, the numeric scales developed for each attribute are summed to produce a total metric score for each product evaluated. In spite of the power of this methodological innovation, few anticipated that Luce and Tukey's (1964) conceptualization of the additive conjoint model would expand to include the new classes of research methodology that are present today. The purpose of this paper is to identify the key evaluative dimensions for the design, measurement, analysis, and simulation of conjoint problems. During this discussion, the three approaches to modeling conjoint data are identified. These approaches are operationalized in the PC-MDS software package.

Conjoint Analysis Models:

Again tracing the development of conjoint models, these techniques began as non-metric compensatory models. This model type assumes that an abundance in one attribute "compensates" for deficiencies in another attribute. The additive conjoint model further assumes that for the set of objects being evaluated, (1) they are at least weakly ordered (may contain ties), (2) each object is represented as an additive combination of separate utilities that exist for the individual attribute levels, and (3) that the derived evaluation model is intervally scaled and comes as close as possible to recovering the original non-metric (rank order) input data.

The MONANOVA Model

The non-metric additive methodology was initially applied to full factorial designs, where the respondent is required to evaluate objects representing all possible attribute combinations. The MONANOVA (Kruskal, 1965) program represents a model of this genre.

Figure 1
Alternative Conjoint Analysis Methodologies

Stimulus Construction:	Fractional Factorial Design; Full Factorial Design; Paired Attribute
Data Collection:	Full Profile; Two Factors at a Time
Model Type:	Compensatory and Non-Compensatory Models Part Worth Function; Vector Model; Mixed Model; Ideal Point Model;
Measurement Scale:	Rating Scale; Paired Comparisons; Constant Sum
Estimation Procedure:	Metric and Non-Metric Regression; Linear Programming; Non-Metric Trade-off LOGIT; PROBIT
Simulation Analysis:	Maximum Utility; Average Utility (Bradley-Terry-Luce); LOGIT

MONANOVA is based theoretically in the analysis of variance model. The idea is one of finding a monotonic transformation (fm) such that a monotone vector of values fm(δij) can be explained as closely as possible by the additive main effects model. The analysis of the monotone vector fm(δij) produces a 'fitted value' for each cell. For the three-attribute example in Figure 2-A, the fitted value for each cell is a function of the grand mean μ, plus the effects of Price, Seats and MPH ($\mu + \alpha + \beta + \Gamma$). When the fitted values are put into a vector Pa(M), the sum of squares deviation of the fitted and monotone vectors (fm(δij)-Pa(M)) are as small as possible.

The limitations of the MONANOVA model are almost entirely in the size of design that can be handled in the data collection process. For example, the 27 evaluations associated with the three attribute model would be tripled to 81 by the addition of a fourth attribute having three levels.

The Trade-off Model

Trade-off models were introduced as an alternative means of data collection and conjoint analysis (Johnson, 1974). Johnson showed that the sequential evaluation of two attributes at a time provides efficiency in data collection at no cost in reliability. The trade-off analysis approach significantly reduces the size of the evaluation task, both in total magnitude and by limiting the evaluation task to two attributes at a time. While the number of judgments is the same for three-attribute MONANOVA and trade-off models, expansion to four or more attributes creates economies in the total number of separate judgments and in the judgment task (Figure 2-b). Trade-off models produce an estimate of the relative utilities of each of the levels of the two attributes considered. One formulation estimates the utilities for each attribute set as:

$$U_{im} = 1 - L_i \cdot \frac{\sum_{n=1}^{L_j} R_{imjn}}{\sum_{n=1}^{L_j} \sum_{m=1}^{L_i} R_{imjn}}$$

where:

U_{im} = the estimated utility value of attribute i, level m.
R_{imjn} - the rankings for level m of attribute i, level n of attribute j.

Trade-off algorithms are generally based on iterative algorithms which calculate the inconsistencies between the input rankings and the computed utilities. The iterative procedure continues as utility values are changed and re-evaluated for consistency.

OLS Regression Conjoint Model

The Ordinary Least Squares regression approach to conjoint analysis offers a simple, yet robust method of deriving either part-worth, vector, or ideal point models of respondent utilities. The attractiveness of the OLS model is in part a result of the ability to shift from choice rankings to behavioral-intention ratings scales. Further, the ability to implement larger numbers of attributes (through fractional factorial designs) has made this methodology the de facto standard for conjoint analysis (Cattin and Wittink, 1986). The objective of OLS conjoint analysis is to produce a set of part-worth (vector or ideal point) utilities that identify each respondent's preference for each level of a set of product attributes.

The first step in the analysis is to develop either a full or fractional factorial design. A full profile approach is demonstrated in Figure 2-C for our three-attribute example. The use of fractional factorial designs permits the estimation of a parameter for the main effect of each attribute included in the analysis. For example, 3 levels of Price, 3 levels of Passenger Capacity, and 3 levels of Maximum Speed may be evaluated. This design, when analyzed, would produce estimates of the utilities for each of the 9 attribute levels. The utilities are additive and interactions may be considered by the PC-MDS version of the OLS conjoint program.

In application, the OLS model solves for utilities using a dummy matrix of independent variables. Each independent variable indicates the presence or absence of a particular attribute level. The dependent variable is the respondent's evaluation (often a rank value) of one of the (9) products described (by the independent variable).

The OLS Conjoint program produces a file containing the individual attribute level utilities as output for each respondent. The average utilities for each attribute level for all respondents, and the relative importance of each attribute are also produced. In addition, a simulator produces choice share estimates using the first choice model and the average choice (Bradley-Luce-Terry) model (Figure 3). The first choice model adds the respondents' utilities associated with attribute levels defined as making up a new product. After the total utilities for the simulated products are obtained, the product with the highest utility is selected and receives a value of 1. Ties receive a .5 value. After the process is repeated for each respondent, the cumulative "votes" for each product are evaluated as a proportion of votes for each product in the universe.

The Bradley-Terry-Luce model estimates choice probability in a different fashion. The choice probability for a given product is based on the utility for that product divided by the sum of all products in the simulated market.

The PC-MDS version of the OLS Conjoint program contains three distinct operational components (a pre-processor, an analyzer, and a simulator). The components may be used together to form a complete analysis package, or the simulator may be used independently after a file containing subject utilities has been created.

Conjoint Model Reliability

The conjoint models presented thus far have been of the compensatory type, using either metric or non-metric approaches to the estimation of the part-worth values. These OLS and non-metric approaches have to date produced comparable results, though the OLS model provides estimates without risk of solution degeneracy and local optima encountered with the nonmetric methods (Green and Srinivasan, 1978).

The literature published to date indicates little differences in external reliability for the trade-off and full profile models (Jain, Acito, Malhotra and Mahajan, 1979; Leigh, Mackay and Summers, 1984). Acito (1979), however shows that the number of profiles evaluated is positively related to reliability, while an inverse effect occurs for the number of attributes per profile. Huber and Hansen (1986) observed similar results in the area of non-compensatory models, as minimizing affective differences and decreasing dimensional complexity improved the quality of responses.

These findings, while reasonably consistent, are subject to differences that result from the diverse methodologies employed, including the product evaluated, situational contexts, experimental designs, data measures, data collection methods, analytical methodology and measures of reliability employed.

Conjoint Model Validity

Validation of conjoint models has been a topic of limited research. The majority of validation research has been for internal cross validation (holdout samples) rather than the external validation (how well do predicted results compare to market results). Green (1984) in a comparison of the traditional full profile conjoint model with hybrid models concludes that "the traditional conjoint model does at least as well as either the hybrid or self explicated models. Whether the traditional conjoint model's superiority over hybrid models is sufficient to justify the greater data collection complexity is still to be determined."

Summary

It is somewhat surprising that an "absolute best" conjoint model has not emerged in the 25 years since Luce and Tukey introduced the theoretical basis of conjoint analysis. This lack of solution is largely explained by the trade-offs associated with new methodologies that require less respondent information but at the same time make more extensive parameter estimates. Klein (1976) shows that the introduction of non-compensatory models further accentuates this dilemma, as respondents increasingly make "mistakes" in their choice decisions.

The results of the empirical investigations found in the literature suggest that the data type (trade-off or full profile) do not affect the internal validity of the parameter estimates. The full profile does, however allow an increased level of realism associated with a complete product and with interactions between the attribute levels. Further, the full profile design is well suited for telephone-mail-telephone interviewing techniques where visual cue cards are a part of the interview. In contrast, the tradeoff models reduce the complexity of the data collection task, making it applicable for mail questionnaires. Because differences in parameter estimation techniques do not appear to affect results, the most appropriate data collection procedure is contingent upon the specific application under consideration.

References

Acito, F. (1979), "An Investigation of the Reliability of Conjoint Measurement for Various Orthogonal Designs," in Proceedings Southern Marketing Association 1979 Conference, R. Franz, R. Hopkins and A. Toma (eds.), Univ. of Southwestern Louisiana, pp 175-178.

Cattin and Wittink (1986), "Commercial Use of Conjoint Analysis," a paper presented at Association for Consumer Research, Toronto, Ontario, Canada.

Green, P. E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research" Issues and Outlook," Journal of Consumer Research, Vol. 5, (September), pp 103-123.

Green, P. E. (1984), "Hybrid Models for Conjoint Analysis: Expository Review," Journal of Marketing Research, XXI (May, 1984), pp 155-169.

Huber, J., and D. Hansen (1979), "Testing the Impact of Dimensional Complexity and Affective Differences of Paired Concepts in Adaptive Conjoint Analysis," in Advances in Consumer Research, Vol. 12, M. Wallendorf and P. Anderson (eds.), Provo, UT., pp. 154-158.

Jain, A. K., F. Acito, N. K. Malhotra, and V. Mahajan (1979), "A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models," Journal of Marketing Research, XVI (August, 1979), pp 313-322.

Johnson, R. M. "Trade-Off Analysis of Consumer Values (1974)," Journal of Marketing Research, XI (May, 1974), pp 121-127.

Klein, N. (1976), "Assessing Unacceptable Levels in Conjoint Analysis," in Advances in Consumer Research, Vol. 12, M. Wallendorf and P. Anderson (eds.), Provo, UT., pp. 154-158.

Kruskal, J. B. (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data," Journal of the Royal Statistical Society, Series B, 27, pp. 251-263.

Luce, R. D. and J. W. Tukey. "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement," Journal of Mathematical Psychology, 1 (February 1964), pp 1-27.

Figure 1
TABLE OF AVERAGE UTILITIES

ATTRIBUTE/LEVEL:	1	2	3	4
INSTRUMENT PRICE IMPORT.%:32.30	3000 .00	6000 -.53	15000 -2.54	30000 -4.07
TEST PRICE IMPORT.%: 7.81	10.00 .00	20.00 -.13	30.00 -.32	40.00 -.98
SPECIMEN TYPE IMPORT.%: 1.84	NONE .00	URINE -.23		
SAMPLE SIZE IMPORT.%: 5.21	1 ML .00	10 ML -.66		
TEST MENU ITEMS IMPORT.%:46.70	1 TEST .00	2 TEST 2.30	4 TEST 5.05	8 TEST 5.88
REAGENT FORMAT IMPORT.%: .36	DRY .00	LIQUID -.05		
TEST TIME IMPORT.%: 5.78	10 MIN .00	20 MIN .03	30 MIN -.34	40 MIN -.73

Figure 2-A
Full Factorial "MONANOVA" Model Data

	2 Seats			4 Seats			8 Seats		
Top Speed	$7K	$14K	$21K	$7K	$14K	$21K	$7K	$14K	$21K
130 MPH	1	2	5	15	17	22	7	8	11
100 MPH	4	3	6	16	18	23	9	10	12
70 MPH	25	26	27	20	21	24	13	14	19

Figure 2-B
Trade-Off Model Data

	Seating			Price		
	2	4	8	$7K	$14K	$21K
Top Speed						
130 MPH	1	6	4	1	3	5
100 MPH	2	7	3	2	4	6
70 MPH	8	9	5	7	8	9
Seating						
2 Seats				1	2	3
4 Seats				7	8	9
8 Seats				4	5	6
Price						
$7,000						
$14,000						
$21,000						

Figure 2-C
3 x 3 x 3 FRACTIONAL FACTORIAL DESIGN
(9 CARDS)

$7,000 8 Seats 70 MPH	$7,000 4 Seats 130 MPH	$7,000 2 Seats 100 MPH
$14,000 8 Seats 130 MPH	$14,000 4 Seats 100 MPH	$14,000 2 Seats 70 MPH
$21,000 8 Seats 100 MPH	$21,000 4 Seats 70 MPH	$21,000 2 Seats 130 MPH

Figure 3-a
AVERAGE PROBABILITY MODEL (ALL RESPONDENTS)
AVG. PROB. FOR EACH PRODUCT BRADLEY-LUCE-TERRY MODEL

n	1	2	3	4	5	6
47	3.93	3.23	-1.58	-2.71	-.28	-1.58

Figure 3-b
FIRST CHOICE MODEL
FREQUENCY OF RANKINGS AND PERCENTAGES BY PRODUCT

			RANK OF CHOICE					
	ROWTOT	AVGRANK	1	2	3	4	5	6
PROD 1			4.5	10.5	2.0	1.0	4.5	24.5
	47.00	4.36	9.57	22.34	4.26	2.13	9.57	52.13
PROD 2			13.5	5.5	.0	2.0	22.5	3.5
	47.00	3.53	28.72	11.70	.00	4.26	47.87	7.45
PROD 3			3.5	11.5	13.0	10.5	6.5	2.0
	47.00	3.23	7.45	24.47	27.66	22.34	13.83	4.26
PROD 4			14.0	3.5	11.0	9.5	6.0	3.0
	47.00	2.98	29.79	7.45	23.40	20.21	12.77	6.38
PROD 5			7.0	5.5	7.0	15.5	1.0	11.0
	47.00	3.66	14.89	11.70	14.89	32.98	2.13	23.40
PROD 6			3.5	11.5	13.0	10.5	6.5	2.0
	47.00	3.23	7.45	24.47	27.66	22.34	13.83	4.26

PC-MDS 4.1
LIST OF ANALYTICAL PROGRAMS

MULTIDIMENSIONAL SCALING:		**CONJOINT ANALYSIS PROGRAMS:**	
MDPREF	Preference Analysis	MONANOVA	Monotone ANOVA
KYST	Scaling Analysis	TRADEOFF	Attribute Tradeoff
PREFMAP	Preference Mapping	CONJOINT	OLS Analyzer/Simulator
INDSCAL	Individual Differences		
PROFIT	Property Fitting		
CLUSTER ANALYSIS:		**DATA MANIPULATION PROGRAMS:**	
CLUSTER	Howard Harris Cluster	CASE5	Thurstone Case 5
HICLUST	Hierarchical Cluster	DISTRAN	Distance Computation
WARDS	Ward's Method	FMATCH	Cliff Factor Matching
CORRESPONDENCE ANALYSIS:		**AUTOMATIC INTERACTION DETECTOR:**	
CORAN	Lebart,Morineau,Warwick	AID	Survey Research Ctr.
CORRESP	Carroll, Green, Schaffer		Univ. of Pa. Version
MULTIVARIATE ANALYSIS:			
DISCRIM	Discriminant Analysis	REGRESS	Regression Analysis
FACTOR	Factor Analysis	FREQ	Frequency Analysis

SOFTWARE FOR FULL PROFILE CONJOINT ANALYSIS

Steven Herman
Bretton-Clark

We've integrated the time-tested and validated techniques of full profile conjoint analysis into an easy-to-use set of software. Our goals are to make these techniques widely available to research practitioners and to increase researchers' productivity in carrying out this research. Before reviewing details of the software, I'd like to discuss our success in meeting these objectives.

Many of our users are new to conjoint analysis and have become successful practitioners through the use of our software and manuals. Our manuals explain in detail how to design and analyze conjoint studies, and go well beyond instructions for using the software. In addition, the software has been adopted by universities worldwide, and has proven to be a useful teaching tool. Finally, customers rarely run into problems. Few require support, and most questions we receive involve general issues in conjoint analysis rather than specific software issues.

Besides being easy to use, the software is also powerful enough to meet the needs of many of the largest and most experienced research companies, including many at this conference. Now let me describe some its capabilities.

First I'll describe Conjoint Designer and Conjoint Analyzer, which comprise an integrated system for full profile analysis. Then I'll introduce Simgraf, our new simulation package. Finally, I'll present some comparisons between the full profile approach and Sawtooth's ACA approach, and present some conjectures about future developments in computerized conjoint analysis.

Conjoint Designer

The toughest part of conducting a conjoint study is the design stage. Full profile studies generally use experimental designs called orthogonal arrays or orthogonal main effects plans. These designs have important advantages. First, they allow you to gather data about a large number of product attributes with a relatively small number of questions. Second, from a statistical perspective, orthogonal designs are most efficient. Finally, as Professor Huber pointed out at last year's Sawtooth Conference, orthogonal arrays make full-profile studies resistant to various types of "misbehavior" by the respondent.

In the past the use of these designs was limited to a relatively small number of elite researchers. Even those researchers had to invest a good deal of time and trouble to develop these designs, and still more to customize them. In addition, with manual techniques it's all too easy to make a mistake which can jeopardize a study.

We introduced Conjoint Designer in 1985 to solve these problems. It produces these orthogonal arrays, automates the associated tasks involved in producing a finished design, and eliminates errors.

The first step in using Conjoint Designer is to specify the number of features in your study. For each feature, such as price, you enter its name and the number of levels -- in this case, the number of prices to be included. The program determines the smallest orthogonal array that will meet your needs and tells you how many cards you will need. You can then add or delete features, or change the number of levels of any feature, and the program will tell you the new number of cards required, up to a maximum of 81 cards.

After you settle on the final number of features and levels, the program prompts you for the name of each level of each feature (for example, the particular prices or brands under study). It then generates the design and presents you with a number of options.

First, you can view any or all of the cards on your screen, or print them. You can also create a "card image file" -- a text file you can use with your word processor, desktop publishing software, or neighborhood typesetter so that you can print the cards with the fonts and layout of your choice. You can also create a design file for use with Conjoint Analyzer or another analysis program, such as that of Professor Smith.

Sometimes a design results in unrealistic cards. For example, a card may describe a product with many feature options and a very low price. Conjoint Designer gives you several ways to handle this problem. First, you can randomize the design. Second, you can generate a new design with the same specifications. Finally, you can change the card. The last option will also cause the design to deviate from orthogonality, so we allow you to view the correlation matrix for the design and assess the damage. The correlation matrix can also be used to show that the original design was orthogonal.

Conjoint Designer also allows you to create holdout cards for your study. These are used to validate your results, and also to gather data on particular products of interest. For example, if the actual design does not describe some of the current products in the market, you can create holdout cards that do.

More sophisticated researchers can also create incomplete block designs, designs which can estimate interaction effects between respondents, and, much of the time, Paretooptimal designs. These topics are beyond our current scope. Conjoint Designer also offers convenience features such as a built-in directory search capability which can locate any design file, read it back into the program, and modify it if necessary.

Conjoint Designer is menu-driven, easy to use, and very fast. In most cases, the time it takes to produce a finished design is determined by your speed in reading and reviewing the cards.

Conjoint Analyzer

Conjoint Analyzer picks up where Conjoint Designer leaves off, namely, after respondents have completed the conjoint task. It can handle data from either card sorts or rating scales. We recommend the use of rating scales, as they are faster to administer and usually result in higher quality data. Conjoint Analyzer calculates the utility function for each respondent, helps you clean the data, produces high resolution graphs of the results, and performs two types of market simulations. It also performs some secondary analyses not addressed here.

Conjoint Analyzer lets you select from three types of utility models for the features in your study -- the partworth model, the vector model, and the ideal-point model.

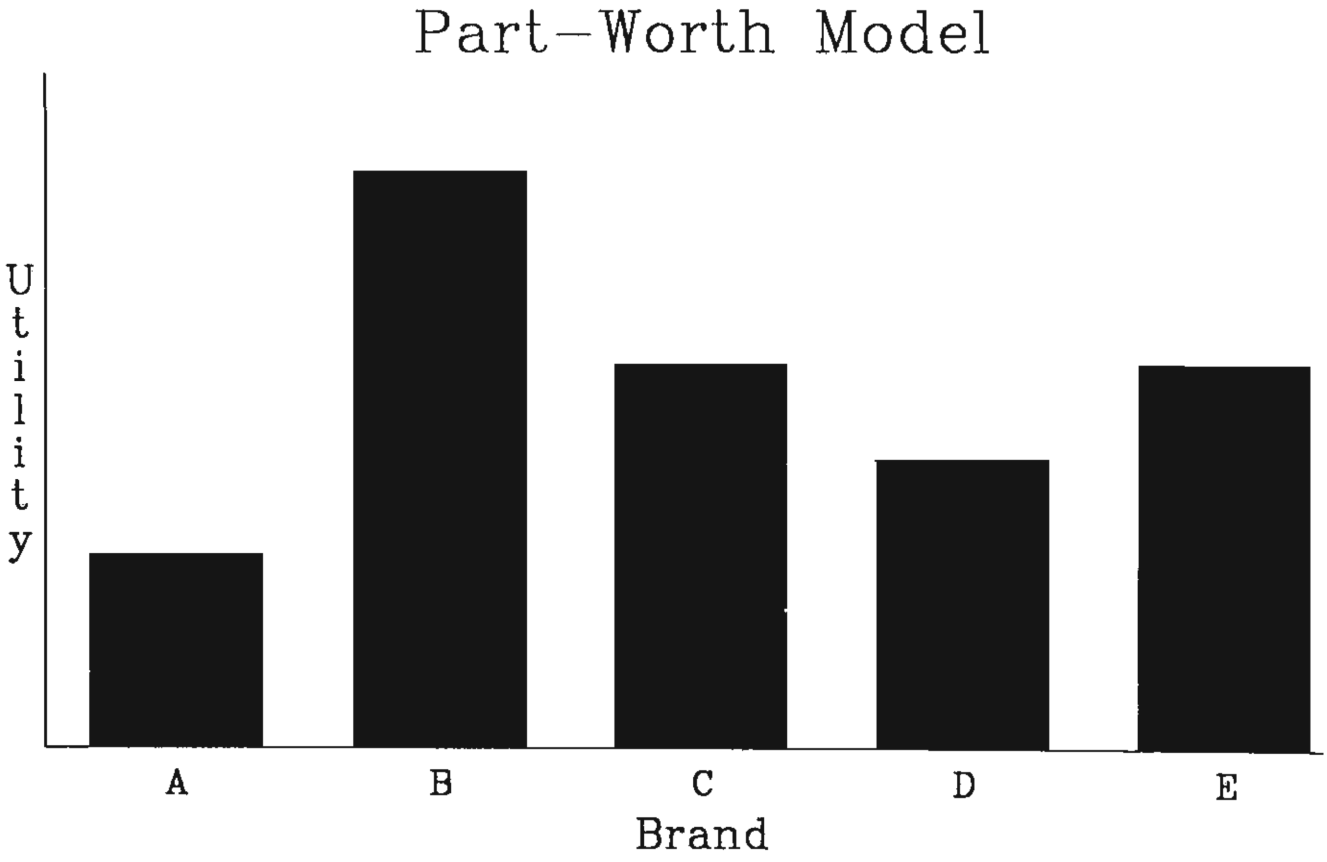


Figure 1

The part-worth model is appropriate for qualitative features like Brand. For each brand in the study, you get a separate estimate of its utility. Most conjoint software only implements this model. However, there's a good deal of evidence that vector and ideal-point models can produce more accurate results with quantitative or continuous features.

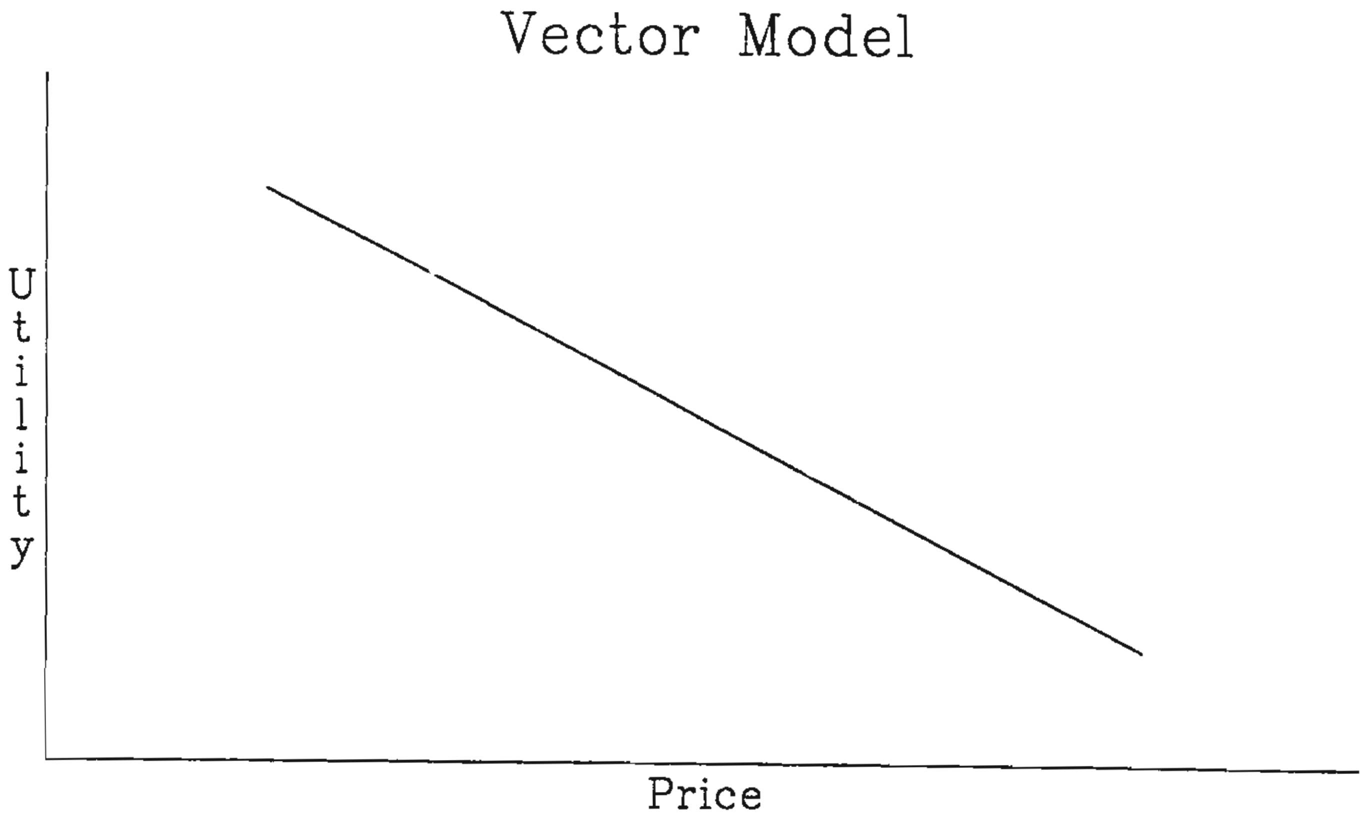


Figure 2

The vector model assumes that a consumer's utility for a product is linearly related to a product feature. For example, increases in price usually result in linear decreases in utility.

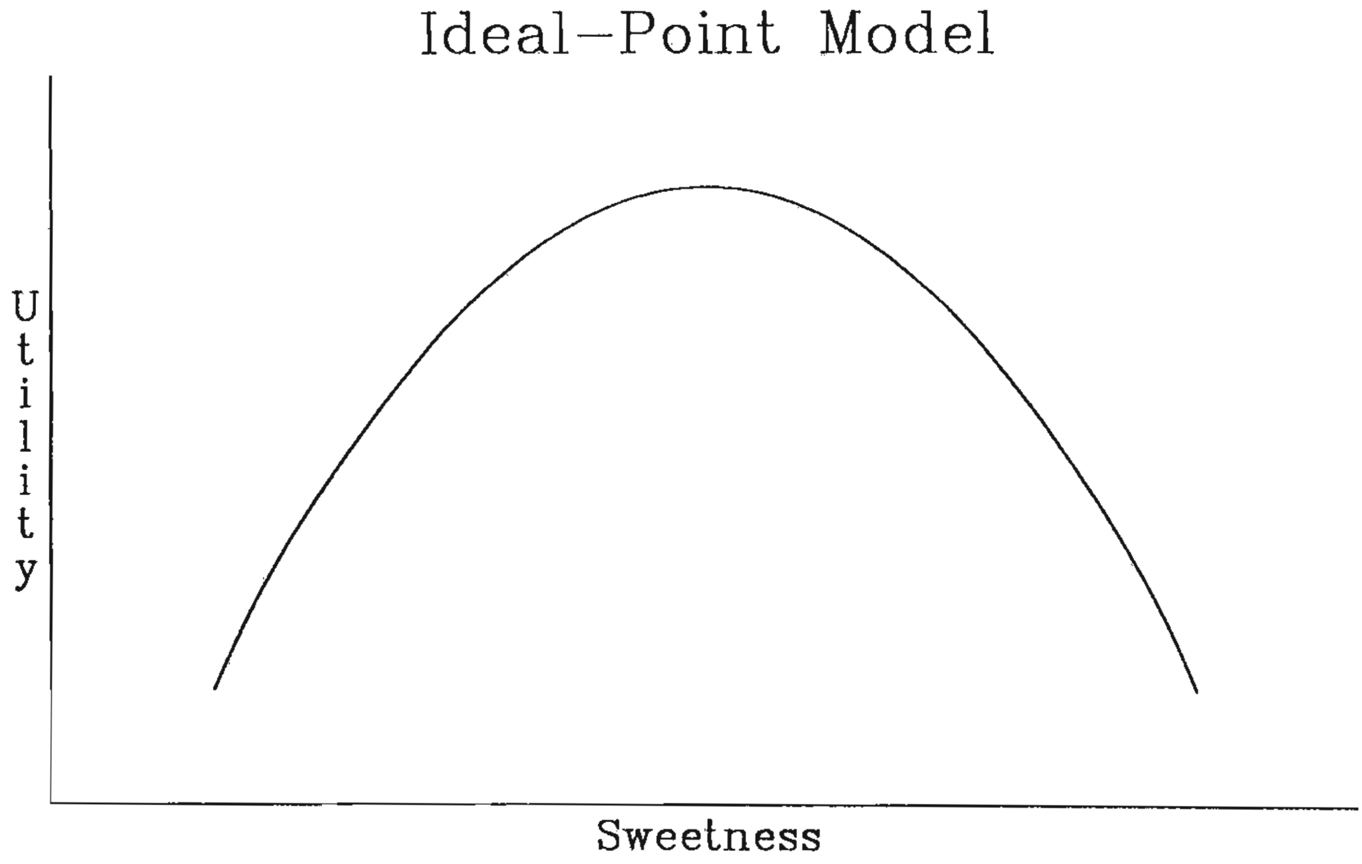


Figure 3

The ideal-point model is appropriate when there is a strong curvilinear relation between a feature and a consumer's utility for a product. For example, most people prefer a specific amount of sweetener in a cola, and dislike a drink with too much or too little sweetener. Most products do not require the ideal-point model.

Use of the vector and ideal-point models will not only increase the accuracy of your results, but will often allow you to reduce the size of your design as well.

After you select the models, you are asked whether any of the qualitative features are "ordered" in terms of consumer preference. For example, low prices are preferred to high ones and more feature functionality is preferred to less. This information is used to clean the data.

Data cleaning is a topic of great practical importance which has received very little attention. Many researchers use a measure such as R squared to determine how consistent a respondent has been. Unfortunately, respondents often behave consistently, but incorrectly. For example, some uncooperative respondents give each card the same (typically low) rating. Other respondents "reverse" the use of the rating scale, mistakenly using low numbers for good products, or vice versa. These kinds of errors can seriously bias your results, particularly at the market segment level.

Conjoint Analyzer employs a simple but unique data cleaning procedure that catches these errors and has some additional benefits as well. The program measures the number of features a respondent reverses, which provides a simple measure of the validity of his data.

After you enter the data about ordered features, the program calculates the utility function for each respondent, counts the number of ordered features he reverses, and files the results in a "utility file." This happens pretty quickly -machines with math coprocessors will generally process several respondents a second.

Following this, you can screen the data for respondents with the highest number of reversals. Reversed features are highlighted, and the relative importance of the feature is also displayed. If most features are reversed, you can "reverse" the respondent's data and thus correct the data. If only one or two important features are reversed, you can mark the data as bad and remove it from further consideration.

The measure of number of reversals also provides another indication of the importance of quantitative models like the vector model. Using these models generally results in a dramatic decrease in the number of reversals. Results for all features are improved, not just the quantitative ones.

In some studies we've done, these models have produced valid results for each individual respondent. The client, who was familiar with each of the large accounts in the study, felt that each respondent's feature preferences were accurately reflected by the utility function -- even seemingly aberrant ones.

After the data are cleaned, you can get high resolution graphs of the overall results. Figure 4 provides an example of the graphic output. These graphs help communicate your results quickly and effectively, and require no effort on your part.

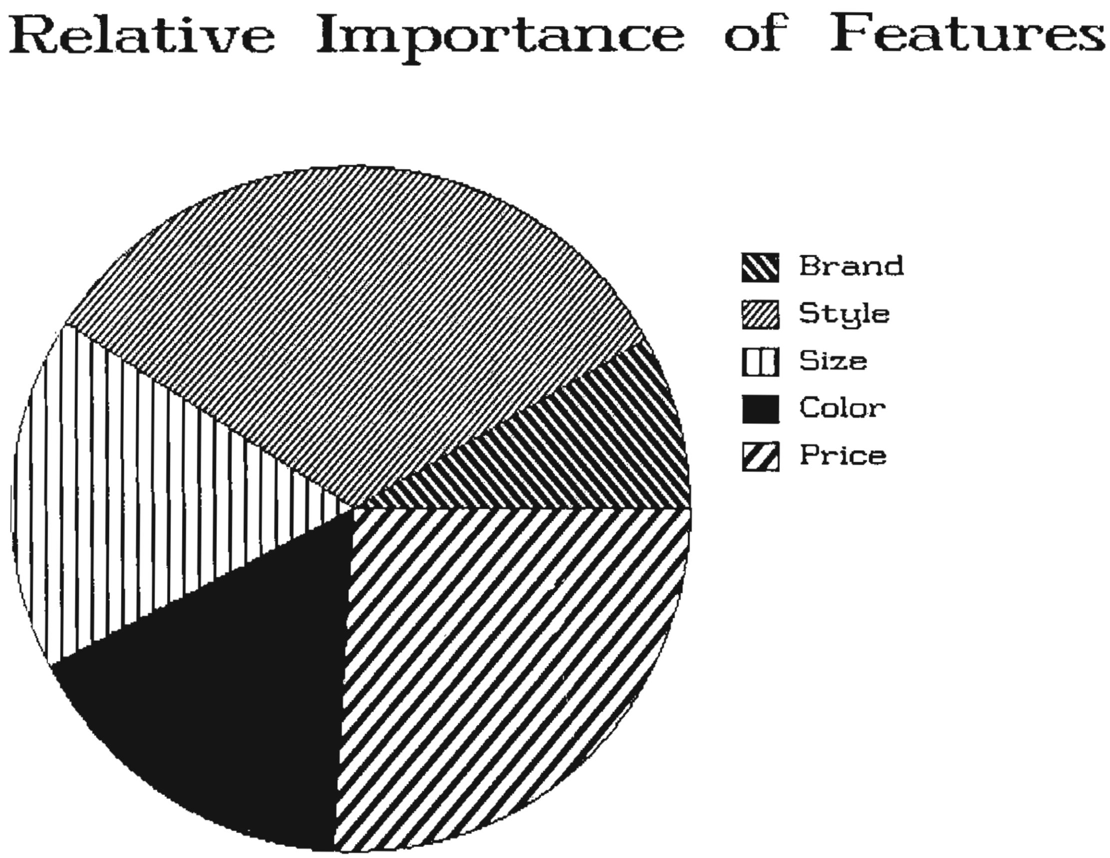


Figure 4

Conjoint Analyzer also includes two types of market simulation models. The First Choice Model assumes that a respondent will select the product with the highest utility to him. The BTL, or Bradley-Terry-Luce model, assumes that the probability a respondent will choose a product is equal to the utility of that product divided by the sum of the utilities of all products in the choice set -- that is, its share of the utilities. These models are pretty standard in the field. Some researchers use a variant of the BTL model involving a logit transform, but results tend to be similar.

The simulator accepts up to twenty products at a time, and makes it easy to add, drop, or modify any product. The output of the simulator consists of "shares" for each of the products. You can also save the results for each individual respondent, which is useful for specialized analyses.

The Conjoint Analyzer package also comes with two other programs. One is a full screen data entry and editing program, which is especially useful for compiling results of focus groups and pilot tests, and permits very fast turnaround of results.

Conjoint Analyzer also includes a program which allows you to modify the utility file in a variety of ways. You can increase or decrease the importance of any feature to conduct alternative simulation scenarios, you can create "export" files which are compatible with general statistical programs, or you can output just the data you need for conducting a cluster analysis.

Simgraf

We recently developed a new stand-alone simulation program called Simgraf, which has several important benefits. First, it incorporates demographic or background variables, so you can assess the impacts of new products on individual market segments, or develop product line extensions to appeal to specific segments.

Second, it minimizes some of the problems associated with the traditional simulation models incorporated in Conjoint Analyzer and other software packages. In general, the First Choice Model tends to exaggerate the share of both popular and unpopular products. The BTL and logit models, on the other hand, tend to underestimate the share of popular products and overestimate the share of unpopular products.

Simgraf incorporates a generalization of the BTL model which allows you to control how "extreme" the model forecasts are. Using low values for the model results in performance like the BTL model, while high values of the model approximate the performance of the First Choice Model. You can tune the model continuously between these two extremes, and thereby calibrate the model to historical data.

We've also attacked this problem by modifying the First Choice Model to include a "choice threshold" parameter. Normally, the model assumes that even the slightest difference in utility between two products is sufficient to cause the consumer to select the product with the highest utility. However, if this difference is less than the choice threshold, the revised model assumes that there is a tie between the products. This procedure makes intuitive sense, and generally produces good results as well.

In addition, Simgraf contains another option, which permits you to calibrate the output of the simulator to known market data. Normally, when you enter descriptions of the current products in the marketplace into a simulator, you don't get back the current market shares of these products. While there are good reasons for this -- for example, a dynamically changing market, omitted variables such as promotions and advertising, and so on -- researchers often want to measure new product introductions from today's baseline. Simgraf contains a "Bayesian" procedure to accomplish this.

As the name suggests, Simgraf produces graphic output as well as numeric output. These graphs, consisting of color-coded stacked bar charts, simplify the interpretation of the data. This is especially true for analyses conducted at the market segment level. Each segment is represented by a separate chart, and these charts are presented simultaneously. Segment level differences are vividly portrayed, and are hard to miss.

In addition, Simgraf keeps a running record of all your simulations in a "log" file, so you don't need to take careful notes, or keep track of multiple files. You can also save results for each individual respondent. All inputs to Simgraf are simple text or ASCII files. Users of Conjoint Designer and Conjoint Analyzer will already have the necessary input files, but users of other conjoint software can easily create these files as well.

Simgraf can also handle larger problems than the simulator available in Conjoint Analyzer. It can handle up to 60 products at a time, and an unlimited number of product features and demographic variables. Product features and demographic variables can each have up to 20 levels.

In addition to introducing enhanced simulation models, demographic analyses, and larger capacity, Simgraf has other benefits. First, it comes with a transferable license, so you can leave a copy with a client following the completion of a study. This way the client can evaluate any new ideas or developments that may arise, and can extract the fullest benefits from the study. In addition, by giving the client the ability to run simulations, you are freed to concentrate on your current studies, rather than maintaining a large inventory of "active" studies.

Comparisons with ACA

We're frequently asked to compare our software to ACA. Since the approaches differ on so many dimensions, and since relatively little research has been published on ACA, this is not easy.

The most obvious difference between our implementation of the full profile approach and the ACA approach involves the way tasks are administered. ACA requires computerized interviewing, while our software does not include this option. Our software is in fact neutral with respect to mode of administration.

The full-profile approach can be administered in a wide variety of ways, including personal interviews, mail surveys, combined mail and telephone surveys, and focus groups. Since focus groups have not traditionally been used with conjoint, and since they provide a quick and relatively inexpensive way of collecting full profile data, a few words are in order.

The first part of the group session involves discussions of the product, its features, uses, and related topics. In the second part of the group, we explain the conjoint task, pass out decks of cards for respondents to rate, and then collect the data. Using our data entry program, you can create a data set, analyze it, and prepare a top line report within a couple of hours.

In general, if you want computerized interviews, ACA is the natural choice. If you don't, our conjoint software is. I should also note that several studies have successfully used computerized collection of full profile data. However, this involves some programming expertise.

Mode of fieldwork, however, is only one consideration. Finkbeiner and Platz of National Analysts compared ACA and the full profile approach (administered by paper and pencil) in terms of speed of administration. They found that the full profile approach was substantially faster to administer.

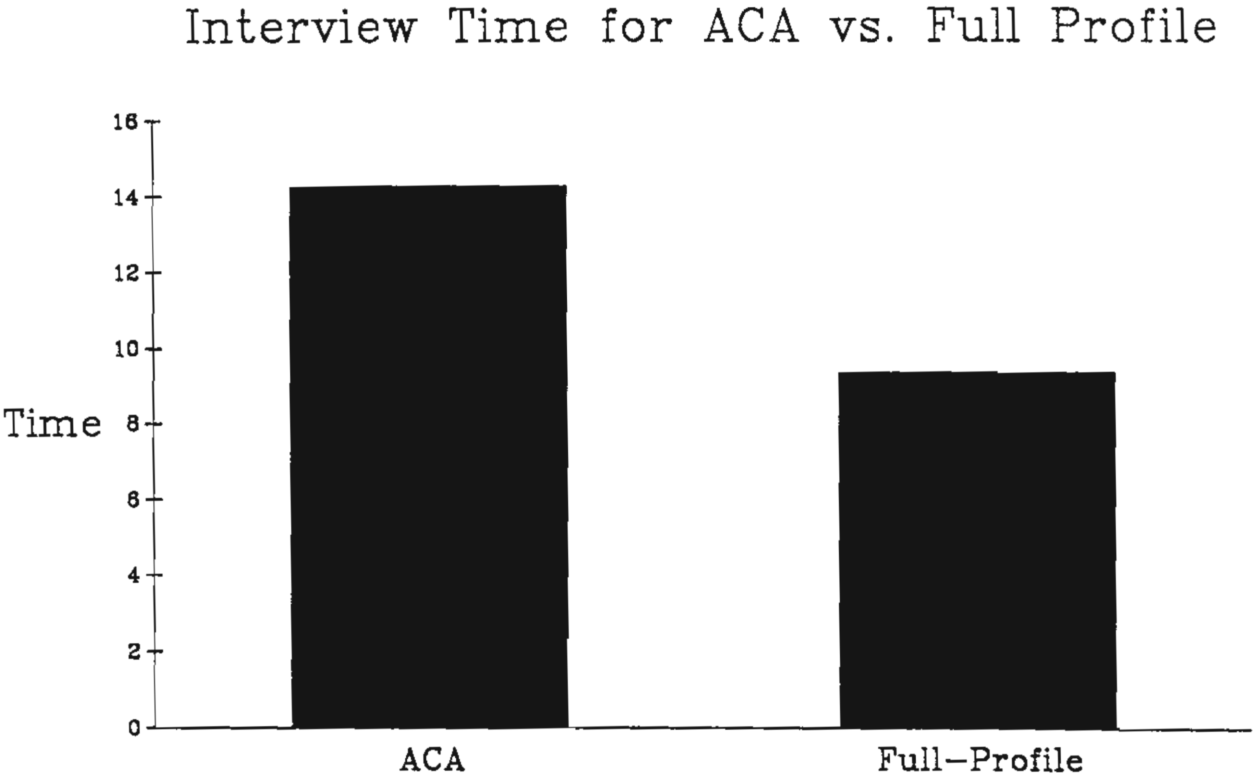


Figure 5

The full profile approach was over 35% faster than ACA. (Although they also conclude that with more than 32 full profile cards ACA might be expected to have a timing advantage.) Increased speed has important practical consequences. You can use the time to gather additional information from the interview, reduce fieldwork costs, or both. However, these results will vary with the size of design, whether pictorial or verbal stimuli are used, and a host of other factors. Further research is needed.

Other considerations involve statistical modeling. Conjoint Analyzer currently implements vector and ideal-point models, both of which have been shown to yield more accurate estimates of utility functions than the simple part-worth model. ACA is limited to the part-worth model.

Other modeling considerations are potentially important as well. For example, interactions between product features are important in areas such as sensory testing and package design. The full profile approach can handle interactions, while ACA cannot. In addition, Hagerty has developed optimal weighting techniques for improving the predictive power of conjoint analysis. The method assumes that each respondent sees the same cards, and therefore is incompatible with ACA. However, our implementation of the full profile approach does not currently address these factors adequately, so the issue may be moot.

Ideally, we would like validity data for these methods under a variety of conditions. Due to the newness of ACA, I know of only three studies that are relevant. First, Finkbeiner and Platz found the methods did equally well in predicting results for holdout data. However, the authors relied exclusively on the part-worth model, which may have handicapped the full profile approach.

Second, two studies have addressed the validity of allowing respondents to eliminate "unacceptable" levels from the conjoint task. This is an optional feature of ACA. Results of these studies suggest that this technique lowers predictive validity, and should be avoided.

ACA and the full-profile approach also differ in the way product descriptions are presented to the respondent. In the full profile approach, respondents are presented with complete products to evaluate, while in the ACA approach, respondents see pairs of partially described products. On the one hand, the full profile approach provides a more realistic task. On the other hand, with a large number of features, the ACA approach may provide a more manageable task.

I'd like to add a few words about conducting full profile studies with large numbers of features. First, it is often possible to do this using bridging techniques, and we will be introducing bridging software later this year. However, this sidesteps the real issue, which involves creating conjoint tasks which closely approximate real world purchasing situations.

For example, with some technical products, we've been able to use the full profile approach with over fifteen features. The respondents were accustomed to looking at spec sheets containing large numbers of product features, and the conjoint task seemed natural to them.

Similarly, pictures often provide a way of incorporating large numbers of features using full profile conjoint. For example, Page and Rosenbaum recently described a full profile study of food processors that Sunbeam conducted. Despite the fact that the study employed twelve product features and used the more difficult card sort procedure, respondents had no difficulty.

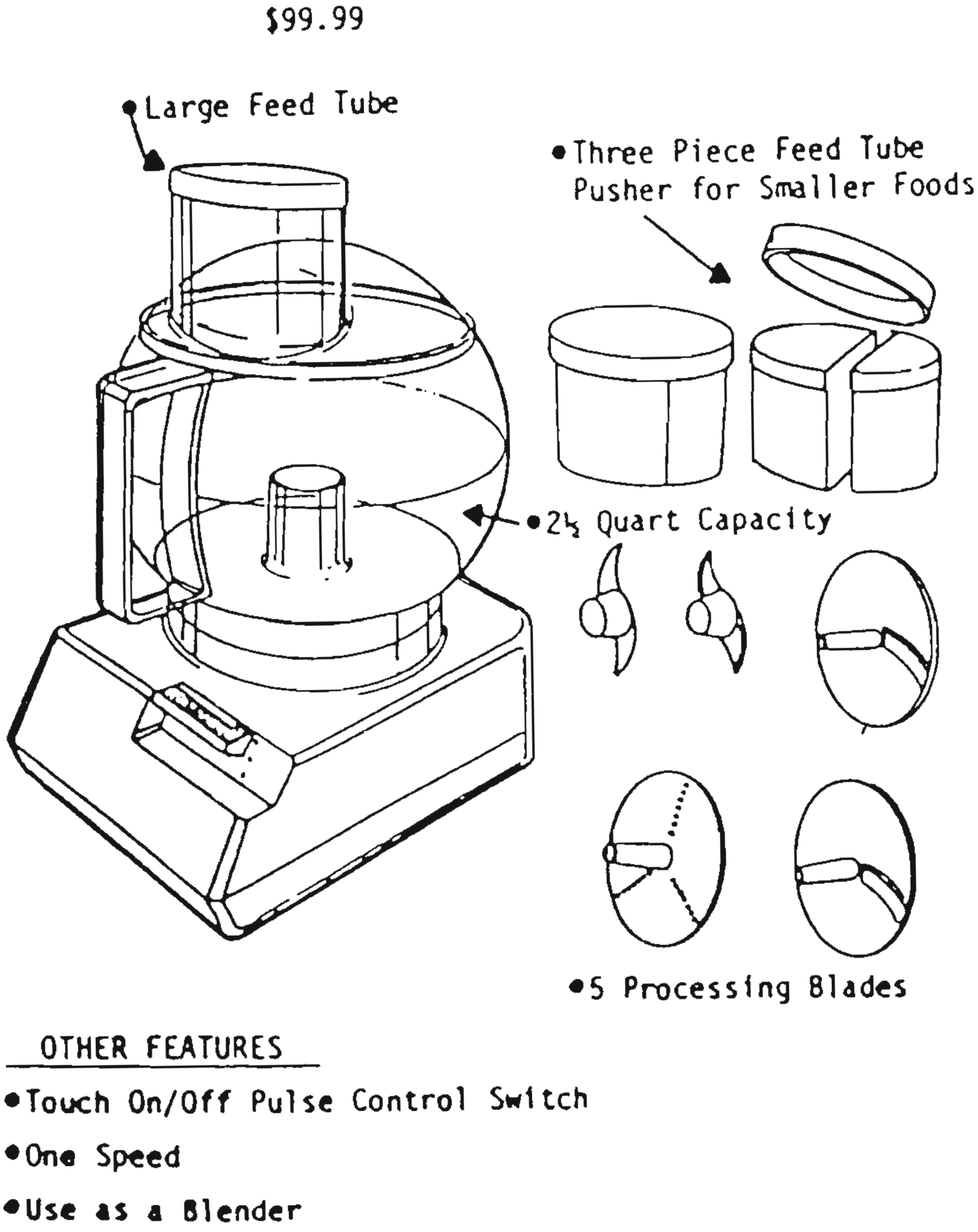


Figure 6

In short, I feel there is no specific limitation to the number of features that can be handled with the full profile approach, <u>as long as the profiles are presented or described in a natural way</u>. To the extent that this is not done, respondents need to perform complex mental transformations to map the stimulus props into their normal framework, and this may create information overload.

Future Developments in Computerized Conjoint Analysis

Computers will play a major role in creating these realistic conjoint tasks. Today, for example, computers can be used to present the kinds of line art that Sunbeam used for food processors. Soon it will be economical to present respondents with realistic pictures in full color. Within a few years, PCs will also be capable of much more sophisticated tasks.

For example, respondents will be able to view a lifelike image of a car, and then rotate that image to inspect the car from any angle. They'll be able to zoom in on details, including the dashboard and interior. They may even be able to hear the sound of the door slamming. This capability will permit large numbers of features to be tested without straining the respondent.

In fact, the richness of the possibilities will create the need for new research techniques. Not only will respondents be able to view realistic products, but they'll be able to change the color, design, and various other product attributes to suit their taste. Current approaches to conjoint do not permit this.

Similarly, the computer can store vast amounts of data for each product. Respondents will be able to ask about any product feature that interests them. While some respondents may be content with viewing a car, others may want detailed specifications about its power train and chassis. Or they may want to see this information for some products but not for others. Current research techniques simply can't handle these requirements.

I predict that these developments will lead to important advances in our ability to understand consumer preferences and purchasing behavior. Developing and implementing these techniques will provide exciting opportunities for researchers.

References

Finkbeiner, C.T., & Platz, P.J. (1986) "Computerized vs. Paper and Pencil Conjoint Methods: A Comparison Study." In M. Wallendorf and P. Anderson, eds., Advances in Consumer Research, 14, Provo, UT: Association for Consumer Research.

Green, P.E., Krieger, A.M., & Bansal, P. (1987) "Completely Unacceptable Levels in Conjoint Analysis: A Cautionary Note." Wharton School, University of Pennsylvania

Hagerty, M.R. (1985) "Improving the Predictive Validity of Conjoint Analysis," Journal of Marketing Research, 23, 16884.

Klein, N.M. (1986) "Assessing Unacceptable Attribute Levels in Conjoint Analysis," In M. Wallendorf and P. Anderson, eds., Advances in Consumer Research, 14, Provo, UT: Association for Consumer Research.

Page, A.L. & Rosenbaum, H. (1987) "Redesigning Product Lines with Conjoint Analysis: How Sunbeam does It," J. Prod. Innov. Manag, 4, 120-137

CONJOINT ANALYSIS BY TELEPHONE

Brent Stahl
MORI Research

I will discuss some of the practical problems encountered during a project in which conjoint data were collected over the telephone.

A PROBLEM: COSTS

Those who have done conjoint analysis projects or have thought about them know they can be quite expensive. From a cost standpoint, the ideal situation would be to use personal computers as interviewing stations at a convention of computer-literate professionals to probe a subject of great interest to them. Often, however, the sample must be drawn from the general population or from a geographically-dispersed special population. Interviewing costs can be substantial if the sample is large and respondents have to be recruited to come to a central location and compensated for their time.

A couple of years ago we at MORI Research had such a project, involving two special populations that had to be further stratified by geography and by income. We were testing a complicated new service with many new features, so it was especially important that respondents become familiar with the concepts before plunging into the conjoint interview.

We wanted to use the Sawtooth ACA (Adaptive Conjoint Analysis) conjoint software on this project for several reasons. One was the relative speed in data analysis it affords. Another was that the questionnaire (conjoint and other questions) would be lengthy (45 minutes), and keeping respondents involved was a potential problem. The novelty of a computer-driven interview and the respondent's awareness that ACA is "customizing" the interview would be advantages here. Another reason was that the client was enthusiastic about ACA--the research manager there had used it to help sell the idea to the brand manager.

These considerations all argued for recruiting respondents to come to central locations. We sat them in front of rented computers which had a lengthy conventional questionnaire (using Ci2, Computer Interactive Interviewing System) in addition to a conjoint questionnaire dealing with 13 attributes.

That project went well--we finished on time and, thankfully, on budget, and the client was pleased with the substantive results. The client came back to us the next year with another project for which conjoint would be appropriate, but their budget would be much smaller than that of the earlier study. It would have to be a telephone study, we were told, with or without conjoint analysis. They asked, "Can you do conjoint over the telephone?"

My boss is a very enthusiastic, "can-do" kind of person, which is highly desirable for one who sells professional services. I'm more of a cautious type and sometimes play the role of Scotty to her Captain Kirk. In this case, when she asked me if we could manage this, I was skeptical that we could do the interviewing by telephone.

The subject was credit cards, and the client wanted to know as much as possible about tradeoffs among such things as annual fees, interest rates, and some "enhancements" that many credit cards offer, such as flight insurance.

While the subject was not as complicated as the earlier study, the problem remained of how to make the conjoint choices understandable to respondents using only oral descriptions. In particular, if we were to use the ACA program, we would be presenting respondents series of paired comparison choices. For each choice, we would ask which of the two product descriptions they preferred and by how much (using a 9-point scale). Since each product description would have to include at least two characteristics ("attribute levels"), our respondents would have to keep track of mentally at least four concepts at a time, as well as a preference scale:

[FIGURE 1]

```
          PRODUCT 1              PRODUCT 2
      +---------------------------------------+
      |   Feature A1      |    Feature A2     |
      |                   |                   |
      |   Feature B2      |    Feature B1     |
      +---------------------------------------+
   1     2     3     4     5     6     7     8     9

STRONGLY                  NO                   STRONGLY
PREFER 1              PREFERENCE               PREFER 2
```

This seemed to be a lot to ask, especially since we would have to ask respondents to make 15 or so of these choices and to be consistent about it. For this reason we decided early on that, rather than attempting to complete the interview in one phone call, we would probably need some kind of phone-mail design.

An obvious question at this point is why use the ACA version of conjoint analysis, as opposed to mailing cards with full-profile descriptions to rate or rank? We wanted to use ACA for essentially the same reasons mentioned earlier--ease of data analysis, client familiarity with the program, and the increased respondent involvement when the computer seems to be customizing the questionnaire.

DEALING WITH PAIRED COMPARISONS

We still faced the problem of how to simulate the appearance of the paired comparison screen that appears in the ACA interview. An initial thought was that we could send respondents a "game board" with a scale across the top, along with two decks of numbered cards. (Each card would represent an attribute level.) At the interviewer's instruction, the respondent would pick out various card combinations to put on the game board and then rate them. This did not seem workable to me. Sorting through cards would become tiresome very quickly for respondents; and frankly, I didn't want to deal with the printing, sorting and stuffing into envelopes of the 20,000 or so cards this would require.

We conducted some focus groups of credit cardholders in order to narrow our list of attributes and to try out some data collection ideas. We found, incidentally, that the ACA program was useful in encouraging discussions of tradeoffs in focus groups. (We used hardware that transfers the computer screen image onto a larger screen.)

We also found that respondents varied considerably in their ability to keep track of concepts and the paired comparison scale. The more sophisticated credit cardholders were very familiar with the concepts we were testing and could state strong opinions without the help of props. Other members of the groups could manage if they had a workbook to refer to that listed the various product characteristics we were dealing with. This encouraged us to go with a phone-mail-phone design for the main study, since we would be using a general sample of credit cardholders.

We started the survey with a 15-minute telephone interview that focused on usage patterns and reactions to various credit card enhancements. We then sent a packet to respondents that included a 4-page written questionnaire about financial attitudes and demographics. The respondent filled this out and saved it for the return call.

The packet also included color-coded pages showing the different scales used during the initial parts of the ACA interview, and a page listing the attributes and levels and descriptions. Respondents used this list for reference during Ci2 questions that were asked just before the ACA section of the interview, and during the first part of the ACA interview. (Before starting the conjoint section, we asked questions designed to familiarize respondents with each of the product concepts and to get them used to referring to the concept list on the page.)

We also included a 20-page booklet with all of the attribute levels listed twice in facing columns on each page. At the top of each page was the 1-to-9 tradeoff scale. The levels, or product features being tested, were numbered from 1 to 21.

These numbers allowed the interviewer to refer to them when instructing the respondent to circle the designated features for each 2x2 tradeoff.

The interviewer would say, "Please circle items 3 and 9 on the left side of the page and items 4 and 10 on the right side. Now, think about just these features and tell me which side you prefer, using the scale at the top of the page."

(Whether respondents could concentrate just on the circled features was an important question that we tested in focus groups.)

The front page of the book was marked to illustrate the task. (See Figure 2, which illustrates the general appearance of the book using automobile features.) Two levels in each column were circled, as was the "6" on the scale, plus an annotation saying that this marking indicates a slight preference for the "card" described in the right column. As the respondent made choices, the interviewer would feed back the information to verify that they were communicating. In most cases the interviews picked up speed as respondents got the hang of it.

In order to keep things simple, we restricted the paired comparisons to 2x2 tradeoffs. In this we were following research reported by Joel Huber at this conference last year which indicated little is gained by including larger attribute sets in the tradeoff exercise. We limited the number of paired comparisons to a maximum of 16 (most interviews required 13 or 14). (The program ends the interview when it thinks the utility values have stabilized.)

We used the back pages of the workbook for the "calibration concept" section of the interview. On these questions respondents react to full-profile descriptions of product concepts that range from least desirable to most desirable, based on their previous answers. The respondents were again asked to circle product features that defined particular credit cards, and then to indicate how likely they would be to acquire each. (This exercise functions partly as a consistency check on the respondent.)

By a more or less happy coincidence we had 8 attributes and 21 attribute levels in our final version of the questionnaire. We wanted the telephone interview to last no more than 20 minutes, and this number of attributes and levels was about the limit for that time frame. It was also about the limit for the number of items that would fit on the workbook pages with adequate annotations and still be easily readable.

(The interview could have been shortened by asking which product features were "unacceptable" to the respondent. However, we didn't want to do that since "unacceptable" features can cause complications in data analysis.)

DATA COLLECTION

The fieldwork for the study went well. We wanted to get conjoint data from at least 75% of those going through the first phase of the study. We ended up with 81% and could probably have reached 85% or so had we not cut off interviewing to meet the report deadline. To encourage responses we used two rather small incentives--a $2 bill enclosed with the questionnaire and a chance to win an electronic gadget or its monetary equivalent--$300. This recovery rate is very similar to others we have obtained using phone-mail-phone designs with long and demanding questionnaires.

Those who did not complete the second part did not differ appreciably in their demographics from those who did. They were a bit less interested in and reliant upon credit cards. This did not bother us too much, because this group would also be less likely to change from its current credit cards or usage patterns.

The second interviews lasted about 21 minutes on average, with some elderly respondents taking longer. The longest lasted almost two hours, but it turned out the respondent wasn't confused, just slow!

I wish we had included direct questions asking respondents how well they understood what was going on. The interviewers did report that with only a few exceptions (mainly elderly), respondents seemed to follow things pretty well. One of our best interviewers said that about 75% had no problems, about 20% had some initial difficulties but did better as the interview proceeded, and about 5% never did understand what we were doing.

This estimate may have been pretty close to the mark. Termination rates were lower than I expected--we had ten during the conjoint interview, and three of these were initiated by the interviewer. The proportion of respondents dropped from analysis because of data inconsistencies (as indicated by low calibration concept correlations (< .75)) was only 4.5%. This rate was actually lower than the previous (in-person) conjoint study we had done for this client.

I expected this rate to be higher, if anything, since we were doing this over the telephone. After further reflection, this seems to be the result of two factors. One was the relative simplicity of the concepts we were testing. The other may be due to the questions being read to the respondent rather than being self-administered. The calibration concept product descriptions are offered in order of likely preference, given the respondent's previous answers, and respondents are told they will probably like the first product least and the last product most. Having this read to you may discourage inconsistent answers more than simply reading the instruction. (However, we instructed interviewers <u>not</u> to encourage "correct" answers.)

In examining the quality of the data we looked mainly for consistency within the conjoint data and between the conjoint and Ci2 datasets. Based on focus groups and other information, we strongly expected that our respondents would be highly segmented according to usage patterns, attitudes, and demographic characteristics. For example, we expected certain subgroups to react differently to enhancements in general, or to certain enhancements in particular. Other groups were predicted to have different priorities regarding the tradeoffs between annual fees and interest rates. These differences held up strongly both in the conventional (Ci2) and conjoint data, thereby increasing our confidence in the data generally.

On one of the pricing concepts respondents seemed to express different priorities in the conjoint questions compared to the CI2 questions that referred to the same concept, albeit in a different way. Although this initially concerned the client, we were not especially surprised, given the concept involved and previous focus group findings.

The Ci2 questions asked what was important in previous behavior, while the conjoint questions asked what would be important in future behavior, given certain conditions. The apparent anomaly was really a result of the price levels tested and the range of prices now available to cardholders we interviewed. A rule of thumb in any kind of conjoint analysis is that the importance of an attribute is partly a function of the range of levels tested. This kind of issue can always come up in conjoint analysis and was not a function of the mode of data collection.

We ordinarily include extra calibrating concept-type questions for verification purposes but decided not to in this case. As mentioned, we had ample data from the rest of the questionnaire to use for consistency checks. A practical problem was that pretest interviews averaged more than 30 minutes, which far exceeded our target of 20 minutes, and interviewers said respondents were not happy with the questionnaire length. This raised concerns about the quality of answers we would get on additional questions.

It turned out that the added length was mainly a function of the type of respondents we had in the pretest and, perhaps, the relative inexperience of our interviewers at that point.

CONCLUSIONS

This study was hardly a definitive test of telephone conjoint research, but it does argue for the viability of this approach. I think in general phone-mail-phone designs will work if the concepts are not too elaborate or abstract, and if the interview can be kept reasonably brief. Respondents have to understand the concepts and the choice scales they are using. I'm not aware of split-sample studies that compare conjoint results among telephone, in-person, and mail data collection procedures, although I'm sure research of this nature has been done somewhere.

An interesting issue is whether conjoint can be done well using only a single telephone interview. Much depends on the complexity and number of concepts involved and the nature of the respondents. I am convinced that most, if not all, telephone respondents in the general population need something visual to refer to--even if they have to write down the various options themselves. Sophisticated respondents can probably deal with a single-telephone call conjoint interview under some circumstances. This suggests possibilities for this approach in business-tobusiness research. However, other reports at this conference indicate that mailed diskettes may be an effective way to conduct conjoint interviews with these populations, especially when a large number of product attributes must be considered.

To return to our telephone study, the client was again quite pleased with the project. One pleasing result was that our perinterview costs were about 55% of those for the earlier in-person conjoint study. While the client is still working out applications of the results, it was encouraging to be told, "You know, this study may actually be useful to us, which is more than I can say about some of the research we get."

[FIGURE 2: EXAMPLE SHOWING GENERAL APPEARANCE OF CONJOINT INTERVIEW WORKSHEET SENT TO RESPONDENTS.]

Example

WHICH DO YOU PREFER?

Slight Preference for car on right

STRONGLY PREFER LEFT				NO PREFERENCE				STRONGLY PREFER RIGHT
1	2	3	4	5	[✓ 6]	7	8	9

1. BODY TYPE: Sedan	1. BODY TYPE: Sedan
2. BODY TYPE: Station Wagon	2. BODY TYPE: Station Wagon
[✓ 3. MANUFACTURER: AMERICAN]	3. MANUFACTURER: AMERICAN
4. MANUFACTURER: JAPANESE	[✓ 4. MANUFACTURER: JAPANESE]
5. MANUFACTURER: GERMAN	5. MANUFACTURER: GERMAN
6. MANUFACTURER: FRENCH	6. MANUFACTURER: FRENCH
7. MANUFACTURER: KOREAN	7. MANUFACTURER: KOREAN
8. MILEAGE: 20 MILES PER GALLON	8. MILEAGE: 20 MILES PER GALLON
[✓ 9. MILEAGE: 25 MILES PER GALLON]	9. MILEAGE: 25 MILES PER GALLON
10. MILEAGE: 30 MILES PER GALLON	[✓ 10. MILEAGE: 30 MILES PER GALLON]
11. MILEAGE: 35 MILES PER GALLON	11. MILEAGE: 35 MILES PER GALLON
12. COLOR: SILVER	12. COLOR: SILVER
13. COLOR: RED	13. COLOR: RED
14. COLOR: WHITE	14. COLOR: WHITE
15. COLOR: NAVY	15. COLOR: NAVY
16. BRAKES: ANTI LOCK	16. BRAKES: ANTILOCK
17. BRAKES: STANDARD	17. BRAKES: STANDARD
18. PRICE: $10,000	18. PRICE: $10,000
19. PRICE: $12,500	19. PRICE: $12,500
20. PRICE: $15,000	20. PRICE: $15,000
21. PRICE: $17,500	21. PRICE: $17,500

CONJOINT ANALYSIS BY MAIL

Dan Cerro
Bain & Co.

Over the past year, Bain & Co. has used ACA (Adaptive Conjoint Analysis) on numerous occasions for projects which have been both national and international in size and scope. We have conducted ACA interviews in person at respondents' offices, or at central locations. We have conducted interviews at conferences and during focus groups, and on a few occasions we have conducted conjoints via mail.

The focus of my paper is on our experiences with conjoint interviews via mail. I will first point out some general guidelines for conducting conjoint by mail and then I will share with you some details of a specific example of this technique.

Because we at Bain definitely favor conducting conjoint in person, decisions to proceed via mail are not made lightly. We have resorted to this technique only in specific circumstances, and I'll pass these circumstances on to you. I hope they will serve as adequate guidelines for your own research.

One key issue in the decision regarding conjoint by mail is consideration of the demographic make-up of the target population. At Bain we try to assess the likelihood that the target group will be able to successfully complete the conjoint when there is no attendant available. In this light we have rejected the idea of mailed conjoints when the population is composed of general consumers (for example, users of laundry detergent or purchasers of some variety of cookies or crackers).

Our use of conjoint by mail has been limited to select groups of highly educated people such as engineers or scientists or even market research personnel. We feel that these groups have two important prerequisites: 1) Due to the nature of their jobs they are likely to be computer users and therefore not computer phobic, and 2) They are likely to have PC's or PC compatibles either on their desks or close by in their companies.

These population groups are not baffled or confused by apparently simple instructions such as "Place the disk in drive A." In fact, from the perspective of the regular computer user, the entire process is seen as fun and interesting and definitely something different.

That is the first point. Make sure your target group is comfortable with computers and has easy access to them. Often a few phone calls or a few minutes conversation with your clients will provide enough clues to let you know whether or not you're barking up the wrong tree.

In addition, we pre-test each study among the actual population to assure that participants will understand the specific task.

Another important consideration when doing conjoint by mail is recruiting. We have developed a method which has resulted in return rates of well over 50%, with 90% percent of the returned diskettes being usable. We take some special steps to get the return rates up that high.

We start with a sample which has been generated by our client to be representative of the population under study. A crew of Bain-trained interviewers then proceeds to call the prospective participants and ask them if they are willing to participate. We are careful to make sure that our estimates regarding how much time they will need to spend are realistic. We have used financial or other incentives but we have not done so in every case. When we do use financial incentives we make sure that the incentive is at least equal to one hour's pay for the population group under study. We have taken pains to insure that the time requested never exceeds one hour. We make sure that the prospects have access to IBM PC's or compatibles and we always request a prompt turnaround on the prospects's part.

Next, we draft a letter that includes detailed instructions for inserting the diskette and activating the program. We generally include a detailed list of the variables and levels, which contains more information about the variables then is available on the ACA screen. An idea which has served us well is a conjoint hotline. We include a phone number the participant can dial to receive help with the conjoint during normal business hours. We have found that 10-20% of the participants make use of this service and even the people who never call this number say that they are more comfortable knowing help is available. We tell people about the hotline availability during the initial call.

Here are some other techniques: Our diskettes are sent by next-day delivery and we include a next-day delivery packet free-of-charge to the participant in the initial mailing. We do not use the mail and we use only delivery packets made for diskettes, for protection. We can, therefore, tell any willing participants exactly when they can expect the diskette to arrive, and there is only a 24 hour delay between their agreeing to participate and the arrival of the diskette. I am sure this quick arrival has helped our return rate.

Since we know when the diskette arrived at the participant's office, it is fairly easy to schedule a follow-up call or a series of follow-up calls to the participant. These calls usually occur within 2 or 3 days of the diskette's arrival. The purpose is twofold. First, we can encourage those who haven't gotten around to the program to please do so. And second, we can assess the participant's reaction to the whole scenario with a few well-placed questions. We have found that people do respond to these calls, since diskettes begin flowing in following them. We never make more than three calls, however, since any more than that could begin to be a nuisance. Conversely, when a diskette arrives, we send out thank-you notes which have been addressed previously.

On a technical note: One small point which we have found to be useful is to set the timer on the program to the maximum amount. It seems that occasionally people begin the conjoint and then for one reason or another have to leave it. This can sometimes lead to data points which are derived from a limited number of screens and therefore probably not as robust as complete interviews. For this reason we inspect the ACA.DAT files to ascertain the extent to which all the screens were seen.

Since we believe that conjoint data are best understood in conjunction with a variety of other approaches, we frequently include some form of survey with the diskettes. We also generally schedule a small sample of in-person interviews from the same population group to check for differences due to interview type. So far we have noted no significant differences due to the interview method.

I'd like to say a word about the general conditions which have prompted our interest in conjoint data collection by mail. We have found that data collection by mail is effective when a fast turnaround time is important. More samples can be collected by mail faster than a team of in-person interviewers can collect the same data, especially when in-person interviews would require a lot of travel time. This is especially true when the population group is composed of very busy or hard-to-see people.

Though Bain generally prefers the personal contact of the in-person interview, we have noted that conjoint by mail is also less costly than in-person interviews and should be considered when the cost of in-person interviews is prohibitive.

In summary then, the important points to consider in a mailed conjoint are as follows:

1) The population works with computers and has easy access to them.

2) Recruiting begins with a phone call asking permission and including accurate time estimates.

3) Incentives are useful (if they are valued by the participant) but not absolutely necessary.

3) A hotline is available.

4) Diskettes are transferred via next day delivery and provision is made for next-day return. Do not use the mail.

5) Follow-up calls are made (but not too many - know when to quit).

6) Set the time limit to the maximum and check the ACA.DAT file.

7) Besides ACA data, collect other types of information (surveys, focus groups) for cross-checking. Collect some in-person data to check for effects due to collection methodology.

8) Consider this technique when time is short or when budgets are minimal.

9) Say thank-you.

Without going into too many specifics, I'd like to tell you about an example of an application of this type of methodology and our subsequent results.

The client in this research was a very high-tech manufacturer of machines used by other manufacturers in the computer industry. The client had been losing share over the past few years to its competitors, both foreign and domestic, and both old and upstart firms. The business objectives were to make a recommendation as to the viability of this particular line of machines.

We believed a conjoint study could assess the client's ability to succeed. We hoped that we could identify the proper aspects of the machine in which they should invest in order to improve their competitive position, and also we hoped to be able to model the market so we could assess their market share potential under various scenarios.

The process of arriving at the proper attributes led to the identification of 19 variables, encompassing thirteen variables which were purely technological in nature and six variables which were non-technical, such as the availability of repair crews and the degree of technological expertise of the salesmen. To save time, only 12 of the 19 variables were admitted to the screens for trade-offs.

The project was complicated by the fact that there is an extremely small universe of people, scattered around the country, who are involved in buying decisions regarding these multi-million dollar machines. As a result of severe time constraints and the expense of in-office interviews, a decision was made to collect the conjoint data via a mailout.

Data were collected from 30 respondents, a small number, but one which encompassed some 40% of the total universe. Focus groups were conducted (which included a card sort conjoint) prior to the mailout, and a small number of in-person interviews (pen & paper) were held.

The results of this study were uncanny in their ability to model the previous year's market shares. In fact, the real figures were only 3-5% off from the model output, well within the standard error. It was amazing to us that despite a small N, a high number of attributes (only some of which were tested), a mailout approach, and a distinct lack of consensus (even within the same company) among the decision makers as to what constituted the ideal machine-- the ACA program was able to accomplish such precise modeling. Since ACA was able to model the previous year's market so effectively, we were confident in projecting the future marketplace based on various configurations of both our client's and his competitor's machines.

In summary, despite theoretical and practical obstacles, ACA helped us to identify the key drivers behind the purchasing decisions (a surprising number of which were non-technical) and to direct future investments down the correct avenues. The ACA data also dovetailed quite nicely with the information collected from the focus groups, in-person interviews, and card sort conjoint. All in all, after weeks of calculations and thought, we felt very comfortable with our final recommendations to our client.

In this case, we felt that our ability to collect accurate and adequate data within the confines of a tight timeline was due to the our method of collecting conjoint data by mail.

COMPLEX COMPUTER INTERVIEWS

Robert Zimmermann
Maritz Marketing Research, Inc.

Let me begin by saying that ACA (Adaptive Conjoint Analysis) is, to my mind, one of the most insightful attempts to utilize the real power of a microcomputer in the collection of marketing research data. Unlike Ci2 (Computer Interactive Interviewing), which simply does better what a good interviewer is already capable of doing, ACA actively intervenes in the flow of the interview at a level completely beyond the capacities of even the most highly-trained interviewer.

I think we are only beginning to realize the potential of this approach. A breakthrough of this kind not only allows us to implement designs we have avoided in the past, but it also inevitably suggests to us wholly new approaches that were previously beyond the range of our imagination. Thus, while our capabilities are extended, so are our expectations, and we continue to push the limits of the possible. Since the technology is new and evolving, none of us really knows precisely what those limits are. I am going to discuss some of those limits, and hope to suggest some directions for extending them.

I am not going to deal directly with solutions that the ACA manual explicitly treats, except to list some them here, to make sure that you are aware of them, and so that you know where I am starting. ACA can be called from Ci2, with a return to the Ci2 module following the execution of the ACA module. You can branch to one of several ACA modules based on the response to a question or questions in the Ci2 module. You can combine two attributes into one to make interactions explicit or to avoid impossible combinations. Similarly, in the latest version of ACA, you can prevent two traits from being combined to construct options for the paired choices. You can limit the number of traits presented in each option pair, and you can limit the number of traits and trait levels available for the construction of the option pairs.

Conjoint analysis can be used in many ways. Sometimes the principle objective is to determine the value structures that underlie behavior. This can be useful in establishing long term development and marketing strategies. However, my assumption is that the primary objective of most conjoint designs in marketing research is to model individual decisions, and in particular, purchase behavior. Such decisions are almost always both hierarchical and segmented.

By hierarchical I basically mean that not all trait strengths are commensurate. When the smallest difference in utilities between two levels of trait A is larger than the difference between the two most extreme levels of trait B, there can be no reasonable expectation of a trade-off. Trait A is super-ordinate; decisions with respect to trait A may affect the decisions or the range of options available for trait B, but nothing with respect to trait B affects decisions on trait A. Inclusion of both traits in the same analysis will provide evidence for their hierarchical relationship, but frequently any additional information will be confused or imprecise, particularly with respect to the subordinate traits.

Segmentation frequently results from hierarchical decision processes. As an example, within the automobile market there is frequently an initial choice which limits the buyer's scope of consideration. This might be a distinction based on price (luxury vs. subcompact), maker (American vs. foreign), or perhaps general type (4-door sedan vs. 4-wheel drive utility vehicle). The consumer rarely actively considers the full range of models and options available. There is an initial screening which cannot be based on specific details, followed by a more detailed choice among the class of vehicles defined by the initial screening. A conjoint can treat the initial screening, in which case the attribute is vehicle class and not model. If the attribute is model, then the conjoint either deals only with one segment, or separate model lists must be presented for each vehicle class, which is functionally equivalent to a segmentation. It is less than useless for a prospective owner of a Mercedes to provide us his valuation of a Yugo or an Isuzu. Both the constraints of the technology and the reality of the actual decision process dictate that not all models be included as levels in the model attribute, but the most important consideration is that this does not reflect the reality of the decision process.

Sometimes the only thing affected by a segmentation is the way the conjoint is set up. Prices paid for commercial office supplies differ so markedly as a function of volume of purchase and distribution channel that it is impractical to include tiny offices and very large offices in the same price sensitivity trade-off, yet when price is scaled as a percent of last price paid, the utility values for equivalent levels of price and brand are remarkably similar across office size. Utility values may be reasonably invariant across different members of a product line, yet the conjoint task will seem awkward and artificial if it isn't tailored to each person's individual purchase preference.

A schema for setting up a conjoint analysis might be somewhat as follows:

1. Determine as precisely as possible what it is that you want to know. In particular, what aspect of choice behavior are you trying to model?

2. Specify explicitly how that choice will be formulated by the different individuals in your target population.

3. Construct your conjoint administration procedures such that each individual has as realistic a set of options as is practical.

4. Let the generated data set determine to what extent utilities are transitive across segments.

It should be noted that in estimating market share, and in many instances of simulation, it is more important that everybody be presented with a relevant choice than that everybody be presented with the same choice, since choices that the respondent would not make under any condition do not contribute to the analysis.

I got into all this by blithely telling a client that, of course, we could branch to separate conjoints. The client came back with a request for eight discrete conjoints in each of three different markets. I then read the fine print in the documentation of the then current version of ACA and discovered that there was methodology only for a branch to one of two separate conjoints. These were presumed to be substantively different conjoints to be analyzed separately.

Since each of the markets was to be tested and analyzed separately, I only needed to deal with branching to eight conjoints. The method that ACA and Ci2 use to communicate with each other in fact permits at least an eight fold branching. However, I discovered to my dismay that when I put the Ci2 module and the eight complete conjoint modules on a floppy disk, I had only room for the data from 2 or 3 subjects. This presented me with the decidedly unpleasant prospect of having to work several hundred disks. Since these conjoints were price sensitivity trade-offs, with only two attributes, it was also obvious that most conjoint problems would not permit eight branches, if this approach was used.

The modular nature of Ci2 and ACA permits considerable condensation, as long as one is familiar with MS-DOS batch processing commands. The first space saver is to take out all the optional ACA introductory screens, where they occur once for each of the eight conjoint analyses, and display them once, just before branching from Ci2. This has the added advantage of giving the respondent something meaningful to consider while CI2 and ACA perform the housekeeping tasks necessary for the interfacing.

Secondly, there were in essence two parallel sets of conjoints. Each set needed to be analyzed separately, while the four data sets within each set reflected only the marketing mix, and could be treated together for simulation purposes. This meant I needed only two data sets. This did not save a great deal of space, since the total amount of data was not reduced, but it made the data much easier to work with.

Since attribute levels are in a separate module from the ACA frames module, there is an additional step which can save considerable space. Presuming each conjoint involves the same dimensions, differing only in the separate levels within the dimension, a discrete phrasing of the statements which call the attributes, used in conjunction with separate calling screens in the Ci2 module, can reduce the eight ACA control modules to one module. The appropriate separate attribute lists are presented to this module using the MS-DOS RENAME command.

This may sound rather arcane, but the resulting interview flows quite seamlessly, and it is really quite simple to set up. The space saved is considerable. In the present example, there was space remaining on the disk for at least 100 respondents. Since we limit each disk to 30 respondents, as a precaution against data losses, it is clear that we have not yet stretched the technical limits of the procedure.

This project presented a second problem. While the problem may be specific to a fairly narrow range of applications, the solution might have more general applicability. The context is the same price sensitivity study noted above. There were a discrete number (2 to 4) national chains which were of principal interest to the marketing strategy. There were 2 to 4 regional chains which generally appeared in only one market. And finally, there was an indefinite number of locally owned competitors; more than 60 were mentioned by respondents in some markets.

The market simulations would manipulate only the national chains. The regional and local competitors were of interest for two reasons: first, as the price charged by one or more of the national chains increased, the market shared by the national chains dissipated to the innominate locals; and second, the national chains might be differentially susceptible to erosion by local competitors, limiting their options in competitive pricing. The easiest way to deal with this amorphous group is to assume that they are either a somewhat passive presence, or if they do change, they respond with a sort of sluggish uniformity. Other models are conceivable, but I suspect they would add to the complexity faster than they added to the insights.

With these assumptions, it is only necessary to deal with each individual's preferred option from the class, since, barring interactions, that is the only local competitor that will influence market share for that individual. Each individual was requested in Ci2 to name the establishment he most preferred, which was not already on the attribute list of levels. He was then instructed to consider this as his personal "other" restaurant, and it was referred to in that way within the ACA module. There is no necessary bar to using an attribute level to refer to a definable group, as long as it is possible to assume that each individual has a consistent single preference within that group. Variety-seeking behavior is a notable exception, but then, variety-seeking behavior creates problems for conjoint analysis in almost any form.

There is a more elegant solution, one which we did not implement, but which I am convinced presents no serious problems. Since the attribute lists are simply ASCII text files, there is no reason why they cannot be recreated for each respondent by a run time module which writes attribute lists in usable format from information passed from the Ci2 module. Both numeric responses and open-end responses are located in definable and accessible places in the Ci2 data files. It would then only be necessary to insert an executable run time module in the recursive batch stream, just before the call to the ACA control module.

I have taken a tangential tack on the problem of the limits of complexity when using ACA. With respect to ACA, complexity takes on many faces. From one vantage point, the interactivity and simplicity of the designing and analysis modules permit people who understand the applied uses of conjoint analysis to design and execute analyses without needing to know very much at all about the statistical niceties of orthogonal sets or loglinear analysis or other shibboleths which frighten the uninitiated.

From another vantage point, ACA presents the respondent with a much simpler and more natural task than did the traditional set of orthogonal contrasts. To me this is the true power of ACA. I have dealt with evaluational data in educational and clinical psychiatric settings, as well as in a marketing research setting. I am convinced that the most critical point in the whole evaluation procedure is the making and recording of the evaluation, in situ. One should try never to accept ease of set-up or ease of analysis in a trade-off with lessened validity in data collection. Always make the rating process as unobtrusive and transparent to the respondent as possible.

ACA permits us to model choice situations that are far more complex than any that exist in real life. Not that people are not continually confronted with excruciatingly complex situations. It is rather that because they are complex that the problems are not dealt with as some kind of phenomenologically concurrent decision matrix. The psychological literature strongly suggests that if you go beyond five traits with five levels each, you have exceeded the range of a significant portion of the population, at least with respect to a direct choice context. (It is perhaps less clear that this limitation applies to the value structure underpinning choice.) I feel we are on risky grounds if we exceed those limits. In immediate choice contexts, individuals are likely to form choice hierarchies or ignore low value options when confronted with more complex choices.

We should not be trying to push the limits of ACA in terms of number of attributes or number of levels within an attribute. We should be pressing the limits of the technology by making it do all it can possibly do to create an interesting, realistic, valid rating context. All the complexities that I have introduced above have at least one thing in common--they are designed to present a simpler, more natural, more realistic choice to the respondent, a choice tailored to reflect as nearly as practical the frame of reference he brings to the decision context. Doing so should provide added validity and clarity to the resulting data. The respondent's simplicity unfortunately sometimes becomes our complexity. ACA is itself a striking example. A procedure too complex to be administered without a computer has significantly reduced the perceived complexity of the respondent's task.

OVERVIEW OF PERCEPTUAL MAPPING

William D. Neal
Sophisticated Data Research, Inc.

INTRODUCTION

Perceptual mapping is one of the few marketing research techniques that provides direct input into the strategic marketing planning process. It allows senior marketing planners to take a broad view of the strengths and weaknesses of their product or service offerings relative to the strengths and weaknesses of their competition. It allows the marketing planner to view the customer and the competitor simultaneously in the same realm.

Perceptual mapping and preference mapping techniques have been a basic tool of the applied marketing research profession for more than twenty years now. They are among the few advanced multivariate techniques that have not suffered very much from alternating waves of popularity and disfavor. Although I observed a minor waning of the use of the techniques in the early 1980's, they are now as popular as ever.

And although these techniques have been used extensively over a large number of applied research studies, and for a very wide variety of product and service categories, and have been subjected to extensive validations, there still remain some very basic issues as to the procedure's applicability and usefulness.

In addition, there remain many outstanding issues concerning the proper procedures and algorithms that should be used for perceptual mapping.

So, I see that my main task at this conference is to raise the issues, as I see them. I am taking a rather naive approach. That is, I will approach these issues from the research manager's point of view, and not the statistician's. These issues represent the kinds of questions that my clients ask me and my staff. Obviously, I have some biases, but I will try to minimize those, and concentrate on the issues.

I know that many of these issues will be addressed at this conference, both in formal presentations and in informal discussions. I am raising the issues in the hopes that this introduction will encourage greater investigation, increase validation activities, encourage additional, practitioner-oriented publishing activities, and provide fuel for additional conferences of this type.

WHAT'S IN A NAME?

So, let's start with the first issue. Just what is perceptual mapping? Or, is it preference mapping? Or, is it structural segmentation? Or what? Here is a list of some of the names that I have seen this procedure called:

- Perceptual Mapping
- Preference Mapping
- Structural Segmentation
- Brand Mapping
- Behavioral Mapping
- Strategic Product Positioning
- MDS Mapping
- Market Mapping
- Product Mapping
- Goal Mapping
- Image Mapping
- Semantic Mapping

Well, if the only difference between these various names is the selection of a particular attribute set, then I suggest that we rename the technique to just plain old Multivariate Mapping, ore even just mapping. If one wishes to distinguish algorithms, then the proper descriptive prefix can be used, such as discriminant analysis-based multivariate mapping. Or, if one wishes to distinguish the types of attributes used, then an appropriate suffix like multivariate mapping of consumer product preferences would be more appropriate. All are far more descriptive and certainly reduce confusion.

If there are true differences between these various names and the idea of generic multivariate mapping, then we are obliged to make those distinctions and perpetuate that nomenclature throughout the profession. As it stands now, the name perceptual mapping is confusing to both marketing managers and many research professionals. Currently, most marketing managers assume that there is a fundamental difference between perceptual mapping and, say, preference mapping. Is there really?

ISSUES AND PROBLEMS WITH CURRENT ALGORITHMS IN GENERAL USE

Following are the three major classes of algorithms that are generally in use for perceptual mapping in the applied marketing research arena. Included is a brief discussion of their strengths and weaknesses, and some outstanding questions, from a user's viewpoint.

A. <u>Discriminant analysis</u> is still the most popular algorithm in use today for applied multivariate mapping. The procedure is widely available. The algorithm is robust in that the assumptions concerning the continuity of the data, and the data distributions can be relaxed to a considerable extent.

The inputs to discriminant analysis consist of individual respondent ratings of products across attributes. The basic assumptions are that the rating scales are continuous and normally distributed. However, in using the technique for mapping purposes, these assumptions can be relaxed to the point that products simply rank-ordered on attributes will usually provide sufficient information for mapping purposes.

Discriminant analysis is much like regression analysis in that it uses a least-squares approach in an attempt to fit a linear model to the data. However, the dependent variable is nominal. That is, for mapping purposes, the dependent variable is the product set being rated. Thus, each product rated by each respondent is an input record, so if a respondent rated five products, that generates five input records.

Discriminant analysis then calculates the coefficients to a set of standardized linear equations, called discriminant equations, that explain the differences between the product ratings. Or, said a different way, explains the variance between ratings of different products.

The formation of the linear equations follows an order, such that the first equation explains the most variance, the second explains the most variance remaining after accounting for the variance explained by the first, and so on until you reach a limit of one less than the number of products being rated, or the number of variables, whichever is less.

These linear equations are further constrained so that each one is uncorrelated to the other. That is, they are orthogonal.

These two properties, the successive optimization of the variance explained, and the orthogonality of the equations, form the basis for mapping, because one is assured that the first linear equation, which typically defines the X axis of a map, explains the most variation between products, and the second linear equation, or Y axis, explains the most variance between products, after accounting for the variance explained by the X axis (given the limitations of the least-squares procedure). And the X and Y axes are orthogonal.

In most cases, the first two equations account for the majority of the variance between product ratings, and are the only significant dimensions. Later, we will discuss significant dimensions beyond two.

Assuming for the moment that there are only two significant dimensions, the calculated coefficients of each variable in each equation define the X and Y coordinates of the attribute on the map.

The X and Y coordinates of each product are calculated by substituting the mean attribute ratings of each product into the two discriminant equations, and calculating the results.

The linear discriminant equations allow the researcher to easily plot additional products or concepts into the derived space. These equations also allow the researcher to explore the distributions of specific customer groups in the derived space.

Most widely available discriminant analysis algorithms provide a variety of useful statistics to the researcher, such as eigen values to show you the variance explained by each equation, tests of significance for each equation, multivariate F statistics to show the significance of the group differences, and correlations or loadings between each attribute variable and the discriminant functions.

The procedure also has a few drawbacks.

Obviously it requires individual ratings of individual products (or services, or firms) on each of a selected set of attributes. Consequently, there is a perpetual problem with what to do with missing data points. Although I have read a dozen papers on handling missing data in discriminant analysis, there seems to be no consensus short of case-wise deletion. Yet, the realities of today's marketing research industry often makes this an unacceptable solution. Is mean substitution an appropriate solution? How does mean substitution affect the calculation of the discriminant functions? What are realistic limits on the amount of mean substitution to use? What are realistic and easy to execute alternatives to mean substitution?

The procedure is dependent on the selection of the appropriate attribute set. The omission of important discriminating attributes may lead to false conclusions concerning the dimensionality of consumer ratings of product differences.

Also, the procedure highlights those variables that discriminate between products, and will not display on the map attributes that may be extremely important, even dominating product choice, but that do not differentiate between products. Alternatively, situations often develop where a particular variable discriminates between products, but is not important in product choice.

Often, the selected set of attribute variables is highly correlated. Consequently, there is no control over the number of attribute variables, or over which attribute variables enter the discriminant solution and define the relevant space. To overcome this situation (note: this is NOT dependent on stepwise algorithms), multiple passes, forcing inclusion of variables in which there is a high interest, are often required. This can be costly.

The inclusion or exclusion of one of the products or firms being rated often changes the dimensionality of the space, especially when the set of firms or products under consideration is small or radically different from other products. It is often difficult to convey this situation to research managers and senior marketing management. A radically changing product space detracts from the confidence that senior marketing managers have in the procedure. Is there some way of overcoming this, short of adding more products simply to stabilize the space? That solution is often not viable in researching industrial products or emerging consumer product categories.

B. R-Type Factor Analysis is seldom used as a mapping procedure in today's applied marketing research field, although in the 1970's it was the preferred mapping procedure among many applied researchers. And, there are a few empirical studies that report it to be superior to discriminant analysis. Although you have the same problems with what to do about missing data and with selecting the relevant set of variables as you have with discriminant analysis, at first glance this procedure seems to overcome two of the problems with discriminant analysis: All variables are shown on the map, and the inclusion or exclusion of products has no effect on the extracted dimensions.

The inputs to factor analysis are very similar to those for discriminant analysis: product ratings across attributes. However, an additional ingredient is required. You must also collect an importance rating from each respondent for each attribute. These importance ratings are usually the basis for developing the mapping space. The basic assumptions concerning the distribution and continuity of the rating scales should not be relaxed. Some researchers use the attribute ratings of each product as input to the factoring algorithm, and ignore the importance ratings.

At this point the two procedures part ways. Unlike discriminant analysis, where the variance between product ratings is addressed, factor analysis attempts to explain the correlation between importance ratings (or product ratings) of the variables. That is, the first factor equation is that linear equation that explains the maximum amount of correlation between the variables, and the second extracted equation explains the most of the remaining correlation, and so on, until 100% of the correlation is explained with a number of factors equal to one less than the number of variables. The extracted factors are linear equations which have a coefficient for each variable. These coefficients are commonly referred to as factor loadings.

The output of factor analysis does meet the basic criteria for developing a map. The first two dimensions explain the maximum amount of variance (i.e. correlation) between the importance ratings of the variables (not the ratings of the products), and they are orthogonal. Thus, to define a variable location on the map is a simple case of using that variable's loading on the first factor as the X coordinate, and its loading on the second factor as the Y coordinate.

Factor analysis is an interdependence procedure; thus the various differences in product ratings are ignored until after the factor equations are derived. Product locations in the derived space are calculated by plugging the average standardized product scores on each attribute into the two factor score equations and calculating the X and Y coordinates.

The extraction of factors is highly sensitive to the number of correlated attributes. The addition or deletion of an attribute may dramatically alter the dimensionality of the derived space. In addition, extraction of factors is dependent on the intercorrelations between variables, and does not necessarily optimize the separation between products, like discriminant analysis. Furthermore, a single variable that may be considered extremely important and dominating the selection of products, like safety, may not show up as a dimension on a map, simply because it is not correlated to any of the other measures.

Myers and Tauber (Market Structure Analysis, AMA, 1977) recommended overcoming this problem through the use of a "weighted covariance approach," where the input to the factoring program is a matrix of product covariances, weighted by regression scores derived from regressing the importance ratings against product choice. But this has proved to be a bulky and difficult procedure to implement, and there has been little empirical validation.

However, the entire notion of "mapping" products into an extracted factor space, even using the weighting covariance approach, was argued as a blatant misuse and misinterpretation of the basic concepts of factor analysis by David Stewart in his extensive dissertation on "The Application and Misapplication of Factor Analysis in Marketing Research" (JMR, Vol XV111, FEB 81, p. 51-62).

C. <u>Non-metric scaling procedures</u> are still used quite often for multivariate mapping. However, I am only going to concentrate on one of those, and briefly describe the others.

1. <u>Correspondence Analysis or Dual Scaling</u> is gaining in popularity, mainly because there has been a considerable amount written on the technique over the last few years, it is an extremely robust technique, it has simple data collection requirements, and the algorithms are becoming widely available.

   Because of its ability to use summary distributions of nominal data, correspondence analysis is often used as a post-hoc mapping procedure for studies that did not originally contemplate multivariate mapping. The procedure puts no significant demands on the distribution of the data.

In addition, the point-point maps produced from correspondence analysis are directly generated by most of the programs and they are much easier for general marketing managers and creative promotional personnel to understand.

Inputs to correspondence analysis can be as simple as a summary table of respondent checks as to whether a product has a certain characteristic or not. Almost any data collection procedure imaginable can be used to provide inputs to correspondence analysis. Respondents can be asked to name a single brand most associated with an attribute, or occasion, or store. Even open-ended questions can be used by asking respondents to name the qualities most associated with a brand, or store, or personality. There are no restrictions as to how many or how few items a respondent associates with a product.

The data input to the program is a matrix of counts of how many times a product, service, or firm is associated with an attribute, usage occasion, need, or whatever.

Consequently, the data collection process is highly simplified. This has considerable appeal in light of the industry's intense interest in "respondent abuse" and declining response rates.

Correspondence analysis has a unique ability to integrate a large amount of data from divergent perspectives on a single map. For example, brands, product attributes, needs fulfillment, and usage occasions can all be shown on the same map.

One of the drawbacks of the technique is that it uses only summarized distributions of nominal data for most of the algorithms that are currently available. Thus, a considerable amount of the variance associated with a database of individual responses is sacrificed. Another drawback is that metric data distributions must be "nominalized" to be used in the procedure.

The exception to the restriction of summarized data is Benzacri's SPAD program that few researchers have access to. SPAD allows you to input either the individual observations, or ratings, or the summarized data. Interestingly, you will often get differing amounts of explained variance, and/or different product and attribute locations on the map, depending on whether you use the individual observations or the summarized data. Frankly, I'm not sure why this happens.

If there are metric distributions that must be converted to nominal variables, the selection of the appropriate break-points is critical and has a considerable effect on the amount of explained variance and the extracted dimensions of the correspondence map. We need a solution to this situation, and guidelines on proper procedures for nominalizing metric data.

2. KYST, PROFIT, INDSCAL, TORSCA, PREFMAP, PROXIMITY, ALSCAL, SSA-1 thru SSA-4, MRSCAL, MINISSA, MINITRI, PARAFAC, and MDSCALE, (to name a few) all fall into a class of mapping procedures called non-metric multidimensional scaling procedures. However, in actuality, some of these algorithms are more metric in nature than non-metric. Although conceptually different from correspondence analysis, for the most part they have been replaced with correspondence analysis because the data collection procedure is as easy for one as the other.

Some of these methods release the researcher from having to specify the appropriate attribute set altogether, and instead rely on how consumers judge the products in question to be similar, or dissimilar. The data collection process is often an unstructured sorting task; respondents are asked to sort products into piles that are similar, or simply rank order products based on either their overall similarity, or their similarity for a given attribute.

Orthogonal scales are then derived to explain perceived differences between products. The derivations are based on minimizing stress in the fewest dimensions possible, while preserving respondents' order of similarity. The nature of the dimensions are often determined by simply inspecting the manner in which each product is aligned with each dimension.

For some of these procedures, explanatory variables can be depicted on the map by asking consumers to correlate the similarity of a given attribute, or usage occasion, to the products.

The procedures for the most part are quite sensitive to the number of products in the data set. The addition or deletion of one product will often change the dimensionality of the space.

In addition, several of these algorithms require complicated, and often conceptually difficult, data transformations to work correctly and they are quite sensitive to the types of transformations undertaken. (see "Multidimensional Scaling," by Kruskal and Wish, Sage University Press, 1978.)

CURRENT ISSUES IN PERCEPTUAL MAPPING

A. Defining and limiting the relevant space

How is the relevant space limited? There are three types of limitations that must be placed on the relevant multivariate space that will be analyzed and mapped. They are:

1. Limits on the population that is to be surveyed. This seldom poses a serious problem because it tends to be self-defining in terms of users, or purchasers of the products, services, or firms in question. However, there are questions as to how familiar a respondent is with a product or brand. This will be discussed in a later section.

2. Limits on the relevant set of variables that will be used to define the perceptual space. In my opinion, this is the most critical area for setting limitations, except for those using the scaling methods based on overall product similarities. The major question to the applied researcher is what variables are to be used to orient the perceptual positioning of the various competitors. A nearly unlimited set of variables is available.

   The selection of the relevant variable set determines the type of map that will be produced. That is, the map could be based on such things as purchase behavior, organizational images, product usage behaviors, product attribute characteristics, brand images, consumer goals, consumer needs, convenience issues, or some combination of these.

   This is a critical decision and requires the agreement of senior marketing management to concur with the appropriate attribute set. Determination of the relevant set requires the professional marketing researcher to critically examine previous research in the category, conduct qualitative research, and creatively select those variables that will provide senior marketing managers with the insight necessary to form marketing strategy.

   The problem is that we all have seen empirical evidence that the relevant set of attributes changes dramatically from product category to product category, and even within a product category over time. Yet, there is no substantial body of knowledge to tell us what is the relevant set of variables that should be used in any one category. We are left to re-inventing the wheel every time we approach a new product category with multivariate mapping. This severely detracts from the general adaptation of multivariate mapping procedures at the strategic marketing planning level.

3. Limits on the relevant set of products, services, or firms that will be mapped into the multivariate space is also a major issue. Although I don't believe that this is as critical an issue as the selection of the relevant variable set, it is still a serious one. A balance is required.

   In this era of market fragmentation and the rapid emergence of new product categories and sub-categories, brought on by an acceleration of differentiated products flooding the market place, the selection of the relevant competitive set of products or services is ever-changing.

If the relevant set of products, services, or firms is too broad, we may fail to uncover those truly discriminating variables that may reveal an opportunity for a competitive advantage. That is, some non-competitive products may so skew the spatial dimensions of the map that differences among the true set of competitors may be hidden or overlooked.

On the other hand, the selection of too narrow a competitive set may destine the marketing planner to focus on the wrong competitors and wrong dimensions. As an example, department stores for years focused on competing department stores as the relevant set, ignoring the single-merchandise-line specialty stores and the deep discounters - until the department stores' bottom lines started gushing red ink.

Given the rapid nature of change in the competitive set for most product and service lines, we could not rely on a body of literature to solve this problem. What is needed is a set of generally accepted procedures for determining the relevant competitive set at any point in time.

Permit me to continue the discussion of issues in multivariate mapping in a more abbreviated manner. I will limit my remarks from here on to discriminant analysis-based multivariate mapping, since that is what most of us are using, although many of these issues apply to other algorithms as well.

B. Are there particular product categories or merchandise lines or firm-types where discriminant analysis-based mapping works better? If so, then what are the characteristics of those product categories or industries?

C. Is "high-involvement" in the respondent rating process a necessary prerequisite for multivariate mapping? What level of familiarity is necessary and sufficient to include a set of ratings into the definition of the relevant multivariate space?

D. Extracting the dimensions.

1. What are some good rules of thumb for determining how many dimensions to use? How much variance needs to be explained for us to be comfortable? How should we handle dimensions with low variance explained, but that test as significant?

2. How do you display more than two dimensions? What procedures and graphics algorithms are available? What graphics procedures best convey the information in the multivariate space to managers and creative professionals?

3. If you are forced to use a two-dimensional map but have three or more significant dimensions, how do you adequately show those attributes that are heavily loaded on the third dimension? Do you eliminate those from the display? If you do eliminate them, what criteria should you use?

4. What actions should you take when the first extracted dimension explains much more variance then the second dimension? Is it appropriate to display those two dimensions as equal axes in the map?

E. Plotting the variables in the derived space raises some interesting questions.

   1. Should variable coordinate weighting be used to show differences in the amount of variance explained by each axis?

   2. If so, what should be used as the appropriate weights - percent of variance explained by each axis, eigen values, partial F values, or something else?

F. Plotting the firms/products in the perceptual space

   1. How should we show which products or firms are significantly different from others on the map?

   2. Does anyone attempt to draw confidence limits around the mapped points anymore? Shouldn't we?

G. What about "ideal" points?

   1. Should "ideal" points be used at all?

   2. If so, what is the best way of doing that?

      a. Use attribute importance ratings and treat these as another product rating? In other words, do we permit importance ratings to assist in the definition of the relevant space?
      b. Or, should we calculate standardized mean importance ratings and plug those values into the previously extracted dimensional linear equations to calculate the coordinates of the ideal point?

      c. Should we use a respondent's highest rating of any firm/product on each attribute and use that as a proxy for his "ideal" product?

d. What about using each respondent's preferred firm/product and simply duplicating that rating as the set of "ideal" ratings under the assumption that the respondents will purchase or use those products closest to their ideal?

e. Is it appropriate to map a "generalized" space, then segment the sample on importance ratings or product preferences, then impose the mean ratings of those segments as multiple "ideal" points on the map?

f. What other methodologies are there for generating "ideal points"?

g. What do you do when any one of these procedures dramatically skews the map?

H. Is longitudinal mapping a valid concept? What are the critical issues in overlaying maps? What are the best methods for doing this?

1. Line up "index" points from successive time periods so as to minimize the variance between them? Should the index points be the vector of importance ratings, or some other measure?

2. Select a very stable vector that consistently discriminates between at least two of the products or firms, and minimize the variance between their positions over successive time periods?

3. Use both of these methods in combination?

4. Re-generate the dimensions with each attribute from each time period representing a separate attribute, and each product from each time period representing a separate product?

5. Always use the original space, and simply plug in the standardized means for each product from successive time periods into the linear dimensional equations and calculate the new coordinates?

6. What other procedures are being used?

I. How can you incorporate volumetric data into multivariate mapping? In other words, how can you show the marketing manager where the greatest demand exists on the map? Or, where the opportunities are.

a. Are scatter plots of grouped respondent locations the only thing available?

b. Or, can we develop a surface-plot over the mapped space that will depict such things as dollars spent, or number of items bought, or even number of times visited? What methods are being used now? What could be done with the new graphics packages combined with multivariate "smoothing" routines to superimpose surface plots over the derived space?

CONCLUSIONS

Needless to say, there are still many outstanding issues and further development opportunities with multivariate mapping procedures. I'm sure that there are others besides these. I would like to challenge you to address these issues, share them with your peers, publicize solutions to them, freely subject them to validations, and give us more specificity in executing this most powerful and useful marketing research procedure.

COMPARING PERCEPTUAL MAPPING AND CONJOINT ANALYSIS:

THE POLITICAL LANDSCAPE

Joel Huber
Duke University

John A. Fiedler
POPULUS, INC.

The Choice between Conjoint Analysis and Perceptual Mapping

Choosing between conjoint analysis and perceptual mapping may appear to many to be trivial--a classic no-brainer. It is like the choice in a world championship between the LA Rams and the New York Mets. The outcome trivially depends on your choice of turf and rules of play. Perceptual mapping is played on a turf of image products, such as cigarettes and bourbon, and its rules specify that the competitive structure can be reduced to two-dimensional competitive maps, something that is only possible if the perceptions on products on attributes are strongly correlated with one another. By contrast, conjoint analysis plays on a conceptually different field. The soft turf of image products is replaced by the hard surface of functional products such as computers or forklift trucks. Further, the rules of conjoint keep the attributes sharply distinct, so that the impact of a change in any one of them is clearly discernible. Finally, the outcomes of the two systems are quite different. Perceptual mapping forms elegant spaces which locate consumers' perceptions of the brand, while leaving obscure the relationship between attribute levels and preferences. Compare those maps with the partworth functions of conjoint analysis, which move effortlessly from attribute levels to preferences, apparently finessing the issue of perceptions altogether.

While perceptual mapping and conjoint analysis techniques have been traditionally quite different, the Adaptive Perceptual Mapping (APM) program of Sawtooth Software makes them much more similar. What is novel about the APM approach is that it forms maps at the individual level, and then uses these to predict preferences in a choice simulator. When its individual-level model is compared with the individuallevel model in conjoint, the differences between the two become much less pronounced.

My plan today is to examine the similarities and differences between an individual-level perceptual map and a conjoint analysis. I will then describe a study in which both techniques are used to predict straw votes in the current presidential race. While the winner is the one that predicts the most votes for each individual, the main insights from this study will involve distinguishing when one system will be more appropriate than the other, and why.

Differences Between Conjoint Analysis and Perceptual Mapping

Table 1 summarizes the differences in inputs, outputs, and assumptions between the individual models of choice reflected in the two systems. As we contrast perceptual mapping and conjoint analysis, we will focus on the particular versions, ACA (Adaptive Conjoint Analysis) and APM, although the conclusions apply to any conjoint system and any perceptual mapping system that is estimated at the individual level.

TABLE 1

DIFFERENCES IN INPUTS, OUTPUTS AND ASSUMPTIONS BETWEEN PERCEPTUAL MAPPING AND CONJOINT ANALYSIS

UNIQUE TO PERCEPTUAL MAPPING	SHARED	UNIQUE TO CONJOINT ANALYSIS
	INPUTS	
IDENTIFICATION OF IDEAL ATTRIBUTE LEVELS	PERCEPTIONS OF PRODUCTS ON ATTRIBUTES	RANKING OF ALL ATTRIBUTE LEVELS
		TRADEOFFS AMONG PROFILES
GENERAL IMPORTANCE OF ATTRIBUTES		IMPORTANCE OF BEST VS. WORST ATTRIBUTE LEVELS
	OUTPUTS	
PERCEPTUAL MAP CONTAINING PRODUCTS AND IDEAL POINTS	SIMULATION OF PRODUCT CHOICES	PARTWORTH VALUES OF EACH LEVEL OF LEVEL EACH ATTRIBUTE
	ASSUMPTIONS	
UTILITY IS SYMMETRIC AROUND IDEAL POINT	EACH ATTRIBUTE LEVEL MAPS INTO UNIQUE UTILITY	NO CONSTRAINT ON FORM OF PARTWORTHS
ATTRIBUTE WEIGHTS ARE MODIFIED BY PRINCIPAL COMPONENTS:	ATTRIBUTE WEIGHTS ARE INDEPENDENT:	ATTRIBUTE WEIGHTS ARE MODIFIED BY CONJOINT JUDGMENTS:
MORE CORRELATED ATTRIBUTES GET MORE WEIGHT	THE LEVEL ON ONE ATTRIBUTE DOES NOT CHANGE THE UTILITY OF ANOTHER	MORE IMPORTANT ATTRIBUTES GET MORE WEIGHT

In terms of inputs, both systems need to collect respondents' perceptions or ratings of products on attributes. In perceptual mapping, these ratings are the basic material that is used to form the maps, while for conjoint analysis they allow one to use the partworth utility functions to estimate the utility of the products that have been rated. Both systems also collect some measure of the importance of each attribute. In the APM system this measure tends to be vague and global while in ACA it is anchored at the best and worst levels of the attribute. Despite strong conceptual differences, both importance measures correlate very highly in practice. A big difference between the systems is the way they assess the utility of each attribute level. Perceptual mapping directly assesses the ideal level of each attribute and measures utility as a weighted deviation from that ideal. Conjoint analysis, by contrast, asks respondents to evaluate profiles or product descriptions and uses these judgments to infer the values of the attribute levels.

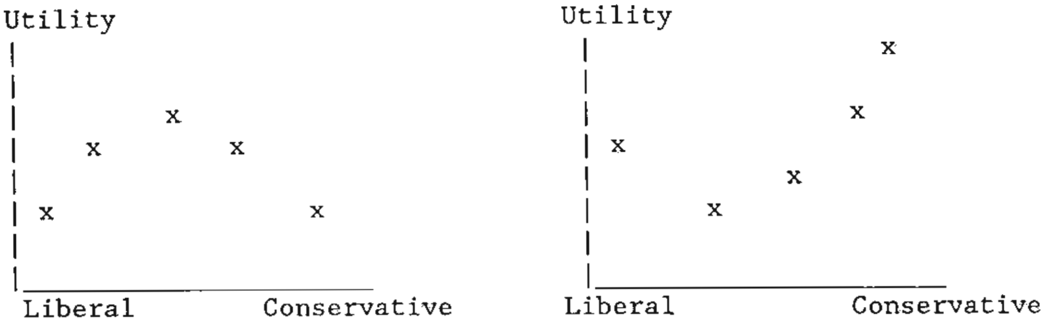
The two systems use these inputs to produce apparently different outputs. As mentioned earlier, both ACA and APM define utility at the level of the individual and perform "what if" simulations. These simulations allow the analyst to estimate what would happen to market share if market composition, or people's perception of a brand, changed. In terms of unique outputs, APM derives images of competitive structure through perceptual maps. These summarize both the perception of the competing brands and the location of ideal points within one space. Conjoint analysis, cannot directly produce maps but generates partworth functions that allow the analyst to visualize how much value an individual or a segment attaches to various attribute levels.

It needs to be stressed that these output differences are not very binding. Thus, while conjoint analysis doesn't produce spaces, the information is there and, with commonly available discriminant analysis software, one could produce perceptual maps. Further, ideal points or vectors reflecting the partworth utility functions could be positioned in this space. Thus the input to most conjoint analyses can be used to produce perceptual spaces. For its part, the information in perceptual mapping can generate individual or aggregate partworth functions. Although, as is considered next, each of these will be in the inverted "U" shape rather than the unconstrained form of the conjoint partworth functions.

Thus, although the two techniques do not differ critically with respect to their inputs or outputs, they do differ in important ways in their assumptions--and it is these assumptions that are likely to make a difference in the predictive power of the models. Two important differences are the form of the utility function and the way of weighting the different attributes.

FIGURE 1

ILLUSTRATING PARTWORTH FUNCTIONS FOR
PERCEPTUAL MAPPING AND FOR CONJOINT ANALYSIS



Symmetric Ideal Point

Anti-ideal Point

Both conjoint analysis and perceptual mapping assume that each attribute level maps into one utility value and that this utility is independent of the levels of the other attributes. Perceptual mapping further assumes that the form of this mapping is an ideal point with preferences decreasing symmetrically as one moves away from the ideal level. Conjoint analysis puts no constraint on the shape of this function. The partworth functions in Figure 1 illustrate this difference. In the left graph the ideal candidate is between liberal and conservative, and the least liked candidates are at the extreme. The right hand graph illustrates a respondent who dislikes moderate candidates relative to either strong liberals or conservatives. The important point is that this latter pattern of preferences (sometimes termed an anti-ideal) cannot be represented by an ideal point model. Only conjoint analysis could capture these utilities. Thus the utility function in conjoint analysis is more general than for perceptual mapping.

It should be emphasized, however, that this greater generality of the conjoint methodology is not always an advantage. To the extent that most of the partworths can be closely approximated by a symmetric and positive ideal, then utility functions constrained to that shape are less affected by respondent error and results in more precise utility estimates. Conjoint analysis has an advantage in the case that significant numbers of partworths are bowl-shaped or jagged and thus cannot be approximated by the positive ideal point.

A second important difference between the two techniques is the way attributes are weighted. The explicit weights which are collected by both techniques have one well-known disadvantage: they tend to overweight the less important attributes. For example, suppose a person indicates that knowledge of international affairs is moderately important in a political candidate. It has often been found that these moderatelyimportant attributes are given very little weight when actually selecting a candidate. Generally speaking, almost any attribute seems important in isolation, but its actual importance may be far less when it has to be traded off against other attributes. Conjoint copes with this problem by altering the weights to correspond to one's judgments of profiles, since in those judgments one tends to place most weight on only a few attributes.

Perceptual mapping deals with this "bias" another way. By first performing principal components on the ratings data, many attributes are replaced by a few components. These few components tend to have their highest loadings on a few attributes. Those attributes that get the most weight tend to be "central" in the sense of being most highly correlated with the others. Thus both perceptual mapping and conjoint tend to limit the impact of unimportant attributes. However, perceptual analysis uses correlations with other attributes as its criterion for reweighting while conjoint uses the judgments on profiles.

In summary, a close examination of the individual choice models of conjoint and perceptual mapping reveals that they are not that disparate. While they have somewhat different inputs and assumptions about the relationships between the attributes and utility, both present reasonable theoretical models of choice. Accordingly, the appropriate question is which system works best at predicting choice, and that is the topic of the next section.

A STUDY TO COMPARE PERCEPTUAL MAPPING AND CONJOINT ANALYSIS

To compare the two systems we built one large questionnaire that provided the inputs needed by each. We then asked a number of holdout choice questions to test their relative ability to predict each individual's choices.

TABLE 2

A STUDY OF POLITICAL PREFERENCES

8 PRESIDENTIAL HOPEFULS

REPUBLICANS	DEMOCRATS
George Bush	Jesse Jackson
Robert Dole	Gary Hart
Jack Kemp	Michael Dukakis
Pat Robertson	Albert Gore

For the category we selected the hotly contested presidential race in the fall of 1987. We selected this area because of the interest in the candidates and because we could not predict whether conjoint or perceptual mapping would do a better job. Candidates were reasonably well-known, although the race was still fluid enough that voters could conceive of candidates as "bundles of attributes." We used the four Democratic and four Republican hopefuls shown in Table 2. We needed attributes that could be considered continuous and could also be broken into discrete levels for conjoint analysis. These attributes and their levels are shown in Table 3.

TABLE 3

A STUDY OF PRESIDENTIAL ASPIRANTS

9 ATTRIBUTES, 4 LEVELS FOR EACH

Ability to get things done
FAIR, GOOD, EXCELLENT, OUTSTANDING

Ability to inspire confidence in the White House
FAIR, GOOD, EXCELLENT, OUTSTANDING

Has a clear vision of the future
LITTLE, SOMEWHAT, VERY, EXTREMELY

Lets ideals, rather than politics
dictate solutions to problems
IDEALS ALWAYS, SOMETIMES, POLITICS SOMETIMES, ALWAYS

Emphasizes workers and their welfare over
economic growth in business
STRONG ON WORKERS, MODERATE, MODERATE ON ECONOMIC, STRONG

Has a conservative political ideology
VERY CONSERVATIVE, SOMEWHAT, SOMEWHAT LIBERAL, VERY

Is a strong advocate of protectionism
STRONG FREE TRADE, MODERATE, MODERATE PROTECTIONIST, STRONG

Emphasizes world affairs over domestic affairs
STRONG ON WORLD, MODERATE, MODERATE ON DOMESTIC, STRONG

Emphasizes religious and moral values
in American life
RARELY, OCCASIONALLY, SOMETIMES, OFTEN

The inputs to each system are summarized in Table 4. For perceptual mapping we collected general importance of each of the nine attributes on a 5-point scale ranging from "Extremely important to have this" through "This really isn't important to me," to "Extremely important NOT to have this." Then respondents rated each of the 8 candidates and a ninth "ideal candidate" on the 9 attributes.

TABLE 4

A STUDY TO COMPARE CONJOINT AND PERCEPTUAL MAPPING

INPUTS

PERCEPTUAL MAPPING	CONJOINT ANALYSIS
Identification of Ideal Attribute Levels	Rankings of Attribute Levels
General Attribute Importances	Importance of Best vs. Worst Levels
Perceptions of Candidates On attributes	Tradeoffs between Profiles

We then collected the conjoint input. Each respondent rank-ordered preferences for the four levels of the attributes. Then we, or more appropriately ACA, asked for the importance of the difference between the best and the worst level of each attribute. Finally, in the core of the conjoint section, respondents indicated relative preferences for 18 pairs of profiles each defined on two attributes. An example of that tradeoff question is shown in Table 5.

TABLE 5

AN EXAMPLE OF A TRADEOFF QUESTION

```
STRONG
                  INDICATE  YOUR  PREFERENCE
PREFER
       |
 TOP   |
       |      STRONG  ON  WORKERS'  WELFARE,
  1    |
       |         VERY  WEAK  ON  ECONOMIC  GROWTH
  2    |
       |
  3    |      SOMEWHAT  LIBERAL
       |
  4    |
       |
--5----+---------------------OR------------------------------
       |
  6    |
       |
  7    |      MODERATE  ON  WORKERS'  WELFARE,
       |
  8    |          WEAK  ON  ECONOMIC  GROWTH
       |
  9    |
       |      SOMEWHAT  CONSERVATIVE
STRONG |
       |
PREFER |
       |
BOTTOM          (PRESS  NUMBER  KEY  TO  ANSWER)
```

Utility functions for each system then predicted choice for each candidate at the individual level. The conjoint estimates come directly out of ACA, although the APM model reflects a departure from its standard output. Instead of weighting deviations from the ideal levels, we used unit weights. Further, we did not use principal components to reduce the space but simply added each deviation. Thus, our version of APM is a far simpler than what which is automatically offered. Thus if there is a bias in this study it is against APM. We are currently testing a number of alternative models. However, our experience with the form of the weighting function indicates that it will make relatively little difference (one or two percentage points) in the hit rate.

These two systems were tested against 16 straw votes comparing candidates as if the election were held today. Half of these were pairs of candidates, one from each party. The second group were triples, two from one party and one from another. Each of the pairs gives one prediction, while each triple generates two. For example, if Bush is preferred in a three-way race between Bush, Kemp and Hart, then there are two predictions: Bush over Kemp, and Bush over Hart.

RESULTS

Forty-two registered voters took part in the study, about half from Columbia University and about half from Duke University. Our emphasis here is not on their preferences but on the ability of perceptual mapping and conjoint analysis to correctly predict their straw votes. This should be relatively unaffected by political orientation, although it may be affected by the high level of education in the samples.

Thus for each respondent we have 8 predictions from the pairs and 16 from the triples. The hit rates from the two are given in Table 6.

TABLE 6

HIT RATES FOR CONJOINT AND PERCEPTUAL MAPPING

PERCENT OF STRAW VOTES CORRECTLY PREDICTED:

	CONJOINT ANALYSIS	PERCEPTUAL MAPPING	n=
PAIRS	65%	76%	336
TRIPLES	63%	80%	672

PERCENT OF RESPONDENTS WITH BETTER PREDICTIONS

	PERCEPTUAL MAPPING BETTER	CONJOINT ANALYSIS BETTER	BOTH TIED
STRAW VOTES (n=42)			
PAIRS	52%	21%	27%
TRIPLES	76%	17%	7%

The results reflect an unexpectedly striking victory for perceptual mapping. For pairs there was an 11 point improvement in hit rates using perceptual mapping (65% vs. 76%), while for triples this improvement is 17 points (63% vs. 80%). Further, as indicated by the percentage of individuals who were better predicted by one system over another, perceptual mapping is more than twice as likely to achieve greater accuracy for pairs, and more than four times as likely to do so when predicting triples. The increased gain for perceptual mapping in triples indicates perceptual mapping is particularly effective in making predictions within party since those predictions were only required for the triples.

Why Perceptual Maps Made Better Predictions

The overwhelming success of perceptual mapping over conjoint analysis is all the more remarkable because of the biases against it. In this case the model was not even the normal one but a far simpler one which simply added the deviations from each individual's ideal point. However, the reasons why it won provide important insights into the predictive abilities of both models. We will examine three reasons why perceptual mapping did so well: its position in the questionnaire, the location of the ideal point questions, and finally, and most importantly, some difficulties subjects had with the conjoint questions.

The simplest hypothesis for the reduced effectiveness of the conjoint analysis is that it came after the evaluation of 8 candidates (and an ideal) on 9 attributes. These 81 judgments were not simple or easy. Thus, when respondents got to the conjoint questions they may have no longer been able to put in the required effort.

A second, somewhat more subtle, hypothesis for why mapping did well deals with the location of the ideal candidate question for each attribute. These questions occur right after one has rated all of the candidates. Thus is it easy for respondents to rate their ideal candidate as close to the candidates they like. In other words, the placement of the question makes it easy for respondents to make their ratings consistent with their choices later on. By contrast, in the conjoint task one may have forgotten whether the candidates one likes have a moderate emphasis on domestic affairs, or a strong one. This is a particular problem in this study since the adverb modifiers in the levels, such as "moderate" or "strong" have little meaning except relative to one another. In the perceptual mapping task it is easy to keep this relative ranking straight, whereas in the conjoint profiles it can be quite difficult.

A third problem, related to the second, is that respondents found the conjoint tradeoffs difficult to answer. This problem came in a number of forms. Sometimes the profile attributes were inconsistent with one another, other times the level of one affected the meaning of another, and generally respondents found it hard to evaluate a candidate from a partial description. Since these are very important issues relevant to the validity of any conjoint exercise, they are considered separately.

Attribute conflict is best illustrated in the conjoint question shown in Table 7. In that tradeoff the respondent is asked to evaluate a candidate who is good at inspiring confidence but for whom politics always dictate solutions to problems. For many respondents such a candidate is a contradiction in terms. This may result in confusion, greater error, and occasionally resentment that degrades responses to later questions.

TABLE 7

ILLUSTRATING CONFLICTING ATTRIBUTES IN A CONJOINT TASK

```
STRONG
                      INDICATE  YOUR  PREFERENCE
PREFER

 TOP   |
       |              IDEALS  ALWAYS  DICTATE
       |
  1    |                 SOLUTIONS  TO  PROBLEMS
       |
  2    |
       |              FAIR  AT  INSPIRING  CONFIDENCE
  3    |
       |                 IN  THE  WHITE  HOUSE
  4    |
       |
--5----+----------------------OR-------------------------------
       |
  6    |
       |
  7    |              POLITICS  ALWAYS  DICTATE
       |
  8    |                 SOLUTIONS  TO  PROBLEMS
       |
       |
  9    |              GOOD  AT  INSPIRING  CONFIDENCE
       |
STRONG |                 IN  THE  WHITE  HOUSE
       |
PREFER |
       |
BOTTOM                (PRESS  NUMBER  KEY  TO  ANSWER)
```

A second, and perhaps more dangerous, problem with the tradeoff questions occurs when the level of one attribute alters the meaning of another. This violates the assumption of utility independence. Consider the tradeoff given in Table 8. In that tradeoff one must choose between a strong liberal who is outstanding at getting things done and a conservative who is fair at getting things done. The problem here is in determining the value of "getting things done." Generally, it has a strong positive value. However, if it is attached to a cause in which one does not believe, then its value can be negative. Thus, the utility of one attribute level depends on the level of the other. This utility dependence violates the assumptions of the conjoint model and results in unstable conjoint estimates and poor predictions.

TABLE 8

ILLUSTRATING UTILITY DEPENDENCE

```
STRONG
                INDICATE  YOUR  PREFERENCE
PREFER

 TOP    |
        |       OUTSTANDING  AT  GETTING  THINGS
        |
 1      |           DONE
        |
 2      |       VERY  LIBERAL
        |
 3      |
        |
 4      |
        |
--5-----+----------------------OR-------------------------------
        |
 6      |
        |
 7      |       FAIR AT  GETTING  THINGS  DONE
        |
 8      |
        |
 9      |       SOMEWHAT  CONSERVATIVE
        |
STRONG  |
        |
PREFER  |
        |
BOTTOM  |     (PRESS  NUMBER  KEY  TO ANSWER)
```

A final problem with conjoint questions relates to the others. That is, a number of respondents would examine a profile and then think, "Ah, that's Jimmy Carter," or "That's George Bush." Once identified, it was easy to evaluate the profile. The important point here is the evaluation of a candidate is more primitive or basic to these respondents than is an evaluation of the candidate's attributes. In such cases evaluation does not follow from attributes, but rather the other way around. In such a context, it is perhaps no surprise that conjoint's focus on the utility of each attribute level does less well.

CONCLUSIONS

The important lessons here are methodological, dealing with the meaning of perceptual mapping and conjoint analysis and when each should be used. On these areas there are three important conclusions.

1. PERCEPTUAL MAPPING AND CONJOINT ANALYSIS ARE FORMALLY QUITE SIMILAR MODELS.

This paper began by acknowledging that until recently distinguishing the uses of perceptual mapping and conjoint analysis was on the order of distinguishing football from baseball. However, if one examines the individual choice models that underlie the new versions, they are formally quite similar to one another. Thus, while there are still product classes or problems for which the choice between systems is a "nobrainer," it is important to understand that there is an increasingly broad range of problems for which either or both are acceptable. The critical question then is to determine which system has merit for a given problem.

2. USE CONJOINT ANALYSIS WHEN ATTRIBUTES ARE NOMINAL AND PERCEIVED AS HAVING INDEPENDENT VALUE BY THE RESPONDENT.

A major difference between the formal structure of conjoint and perceptual mapping is that the former permits any shape in its partworth functions. This flexibility implies that nominal attributes, such as brand names or style types, generally can only be represented by conjoint analysis. However, if one has continuous attributes such as horsepower or durability, the increased flexibility of the conjoint functions may lead to greater error relative to perceptual mapping which constrains their shape.

There is a second, more important issue. Conjoint assumes that respondents evaluate products on the basis of each individual attribute. To the extent that this is not done, conjoint will do a poor job of predicting choice. This problem clearly occurred in our study of presidential candidates. A test of whether conjoint would be appropriate involves showing potential respondents tradeoffs and evaluating their response. If the questions are difficult, reflecting conflicting attributes, unstable utilities, or if respondents need to identify the product's identity prior to making an evaluation, then conjoint methodology is unlikely to work.

3. USE PERCEPTUAL MAPPING FOR CONTINUOUS ATTRIBUTES THAT ARE CORRELATED WITH ONE ANOTHER.

Perceptual mapping will be most successful when attributes are continuous and highly correlated with one another. The idea of continuousness stems from the need to represent utility as an ideal point within each attribute. Most continuous attributes can be represented by single-peaked ideal points, although this assumption needs to be checked for each attribute.

The high level of correlation means that a small number of dimensions will account for the large number of attribute judgments, thus permitting a great deal of information to be conveyed in a few maps. It also means that one cannot change perceptions on one attribute without changing perceptions on a group of others. These interrelations become apparent in the reduced space and account for much of their managerial value.

It must be acknowledged, however, that this adaptability comes at a price. Perceptual mapping accounts for brand preferences primarily because the inputs are structured in such a way that ideal points are placed where one's favorite brands are located. This circularity means that there will be pretty good correspondence between utilities assigned to brands and subsequent choices, just as we found in the political study. However, if one plans to use these maps to evaluate new offerings it is important that the new offerings correspond to current offerings in two senses. First the new offerings should be close to current ones. Since ideal points are close to current favorites any new offering that is dissimilar will do poorly in a choice simulator, even though it might do quite well in the marketplace. Second, and perhaps more important, the new offerings cannot upset the current correlational structure. Thus, if one alters a candidate on one attribute without changing the other attributes that are believed to be correlated with it, then regardless of what the model says, very little change will take place.

THE EFFECTS OF FAMILIARITY: WHO SHOULD RATE WHAT?

William McLauchlan
McLauchlan & Associates

"How the single stimulus is perceived is a function not so much of what it is, but rather a function of what the total set and the subset are. The properties of the total set and the subset are also the properties of the single stimulus, so we cannot understand the knowing of the single stimulus without understanding the properties of the set within which it is contained."

W. R. Garner, 1966

"Every respondent must rate at least two products. Every respondent should rate several products."

R.M. Johnson, 1987

INTRODUCTION

Background

Two of the more perplexing questions affecting the design of marketing research studies involving brand/attribute evaluations can be stated succinctly:

- Who should rate what?
- How many ratings should be obtained?

Unfortunately, the answers to these questions are routinely grounded in pragmatism rather than in comprehension of the potential consequences of the gamut of answers. Perhaps even more unfortunate is the fact that there is usually a poor understanding, on a project-by-project basis, of what those consequences might be.

On the pragmatic side, interview length is typically used to constrain the numbers of brands and attributes to be evaluated. The process often goes something like this:

> For budget reasons, we need to keep the interview length to 25 minutes. We want to ask as many questions as we can in the 25 minutes. We need 5 minutes for demos and 5 minutes for awareness

> and usage questions. That leaves 15 minutes for ratings. Assuming one rating every 5 seconds, we are limited to a total of 180 ratings. We can do either one brand rated across 180 attributes, 180 brands across one attribute, or some arbitrary combination in between, provided the total number of ratings is 180.

If the consequences of the above process were better understood, the process would likely change. Consider, instead, the following scenario:

> For budget reasons, we need to keep the interview length to 25 minutes (some things <u>**never**</u> change!). We have 15 minutes for ratings. Our pre-test work tells us that there are really only 20 unique attributes that characterize the category. Conceivably, then, we could have each respondent rate nine brands. Sadly, the category consists of 15 brands.
>
> What are the options? First, we could use an incomplete block design; one that would assure an equal number of evaluations for each brand. Second, we could use brand usage (unaided or aided or both) as the criterion for brand selection, making sure that ratings were based on experience. Third, we could use brand awareness (unaided or aided or both) to select brands; ensuring familiarity, if not experience. Fourth, we could reduce the number of attributes and have each respondent rate more than nine brands. Complicating all of this is our desire to have each respondent rate our own small-share brand.

Each of the brand selection strategies noted above is tenable. The fundamental problem is knowing which approach is best. Garner (1966) has neatly summarized the quandary. In the present framework, it is well known that the ratings a given brand receives are a function of not only the perceptions that respondents have of the brand, but also a function of the set or subset of brands that form the context for the ratings. In other words, brand rating is a comparative process.

In point of fact, the comparative process is augmented when respondents are simultaneously shown all brands to be rated on a given attribute. If the context for evaluating a given wellknown brand is a set of equally well-known competitive brands, the ratings are likely to be different than if the context for the well-known brand is lessor-known brands. (Although it may be possible to remove some context-dependent effects from ratings data, using various "normalizing" or "centering" procedures, it is also possible that the data treatment itself will further accentuate the effects.)

In order to investigate the role that brand familiarity plays in the ratings process, a small exploratory study was undertaken. The purpose of the research reported here was to look at the two key questions raised by the issues discussed above:

- What is the effect of familiarity on product ratings?
- How do results differ as respondents are required to rate increasing numbers of products and attributes?

The product category that was selected for investigation was Quick Service Restaurants (QSR); chosen largely for convenience and the high incidence of respondent qualification.

Given small sample sizes, the results of this research should not be regarded as definitive. Rather, the intent is to demonstrate that designing a study that involves brand ratings across a number of attributes requires careful consideration of both the demands made on respondents and the context in which those demands are made.

Methodology

The research employed APM* (Adaptive Perceptual Mapping); an interactive PC-based interviewing and analysis system for studies involving brand/attribute evaluations. Specifically, the study was designed to explore the effects of familiarity on the attribute ratings afforded to Quick Service Restaurants and the effects of varied numbers of ratings on the perceptual space defined by those ratings.

*APM System is a registered trademark of Sawtooth Software, Inc.

As described below, the sample was divided into five cells. In-person interviewing was used to collect the data. Respondents were intercepted and screened in one shopping mall in Cincinnati. Those individuals who qualified and agreed to participate were brought back to an enclosed room where the interview was administered on a PC. Although an interviewer was present at all times, the respondents keyed-in all answers to the computer generated questions.

Sample Composition/Size

In order to participate in the study, respondents had to meet the following qualifications:

- Visited any QSR within the past three months
- 18 years of age or older

A total of 125 interviews were conducted with respondents meeting these qualifications.

Design

The sample was divided into five cells defined according to the number of QSRs rated and the number of attributes across which ratings were obtained:

	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5
Number of QSRs					
Familiar	5	5	2	5	10
Unfamiliar	-	1	-	-	-
Number of Attributes	10	10	5	5	15

Cells 1 and 2 were included in the design to assess the impact of forcing respondents to rate an unfamiliar QSR on the ratings afforded to the familiar QSRs. Cells 1, 3, 4, and 5 were used to examine the effects of various numbers of brand/attribute ratings on the perceptual spaces defined by those ratings.

The QSRs included in the study are listed in Table 1, along with the share of visits garnered by each restaurant among the sample; the attributes appear in Table 2. All QSRs, with the exception of White Coffee Pot (the "unfamiliar" restaurant), have numerous locations in the greater Cincinnati area. White Coffee Pot is a very small chain of restaurants located primarily in the Mid-Atlantic states.

Table 1

Quick Service Restaurants

Burgers	**Share of Visits** %
McDonald's	22.0
Wendy's	8.7
Burger King	6.6
Hardee's	5.5
White Castle	4.3
Roy Rogers	3.0
Roast Beef	
Arby's	7.3
Rax	4.2
Seafood	
Long John Silver's	3.1
Captain D's	1.4
Mexican	
Taco Bell	5.2
Zantigo	2.5
Chicken	
Kentucky Fried Chicken	4.2
Famous Recipe	2.9
Chili	
Skyline Chili	4.5
Gold Star Chili	3.8
Variety	
Frisch's	9.9
White Coffee Pot*	.9

*White Coffee Pot is not located in Cincinnati

Table 2

Attributes

Has friendly employees
Has good tasting food
Has fast service
Is a better value than other restaurants
Has convenient locations
Is a restaurant that my friends like
Has attractive decor
Is clean
Has a wide variety of items
Has fresh food
Has nutritious food
Has low prices
Has drive-thru service
Is well managed
Is a restaurant I've had experience with
Has clean restrooms
Has a breakfast menu
Has food that looks appealing
Has appealing dessert items
Has the soft drink brands that I like
Uses quality ingredients
Has special meals for kids
Has light or low-calorie items
Has a salad bar

Questionnaire

The following questions were asked during the structured APM interview:

- Familiarity with each QSR (measured by frequency of visits).

- Importance of each attribute in the decision to pick a QSR to visit.

- A series of QSR attribute ratings. The selection of restaurants and attributes was governed by the cell assignment. Familiar restaurants were randomly selected from those most frequently visited, constrained by the total number of QSRs to be rated within each cell. The attributes were randomly selected from those most important to the individual respondent, also constrained by the cell definition.

- Pairwise QSR preference evaluations (measured by the proportion of visits the respondent was likely to make to given QSRs).

Interviewing Locations/Dates

Interviewing was conducted from February 5 through February 9 in Cincinnati, Ohio.

ANALYSIS

The results of the study are presented in two parts. The first section reviews the data related to the issue of familiarity and its impact of brand perceptions (Cells 1 and 2). The second section of the analysis reports results related to the impact of increased numbers of brand/attribute ratings on brand perception (Cells 1, 3, 4, and 5).

Familiarity

Respondents in cells 1 and 2 were required to evaluate their five most familiar restaurants on each of the 10 most important attributes in the QSR selection decision. In addition, respondents in cell 2 were also required to evaluate White Coffee Pot, the QSR not located in the greater Cincinnati area. Cell 1 is hereafter referred to as Familiar; cell 2 as Unfamiliar.

Reduction Of The Perceptual Space. The APM analytical module was used to reduce the perceptual space defined by the brand/attribute ratings. The software uses multiple discriminant analysis on principal components scores as a means of performing the data reduction.

Given the small sample sizes and the treatment of missing data (all missing values are set to zero), it is not surprising to note that variation explained by the decreased dimensionality is relatively low. Regardless, the first two dimensions in each cell are significant discriminators of between QSR differences. As can be seen below, however, greater variation in the ratings is explained by the first dimension in the Unfamiliar cell than is in the Familiar cell.

	Variance Explained	
Component	Familiar	Unfamiliar
	%	%
1	16.3	24.5
2	9.4	9.3
3	8.7	7.2
4	7.5	6.0
5	6.3	5.7

The difference in explained variance between the two cells on the first dimension appears to be a manifestation of the generally poor ratings given to White Coffee Pot on all attributes. The implication here is that the unfamiliarity associated with White Coffee Pot gives the appearance of greater discrimination between QSRs when, if fact, the ratings on the familiar QSRs are less variable. What is being reflected by the higher explained variance in the Unfamiliar cell on the first dimension is a linear combination of ratings for the familiar pitted against the anomalous ratings for the unfamiliar. In other words, familiar QSRs are regarded as more differentiated in the presence of other familiar restaurants and less differentiated in the presence of an unfamiliar restaurant. This finding can be seen in the following two discriminant maps.

Figure 1

5 Most Familiar Restaurants/10 Most Important Attributes

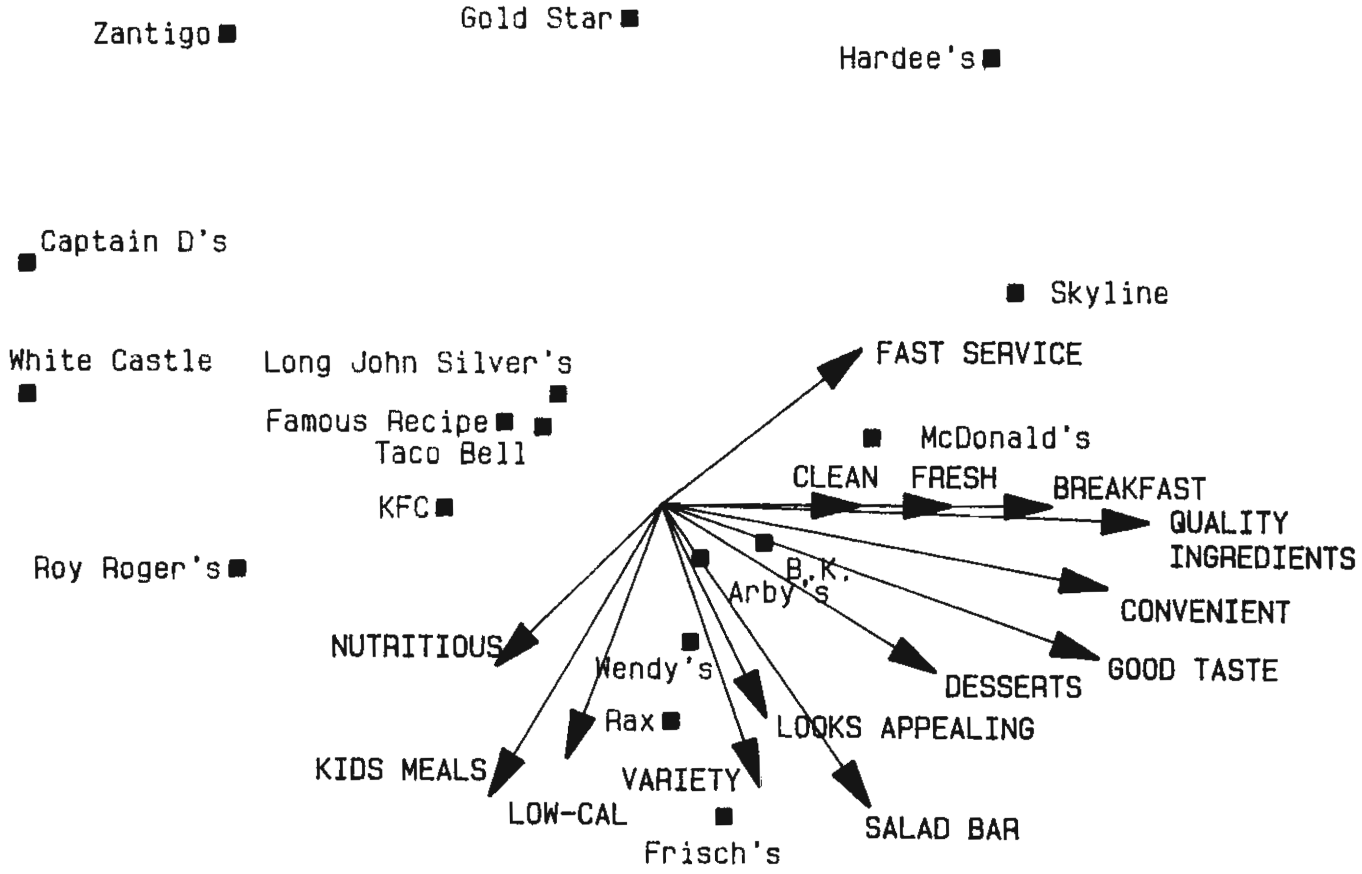


Figure 2

5 Most Familiar/1 Unfamiliar Restaurants
10 Most Important Attributes

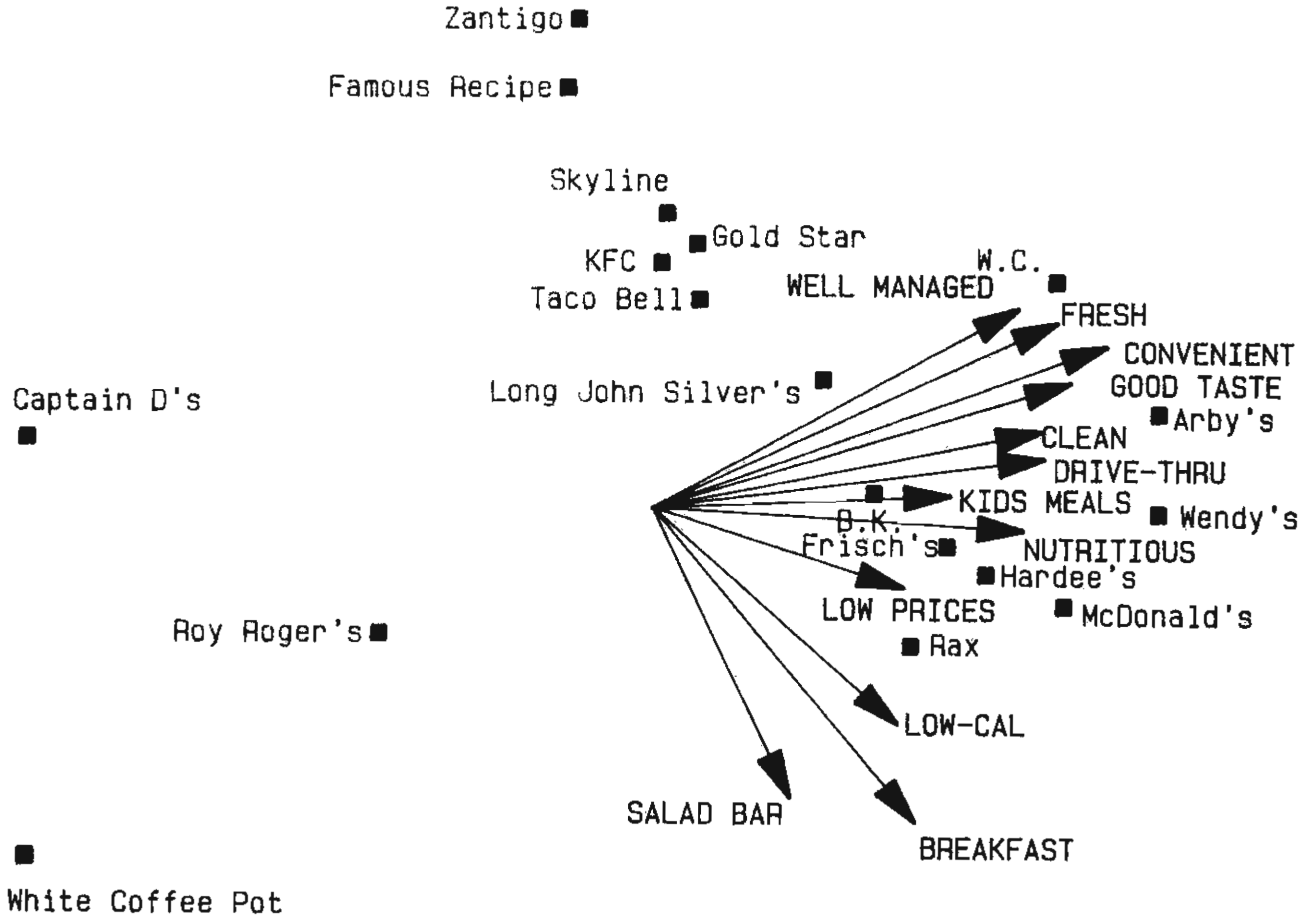


Figure 1, the map for the Familiar cell, reveals greater dispersion of both brands and attributes in the perceptual space than is observed in Figure 2, the map for the Unfamiliar cell. White Coffee Pot is a true outlier.

__Share Of Preference Simulations__. The data from the Familiar and Unfamiliar cells were used as inputs to the APM share of preference simulator. The results are arrayed in Table 3.

Table 3

Share of Preference

	Actual Share of Visits	Estimated Familiar	Estimated Unfamiliar	Estimated Unfamiliar Less W.C.P.
Burgers	%	%	%	%
McDonald's	22.0	20.0	15.8	16.9
Wendy's	8.7	9.6	10.0	10.6
Burger King	6.6	9.1	8.4	9.0
Hardee's	5.5	3.5	4.2	4.5
White Castle	4.3	4.0	4.9	5.0
Roy Rogers	3.0	3.0	1.6	1.7
Roast Beef				
Arby's	7.3	10.8	7.0	7.6
Rax	4.2	7.0	4.4	4.5
Seafood				
Long John Silver's	3.1	5.7	2.5	2.6
Captain D's	1.4	0.1	0.2	0.2
Mexican				
Taco Bell	5.2	2.7	8.2	9.0
Zantigo	2.5	1.2	1.0	1.1
Chicken				
Kentucky Fried Chicken	4.2	5.2	4.6	4.8
Famous Recipe	2.9	0.1	1.8	1.9
Chili				
Skyline Chili	4.5	3.8	8.1	8.6
Gold Star Chili	3.8	2.6	3.9	4.1
Variety				
Frisch's	9.9	11.6	7.5	7.9
White Coffee Pot*	.9	0.0	5.9	0.0

*White Coffee Pot is not located in Cincinnati

As can be seen, White Coffee Pot obtained a 5.9% share of preference among respondents living in a market where all but three individuals in the total sample indicated that they had never visited the restaurant. (Of the three respondents who reported that they had visited White Coffee Pot, two asserted that they visit every day; one professed to visit several times a week! None of the three were in the Unfamiliar cell.)

In general, estimated shares of preference in the Familiar cell more closely resemble actual share of visits data than do the share of preference estimates in the Unfamiliar cell. The QSR in the Unfamiliar cell where the simulated share of preference deviated the most from actual share of visits was McDonald's, the market leader. Further, in the presence of an unfamiliar QSR, the share of preference predictions for QSRs with greater actual shares are more apt to be overstated than are the estimates for those QSRs with smaller shares.

The Impact Of Number Of Ratings On QSR Perceptions

Cells 1, 3, 4, and 5 differed in terms of the number of QSRs rated and the number of attributes on which ratings were obtained:

	Cell 1	Cell 3	Cell 4	Cell 5
QSRs Rated	5	2	5	10
Attributes	5	5	10	15
TOTAL RATINGS	25	10	50	150

In each cell, the most familiar QSRs were rated on the most important attributes. Hereafter, cells will be referred to in terms of the combination of QSRs and attributes (e.g., cell 3 is designated 2/5).

Reduction Of The Perceptual Space. As before, the APM analytical module was used to reduce the perceptual space defined by the brand/attribute ratings in each cell. The variance explained in each cell is shown below. Components that are significant discriminators (p<.05) are indicated with an asterisk.

Component	Variance Explained			
	2/5	5/5	5/10	10/15
	%	%	%	%
1	18.5	12.9*	16.3*	25.2*
2	14.7	10.8*	9.4*	8.6*
3	11.1	8.9	8.7	7.3
4	10.2	7.8	7.5	5.8
5	9.8	7.6	6.3	5.3

The discriminant maps based on the data reduction appear in Figures 1, 3, 4, and 5.

Figure 3

2 Most Familiar Restaurants/5 Most Important Attributes

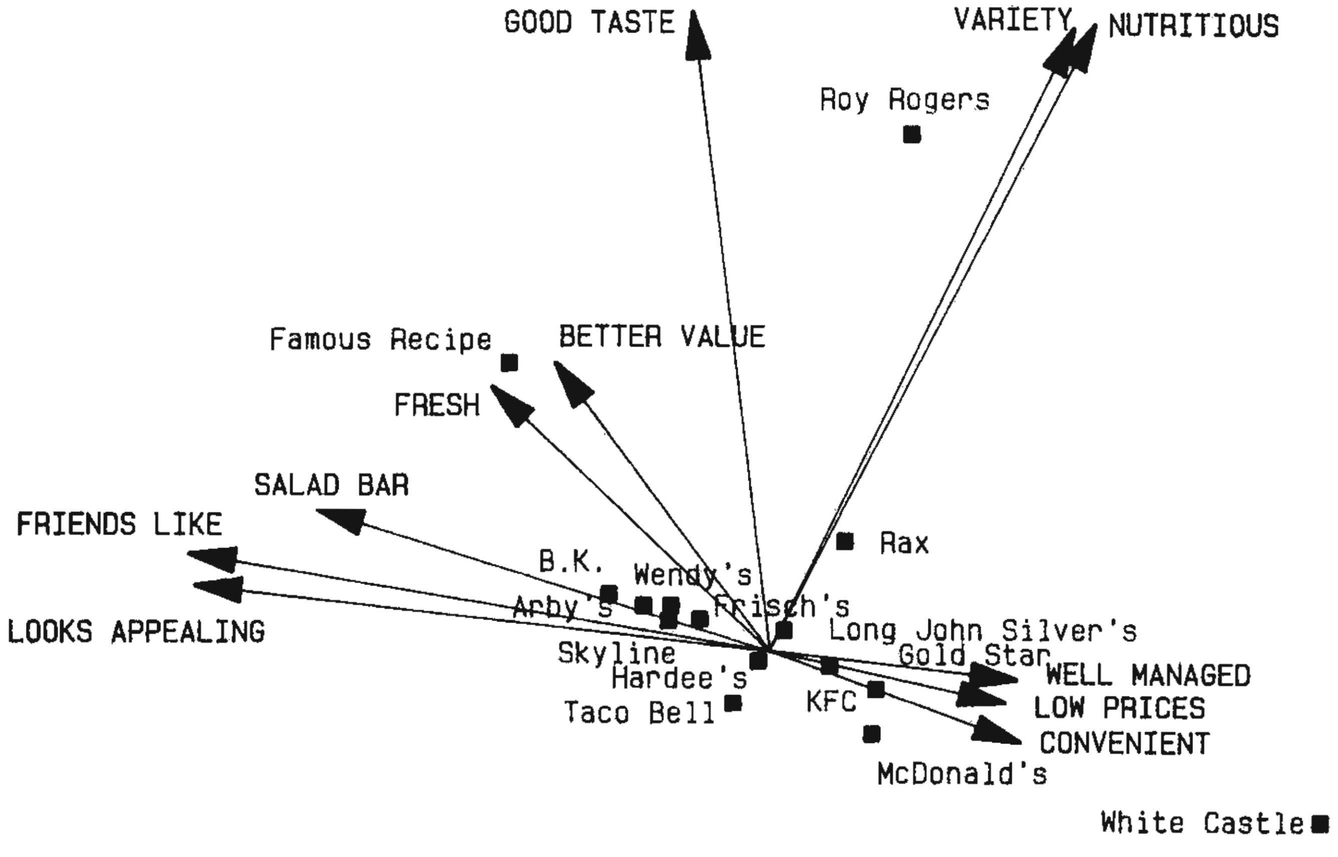


Figure 4

5 Most Familiar Restaurants/5 Most Important Attributes

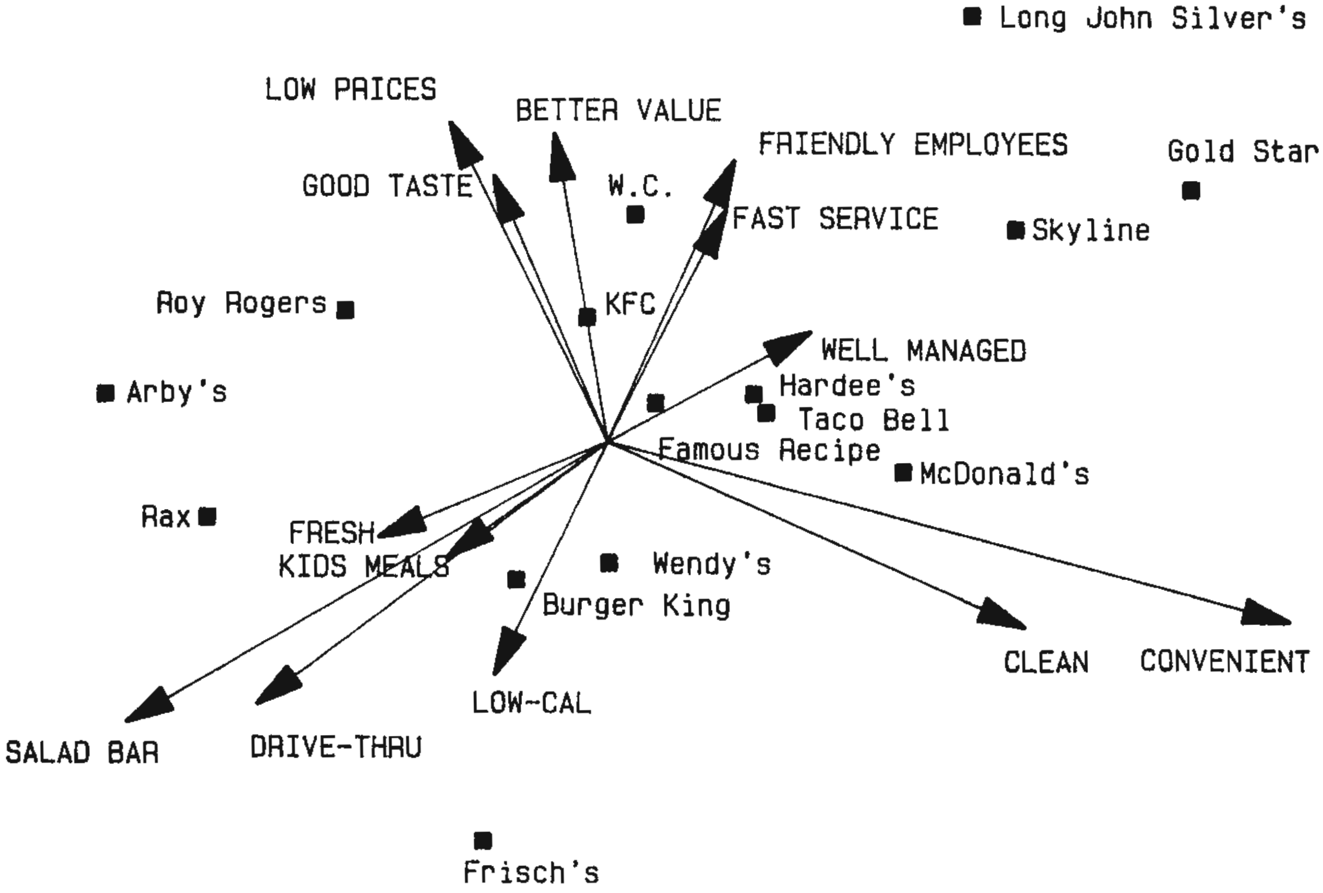
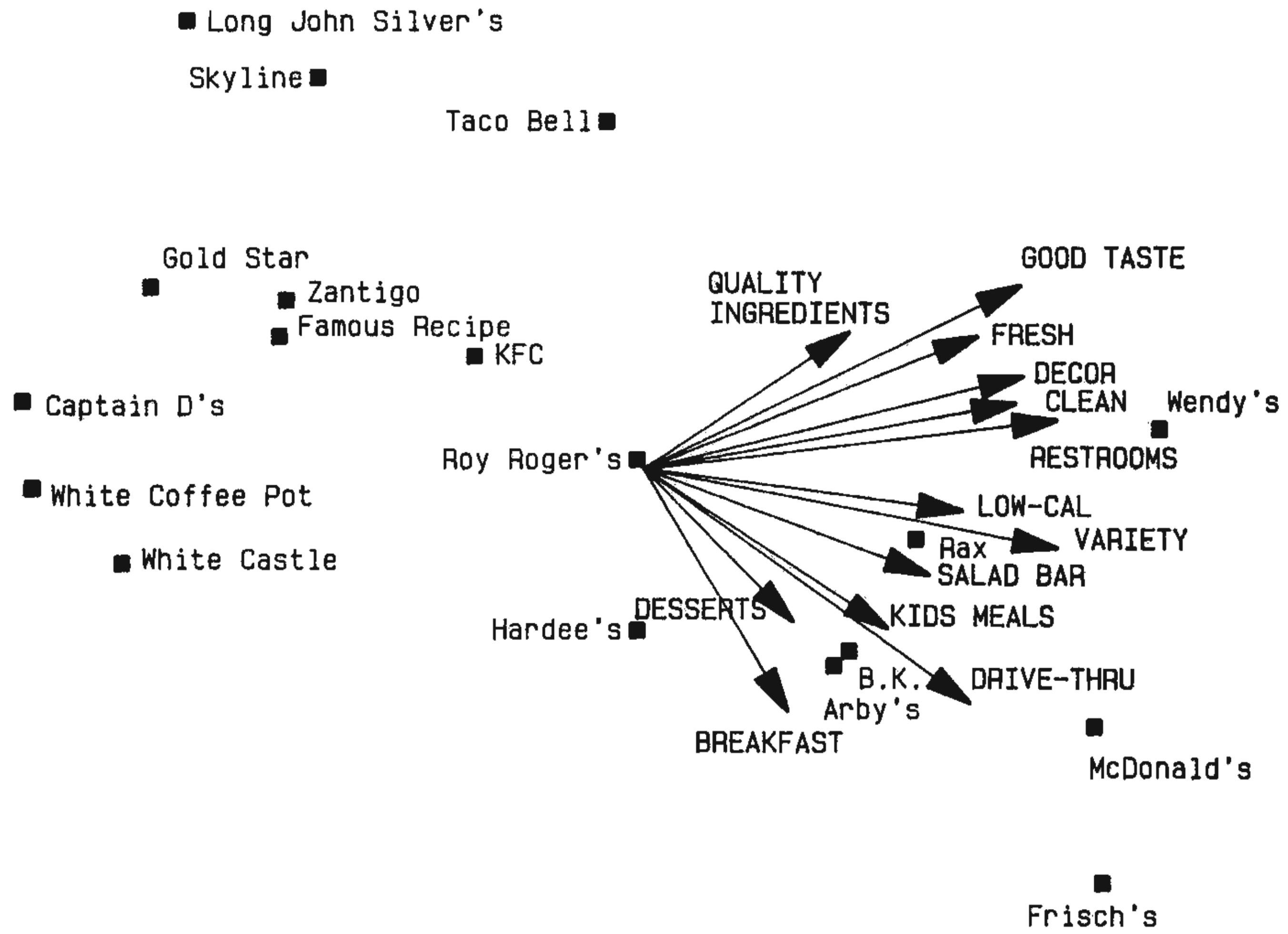


Figure 5

10 Most Familiar Restaurants/15 Most Important Attributes



Keeping in mind that the results are based on relatively small sample sizes, there are several points worth noting. First, in the 2/5 cell, none of the principal component scores discriminate between QSRs. This is another manifestation of few degrees of freedom and the treatment of missing values in the multivariate solution. (None of the univariate F ratios were significant either.) The lack of **identifiable** differentiation is further reflected in the map presented in Figure 3. With few exceptions, all QSRs cluster in the center of the perceptual space.

As the number of ratings increases, the differentiation between QSRs and the corresponding discriminant solutions give the appearance of greater statistical integrity and face validity. In point of fact, the incremental variation explained within each solution is probably biased upwards by the simple addition of ratings, independent of their content. Given the results, it seems apparent that the study could have been improved by the inclusion of a cell where even more ratings were obtained.

Share Of Preference Simulations. The results of share of preference simulations within the four experimental cells are arrayed in Table 4.

Table 4

Share of Preference

	Actual	Estimated			
	Share of Visits	2/5	5/5	5/10	10/15
Burgers	%	%	%	%	%
McDonald's	22.0	34.1	12.7	20.0	12.4
Wendy's	8.7	19.8	14.6	9.6	10.0
Burger King	6.6	3.4	6.0	9.1	6.9
Hardee's	5.5	5.6	3.7	3.5	5.9
White Castle	4.3	3.4	2.1	4.0	3.7
Roy Rogers	3.0	3.3	0.9	3.0	2.2
Roast Beef					
Arby's	7.3	8.7	7.8	10.8	9.0
Rax	4.2	2.2	9.6	7.0	5.8
Seafood					
Long John Silver's	3.1	5.4	3.8	5.7	5.1
Captain D's	1.4	0.0	0.0	0.1	0.6
Mexican					
Taco Bell	5.2	2.2	4.5	2.7	7.3
Zantigo	2.5	0.0	1.5	1.2	1.6
Chicken					
Kentucky Fried Chicken	4.2	3.0	5.3	5.2	6.4
Famous Recipe	2.9	2.2	6.7	0.1	4.4
Chili					
Skyline Chili	4.5	4.6	5.9	3.8	5.6
Gold Star Chili	3.8	1.3	5.8	2.6	2.9
Variety					
Frisch's	9.9	0.9	9.0	11.6	10.1
White Coffee Pot*	.9	0.0	0.0	0.0	0.1

*White Coffee Pot is not located in Cincinnati

If the estimated share of preference for the market leader (McDonald's) is used to gauge the "reasonableness" of the results, the 5/10 cell produces the most satisfactory solution, missing actual share by only two percentage points. On the other hand, the inadequacy of the 2/5 design becomes even more apparent. In this cell, shares of preference for McDonald's and Wendy's are significantly overestimated while the predicted share of preference for Frisch's is grossly underestimated. The 5/5 and 10/15 solutions are quite similar. Both understate the shares of preference for McDonald's and, in general, overestimate preference for most small share QSRs.

DISCUSSION

The present study was designed to measure the impact of familiarity on brand/attribute ratings and to examine how results differ as respondents are required to rate increasing numbers of products and attributes. To generalize about the design of studies that involve brand/attribute ratings, based on the results of the present research, is somewhat dangerous given the small sample sizes involved and some findings that are likely to be category specific. Still, several observations are relevant.

FAMILIARITY

The results of this study clearly demonstrate that forcing individuals to rate a completely unfamiliar brand can lead to degradation of solutions in both data reduction and in the simulation of shares of preference. It is difficult to infer from this study what the discriminant and mapping solutions would be in other product categories if highly unfamiliar products were rated. It seems apparent, however, that there would be less differentiation in respondent perceptions of well-known brands. As a consequence, study outcomes are likely to be context dependent. This may not be problematic if the effect is recognized.

It is important to note that it is not enough to collect ratings from respondents on some large number of products, including ones that are new/unfamiliar, and then "weed-out" the less familiar products in the analysis. As demonstrated by the share of preference simulations in the Unfamiliar cell, both with and without White Coffee Pot, the weeding-out process will not remove the initial contextual effects that unfamiliar brands have had on the ratings for the familiar brands. The damage is already done. In this regard, then, before embarking on a study where brand/attribute ratings are obtained on products that are either new or not well known, careful consideration needs to be given to the manner in which those ratings will be analyzed and used.

How, then, should new or unfamiliar products be dealt with in the design of brand perception studies? The answer to this question is somewhat paradoxical and poses a dilemma. In order to minimize the effects of the unfamiliar on the familiar, the unfamiliar should be made as familiar as possible. This can be accomplished to some extent through the use of well-developed concept boards, the presence of product/package prototypes, and/or home-use testing. The question that remains is whether or not it is appropriate to do so. If the intent of a given study is to measure the perceptions that individuals have of a new or unfamiliar product, it may not be fitting to collect ratings of highly familiar products at the same time.

Although not specifically addressed by the exploratory study reported here, a pertinent consideration that flows from this discussion relates to the degree of familiarity with established brands that respondents should be expected to have. The APM system, unlike traditional paper-and-pencil techniques, makes it easy to use a wide variety of brand selection criteria in the specification of who rates what. If the selection criteria that are implemented are done so in a haphazard fashion, the results could be disastrous. For example, consider a simple selection rule which specifies that each respondent will rate his/her five most familiar products. Consider, as well, a given respondent whose five most familiar products are not very familiar at all. One such respondent in a sample of several hundred will have little impact on the study outcome; more than a few could have the kind of impact reported here.

The Impact Of Number Of Ratings On Perceptions

As demonstrated here, varying the number of ratings that a respondent is required to make can produce dramatically different results. Imagine four marketing researchers; all working for a different QSR and each charged with designing a brand image study. Suppose each arrived at one of the four designs investigated here (2/5, 5/5, 5/10, 10/15). The studies are executed, analyzed and reported. A year later, three of the four researchers are gainfully employed elsewhere while one is promoted. Which one gets the promotion?

The answer to this question is, by no means, obvious. Perhaps the best standard against which the quality of the brand/attribute ratings data should be measured is face validity. In other words, if the results look like they make sense then they probably do.

Sensibility is a key word in all of this. It is not sensible to ask a respondent to sit down and do several hundred ratings, no matter how easy the task is made, and still expect to get reliable data at the end of the interview. Nor is it sensible to get only five ratings on only two brands if the category is characterized by 40 attributes and 25 brands. The results reported here suggest that 5 brands and 10 attributes make sense for QSRs. This may or may not be true if the category were either cereals or automobiles.

Final Thoughts

The intent of this paper has been to stimulate discussion. Those of you that were expecting to be given some definitive rules for selecting brands based on familiarity and for specifying some optimal number of ratings will be disappointed. Those rules do not exist.

What is important to grasp is that issues presented here are real and have a measurable impact on the reliability and validity of your data. As Garner has stated: "we cannot understand the knowing of the single stimulus without understanding the properties of the set within which it is contained." To understand the properties of a set of brands on a set of attributes requires careful pre-testing and experimentation. To do any less is to run the risk of misinterpretation, at best, and invite catastrophe at worst.

References

Garner, W. R. "To Perceive is to Know." American Psychologist, 1966, 21, 11-19.

Johnson, R. M. APM System for Adaptive Perceptual Mapping. Ketchum, ID.: Sawtooth Software, Inc., 1987.

A COMPARISON OF TECHNIQUES FOR "PERCEPTUAL MAPPING" AND AN INTRODUCTION TO BIPLOTS

Robert W. Ceurvorst
Market Facts, Inc.

"PERCEPTUAL MAPPING" IS A GENERIC TERM PERTAINING TO A CLASS OF TECHNIQUES FOR GRAPHICALLY DISPLAYING ROWS AND/OR COLUMNS OF A MATRIX.

EXAMPLES INCLUDE:

- DISCRIMINANT ANALYSIS
- CORRESPONDENCE ANALYSIS
- PRINCIPAL COMPONENTS ANALYSIS
- BIPLOTS
- "DUAL SCALING," "OPTIMAL SCALING," ETC.

NOTE: SOME TECHNIQUES ARE KNOWN BY MULTIPLE NAMES. THERE ARE MANY MORE NAMES THAN METHODS IN COMMON USE.

FEATURES COMMON TO VARIOUS PERCEPTUAL MAPPING TECHNIQUES:

- EACH ROW AND EACH COLUMN APPEARS AS A POINT OR A VECTOR ON THE PLOT.

- THE AXES OF THE PLOT REPRESENT "DIMENSIONS" THAT BEST DIFFERENTIATE AMONG THE ROWS AND/OR COLUMNS OF THE MATRIX.

    THESE DIMENSIONS ARE WEIGHTED AVERAGES OF THE ORIGINAL ROWS AND/OR COLUMNS OF THE MATRIX.

    THE OBJECTIVE IS TO PORTRAY THE MAJOR FEATURES (DIFFERENCES OR RELATIONSHIPS AMONG ROWS AND/OR COLUMNS) OF THE MATRIX IN AS FEW DIMENSIONS AS NECESSARY.

HOWEVER:

ALL PERCEPTUAL MAPPING TECHNIQUES ARE NOT CREATED EQUAL ....

SOME METHODS ARE DESIGNED FOR USE WITH PARTICULAR TYPES OF DATA, e.g.,

DISCRIMINANT ANALYSIS with MEANS;
CORRESPONDENCE ANALYSIS with FREQUENCIES OR PROPORTIONS ("CONTINGENCY TABLES").

OTHER METHODS ARE MORE GENERAL, e.g., BIPLOTS.

IT HAS BEEN SAID THAT "A MAP IS A MAP IS A MAP."

WRONG. CONSIDER:

- LENGTHS OF ROW OR COLUMN VECTORS MAY -- OR MAY NOT -- HAVE A MEANINGFUL INTERPRETATION.

- ANGLES BETWEEN VECTORS MAY -- OR MAY NOT -- HAVE A MEANINGFUL INTERPRETATION.

- DISTANCES BETWEEN ROW POINTS (OR COLUMN POINTS) MAY REPRESENT
    - ACTUAL DISTANCES
    - STANDARDIZED DISTANCES
    - ... OR NEITHER.

- YOU MAY -- OR MAY NOT -- BE ABLE TO PROJECT ROW POINTS (OR VECTORS) ONTO COLUMN VECTORS, AND VICE VERSA. IN OTHER WORDS, YOU MAY -- OR MAY NOT -- BE ABLE TO RELATE THE ROWS TO THE COLUMNS IN THE DISPLAY.

EVEN WHEN TWO OR MORE TECHNIQUES ARE APPLIED TO THE SAME DATA, THE RESULTING "MAPS" GENERALLY REQUIRE DIFFERENT INTERPRETATIONS.

EXAMPLE: THE PLOT THAT TYPICALLY ACCOMPANIES A CORRESPONDENCE ANALYSIS CANNOT BE INTERPRETED IN THE SAME WAY AS A BIPLOT BASED ON THE SAME MATRIX, EVEN THOUGH THE SAME "DIMENSIONS" ARE DEPICTED AND THE PLOTS MAY APPEAR SUPERFICIALLY SIMILAR.

NO ONE TYPE OF "MAP" CAN DISPLAY ALL OF THE FEATURES OF THE DATA IN A SINGLE PLOT.

WE MUST DECIDE WHICH FEATURES WE WANT TO PORTRAY -- e.g., ACTUAL OR STANDARDIZED DISTANCES (IF EITHER), CORRELATIONS AMONG ROWS OR AMONG COLUMNS, ETC. -- AND THEN CONSTRUCT A "MAP" ACCORDINGLY.

MORAL: PERCEPTUAL MAPS MUST BE ACCOMPANIED BY DOCUMENTATION REGARDING INTERPRETATION.

HOW ARE THE VARIOUS MAPPING TECHNIQUES RELATED?

IT HELPS TO VIEW ANY METHOD AS A 3-STEP PROCESS:

1. DETERMINE APPROPRIATE WEIGHTS, IF ANY, TO BE APPLIED TO ROWS AND TO COLUMNS.
2. NUMERICAL EXTRACTION/DEFINITION OF "DIMENSIONS" TO BE PLOTTED.
3. SCALING CONSIDERATIONS LEADING TO PLOT COORDINATES.

ALL OF THE PERCEPTUAL MAPPING TECHNIQUES MENTIONED AT THE OUTSET UTILIZE THE SAME MATHEMATICAL "TOOL" IN STEP 2: A "SINGULAR VALUE DECOMPOSITION" (SVD).

(THIS IS A TOOL FOR IDENTIFYING MUTUALLY INDEPENDENT "DIMENSIONS" THAT BEST ACCOUNT FOR DIFFERENCES AMONG ROWS AND AMONG COLUMNS.)

IT FOLLOWS THAT THE VARIOUS TECHNIQUES DIFFER IN STEPS 1 AND 3.

A COMPARISON OF VARIOUS PERCEPTUAL MAPPING TECHNIQUES

TECHNIQUE	TYPE OF MATRIX	WEIGHTS	PLOT COORDINATES
DISCRIMINANT ANALYSIS	MEANS (rows = groups, columns = variables)	ROWS: SAMPLE SIZES COLUMNS: WITHIN-GROUP COVARIANCE MATRIX	ROWS: "RAW" OR STANDARDIZED GROUP MEANS ON DIMENSIONS COLUMNS: "RAW" OR STANDARDIZED WEIGHTS, OR CORRELATIONS WITH DIMENSIONS
CORRESPONDENCE ANALYSIS	SQUARE ROOTS OF CELL CONTRIBUTIONS TO THE CHI-SQUARE STATISTIC FOR TEST OF INDEPENDENCE OF ROWS & COLS. (or similar matrix based on proportions)	ROW AND COLUMN MARGINAL FREQUENCIES (or proportions)	UNSTANDARDIZED ROW AND COLUMN "SCORES"
PRINCIPAL COMPONENTS ANALYSIS	USUALLY INDIVIDUAL OBSERVATIONS, BUT POSSIBLY FREQUENCIES, PROPORTIONS OR MEANS	OPTIONAL, DEPENDS ON THE MATRIX ANALYZED	ROWS: "RAW" OR STANDARDIZED "SCORES" ON DIMENSIONS COLUMNS "RAW" OR STANDARDIZED WEIGHTS OR CORRELATIONS WITH DIMENSIONS
BIPLOT	ANY	OPTIONAL, DEPENDS ON THE MATRIX ANALYZED	USUALLY "RAW" ROW SCORES AND STAND-STANDARDIZED COLUMN WEIGHTS OR VICE VERSA.

WHAT HAS BEEN "HOT" LATELY? CORRESPONDENCE ANALYSIS

IS IT REALLY A "WONDER DRUG FOR DATA"?

... OR IS IT ONE OF A FAMILY OF TECHNIQUES BASED ON THE SVD, AND APPLICABLE PRIMARILY TO CROSSTAB TABLES?

WHY THE HEAVY PROMOTION?

- o THE NAME "CORRESPONDENCE ANALYSIS" IS RELATIVELY NEW (i.e., "SALEABLE")
- o THE LITERATURE (ESPECIALLY IN ENGLISH) IS RELATIVELY RECENT
- o COMPUTER PROGRAMS ARE RELATIVELY EASY TO WRITE AND REQUIRE ONLY A TABLE OF FREQUENCIES OR PROPORTIONS AS INPUT. SO ...
- o SOME COMPANIES ARE SELLING THE SOFTWARE
- o SOME HAVE BOUGHT OR DEVELOPED SOFTWARE AND WANT TO USE IT

WHAT YOU SHOULD KNOW ABOUT THE USUAL CORRESPONDENCE ANALYSIS DISPLAY ...

DISTANCES BETWEEN ROW POINTS CORRESPOND TO (OR APPROXIMATE) ACTUAL DISTANCES BETWEEN ROWS -- IN A CHI-SQUARE METRIC.

DISTANCES BETWEEN COLUMN POINTS CORRESPOND TO (APPROXIMATE) ACTUAL DISTANCES BETWEEN COLUMNS -- IN A CHI-SQUARE METRIC.

THE METRIC FOR THE ROW POINTS IS NOT THE SAME AS THE METRIC FOR THE COLUMN POINTS. THEREFORE, DISTANCES BETWEEN ROW POINTS ARE NOT COMPARABLE TO DISTANCES BETWEEN COLUMN POINTS.

ROW POINTS CANNOT BE PROJECTED ONTO COLUMN VECTORS, AND COLUMN POINTS CANNOT BE PROJECTED ONTO ROW VECTORS.

IN ESSENCE, EVEN THOUGH ROWS AND COLUMNS ARE DISPLAYED JOINTLY, THEY CANNOT EASILY BE INTERPRETED JOINTLY.

THEREFORE,

WE CAN LOOK AT ROW POINTS, IGNORING THE COLUMNS.

WE CAN LOOK AT COLUMN POINTS, IGNORING THE ROWS.

WE CAN BE MISLED IF WE TRY TO RELATE ROW POINTS TO COLUMN POINTS.

IN OUR EXPERIENCE, THE PRIMARY REASON FOR DISPLAYING DATA VIA A "PERCEPTUAL MAP" IS TO EXPLORE THE RELATIONSHIPS BETWEEN ROWS AND COLUMNS.

IF ROWS AND COLUMNS REPRESENT TWO CLASSIFICATIONS,

- ARE THEY INDEPENDENT?
- IF NOT -- IF THEY INTERACT -- THE DISPLAY MUST ALLOW US TO INTERPRET THE INTERACTION.

IF ROWS ARE GROUPS (DEMOGRAPHIC GROUPS, ATTITUDINAL/ BEHAVIORAL CLUSTERS, BRANDS, ETC.) AND COLUMNS ARE NUMERIC VARIABLES (RATINGS, ETC.),

- HOW DO THE GROUPS DIFFER <u>IN TERMS OF</u> THE VARIABLES?
- PERHAPS, IS THERE A GROUP-BY-VARIABLE INTERACTION? IF SO, DISPLAY IT.

<u>THESE TYPES OF QUESTIONS REQUIRE A JOINT INTERPRETATION OF ROWS AND COLUMNS IN A DISPLAY</u>.

A TYPE OF DISPLAY SPECIFICALLY DESIGNED TO PERMIT A JOINT INTERPRETATION OF ROWS AND COLUMNS IS A

BIPLOT.

IN A BIPLOT, THE SCALINGS FOR ROW AND COLUMN POINTS (OR VECTORS) ARE CHOSEN SO THAT THE ORIGINAL MATRIX MAY BE RECONSTRUCTED FROM THE DISPLAY, IF THE "FIT" IS GOOD.

IN OTHER WORDS, ONE CAN RECOVER THE ROW VALUES IN EACH COLUMN BY PROJECTING ROW POINTS ONTO COLUMN VECTORS, AND VICE VERSA.

THIS FEATURE IS UNIQUE TO BIPLOTS.

WHY HAVE BIPLOTS RECEIVED LESS ATTENTION IN THE MARKETING RESEARCH COMMUNITY THAN CORRESPONDENCE ANALYSIS?

> PRIMARILY BECAUSE THE ARTICLES CONCERNING BIPLOTS HAVE APPEARED IN MORE TECHNICAL JOURNALS NOT WIDELY READ AMONG MARKETING RESEARCHERS (e.g., Technometrics, Biometrika).

SOME FEATURES THAT MAKE BIPLOTS ATTRACTIVE:

- THE TECHNIQUE IS VERY GENERAL --

  ANY RECTANGULAR MATRIX CAN BE DISPLAYED.

  WITH AN APPROPRIATE CHOICE OF MATRICES AND ROW/COLUMN WEIGHTS, BIPLOTS CAN BE USED TO DISPLAY RESULTS OF:

  - DISCRIMINANT ANALYSIS/MANOVA

    -- including "concentration ellipses" around group points (Gabriel, 1980)

  - PRINCIPAL COMPONENTS ANALYSIS

    (any suitable matrix) (Gabriel, 1971)

  - CORRESPONDENCE ANALYSIS (Goodman, 1985)

- BIPLOTS CAN BE USED TO AID IN "MODEL FITTING"

  PATTERNS OF ROW AND COLUMN POINTS REVEAL
  ADDITIVITY OF ROW AND COLUMN EFFECTS,
  "1 D.F. FOR NON-ADDITIVITY" MODELS,
  ROW-BY-COLUMN INTERACTIONS, ETC.
  (Bradu & Gabriel, 1978)

SOME FEATURES OF BIPLOTS (continued):

- FLEXIBILITY IN CHOOSING WHICH FEATURES OF THE DATA TO DISPLAY ...

  ... BECAUSE THERE IS MORE THAN ONE WAY TO BIPLOT ANY GIVEN MATRIX!

  DIFFERENT BIPLOTS RESULT FROM DIFFERENT SCALINGS (CHOICE OF METRICS) FOR ROWS OR COLUMNS.

  BUT ALL BIPLOTS OF A GIVEN MATRIX WILL ALLOW THE RECOVERY OF THE INDIVIDUAL MATRIX ELEMENTS IF THE "FIT" IS GOOD.

THE MATHEMATICS OF BIPLOTS:

LET X BE A MATRIX TO BE DISPLAYED.

THE SVD OF X "FACTORS" X AS: $X = L\ D\ R'$

WHERE L = "LEFT" EIGENVECTORS OF X (EIGENVECTORS OF XX')

R = "RIGHT" EIGENVECTORS OF X (EIGENVECTORS OF $X'X$)

D = DIAGONAL MATRIX OF "SINGULAR VALUES" (STANDARD DEVIATIONS OF THE DIMENSIONS

(WITHOUT LOSS OF GENERALITY, LET THE "DIMENSIONS" BE ARRANGED SUCH THAT THE DIAGONAL ELEMENTS OF D ARE IN DECREASING ORDER. THEN, THE BEST M-DIMENSIONAL FIT IS ACHIEVED BY RETAINING THE FIRST M DIAGONAL ELEMENTS OF D AND THE CORRESPONDING COLUMNS OF L AND R.)

WE CAN RE-EXPRESS THE SVD AS $X = L\ D^a\ D^b\ R'$, WHERE $a + b = 1$.

A BIPLOT IS A GRAPHIC DISPLAY OF

$L\ D^a$, COORDINATES OF THE ROW POINTS

AND $R\ D^b$, COORDINATES OF THE COLUMN POINTS.

SINCE $a + b = 1$, THE INNER PRODUCT OF THE ROW AND COLUMN COORDINATES -- THAT IS, THE PROJECTIONS OF ROW POINTS ONTO COLUMNS OR VICE VERSA -- RECOVERS X, EXACTLY IF ALL DIMENSIONS WITH NONZERO VARIANCE ARE RETAINED, AND APPROXIMATELY OTHERWISE.

DIFFERENT BIPLOTS RESULT FROM DIFFERENT CHOICES OF a AND b.

BUT THE PROJECTIONS ARE THE SAME REGARDLESS OF THE CHOICE.

SOME OBVIOUS CHOICES FOR a AND b INCLUDE:

$a = 0$, $b = 1$:	RELATIVE DISPERSION OF THE DIMENSIONS IS REFLECTED IN THE SPREAD OF THE COLUMN POINTS: LENGTHS OF COLUMN VECTORS (MEASURED FROM THE ORIGIN) REPRESENT STANDARD DEVIATIONS OF THE COLUMNS (THE MOST DISCRIMINATING COLUMNS HAVE THE LONGEST VECTORS): COSINES OF ANGLES BETWEEN COLUMN VECTORS ARE CORRELATIONS BETWEEN COLUMNS OF X. ROW POINTS ARE STANDARIZED -- I.E., ROW POINTS ARE PLOTTED IN STANDARD DEVIATION UNITS.
$a = 1$, $b = 0$:	AS ABOVE, INTERCHANGING ROLE OF ROWS AND COLUMNS. IF ROWS ARE "GROUPS," THIS CHOICE DEPICTS "RAW" RATHER THAN "STANDARDIZED" DISTANCES BETWEEN GROUPS.
$a = 1/2$, $b = 1/2$:	ROW AND COLUMN POINTS ARE TREATED "SYMMETRICALLY," ALTHOUGH DISTANCES ARE NEITHER RAW NOR STANDARDIZED AND ANGLES DO NOT REPRESENT CORRELATIONS.

FOR EXAMPLE,

SAY WE HAVE A MATRIX OF MEANS, WHERE

ROWS REPRESENT BRANDS OF BEER;

COLUMNS ARE ATTRIBUTES ON WHICH THESE BRANDS WERE RATED BY BEER DRINKERS.

HOW MIGHT WE ANALYZE THESE DATA?

- SUBTRACT THE GRAND MEAN (ACROSS ALL ROWS AND COLUMNS) SO WE PRESERVE DIFFERENCES AMONG BRANDS, AMONG ATTRIBUTES, AND THE BRAND-BY-ATTRIBUTE INTERACTION; or

- SUBTRACT OUT THE ATTRIBUTE MEANS, TO "ZOOM IN" ON BRAND DIFFERENCES, AS IN ANALYSIS OF VARIANCE OR DISCRIMINANT ANALYSIS; or

- SUBTRACT OUT ROW AND COLUMN MARGINAL MEANS TO EXPLICITLY FOCUS ON THE BRAND-BY-ATTRIBUTE INTERACTION.

BIPLOT THE RESULTING MATRIX.

BRAND MEANS

VARIABLE	BUDWEISER	COORS	HEINEKEN	MEISTER	MICHELOB	MILLER	OLD MILW	STROHS	MEAN	SD
FOOD	4.76	4.85	4.74	4.18	5.04	4.94	4.36	4.51	4.67	0.30
SPECIAL	3.87	4.33	5.06	3.79	4.88	4.33	3.65	4.06	4.25	0.51
THIRST	4.88	4.93	4.63	4.22	5.03	5.01	4.48	4.64	4.73	0.29
BLUE COL	4.28	4.19	3.73	4.05	4.07	4.28	4.42	4.31	4.17	0.22
TASTE	5.01	5.07	4.95	4.19	5.32	5.17	4.41	4.66	4.85	0.39
VALUE	4.68	4.51	3.90	4.33	4.47	4.73	4.78	4.45	4.48	0.28
SUCCESS	3.82	4.06	4.63	3.75	4.55	4.09	3.66	4.02	4.07	0.35
RUGGED	4.57	4.39	3.81	3.91	4.12	4.23	4.25	4.31	4.20	0.25
WOMEN	3.91	4.29	3.97	3.75	4.46	4.49	3.83	3.96	4.08	0.29
PREMIUM	4.75	4.88	5.31	3.90	5.29	4.89	3.94	4.38	4.67	0.55
IMPRESS	3.34	3.84	4.53	3.34	4.48	3.81	3.15	3.58	3.76	0.52
MEAN	4.35	4.49	4.48	3.95	4.70	4.54	4.08	4.26	4.36	
SD	0.54	0.40	0.55	0.29	0.44	0.43	0.48	0.33		

FOR THE SAKE OF ILLUSTRATION,

WE WILL BIPLOT A MATRIX FROM WHICH THE ATTRIBUTE MEANS HAVE BEEN SUBTRACTED OUT.

SAMPLE SIZES HERE WERE APPROXIMATELY EQUAL (400 each), SO, FOR SIMPLICITY, WE WON'T USE ANY WEIGHTS.

96% OF THE VARIATION AMONG BRANDS CAN BE CAPTURED IN THE FIRST 2 PRINCIPAL COMPONENTS, SO THE 2-DIMENSIONAL FIT IS QUITE GOOD.

> (IF THE FIT HAD BEEN POOR -- e.g., LESS THAN 80% OF THE VARIANCE WAS ACCOUNTED FOR, WE COULD HAVE EXTRACTED AS MANY DIMENSIONS AS NEEDED. HOWEVER, HIGH-DIMENSIONAL DATA IS NOT EASY TO DEPICT WITH ANY TYPE OF PERCEPTUAL MAP.)

HERE ARE TWO BIPLOTS OF THESE DATA ....

BRAND MEANS

VARIABLE	BUDWEISER	COORS	HEINEKEN	MEISTER	MICHELOB	MILLER	OLD MILW	STROHS	MEAN	SD
FOOD	0.09	0.18	0.07	-0.49	0.36	0.27	-0.32	-0.16	0.00	0.30
SPECIAL	-0.38	0.08	0.82	-0.45	0.63	0.09	-0.60	-0.19	0.00	0.51
THIRST	0.15	0.20	-0.10	-0.51	0.31	0.29	-0.25	-0.09	0.00	0.29
BLUE COL	0.12	0.02	-0.44	-0.11	-0.10	0.11	0.26	0.14	0.00	0.22
TASTE	0.17	0.22	0.10	-0.65	0.47	0.32	-0.44	-0.18	0.00	0.39
VALUE	0.20	0.03	-0.58	-0.15	-0.01	0.25	0.30	-0.03	0.00	0.28
SUCCESS	-0.25	-0.01	0.55	-0.33	0.48	0.02	-0.41	-0.06	0.00	0.35
RUGGED	0.37	0.19	-0.39	-0.29	-0.08	0.03	0.05	0.11	0.00	0.25
WOMEN	-0.18	0.21	-0.11	-0.33	0.38	0.41	-0.26	-0.12	0.00	0.29
PREMIUM	0.08	0.21	0.64	-0.77	0.62	0.22	-0.73	-0.29	0.00	0.55
IMPRESS	-0.42	0.09	0.77	-0.42	0.72	0.05	-0.61	-0.17	0.00	0.52
MEAN	-0.00	0.13	0.12	-0.41	0.34	0.19	-0.27	-0.09	0.00	
SD	0.26	0.09	0.50	0.20	0.29	0.13	0.34	0.13		

BIPLOT WITH STANDARDIZED BRAND POINTS:

- THE RELATIVE DISPERSION OF THE TWO DIMENSIONS IS REFLECTED IN THE ATTRIBUTE VECTORS --

  (MORE HORIZONTAL THAN VERTICAL SPREAD)

- LENGTHS OF ATTRIBUTE VECTORS ARE PROPORTIONAL TO STANDARD DEVIATIONS OF THE BRAND MEANS ON THOSE ATTRIBUTES, i.e.,

  LENGTHS OF ATTRIBUTE VECTORS ARE DIRECT MEASURES OF RELATIVE DISCRIMINATION POWER OF ATTRIBUTES.

- COSINES OF ANGLES BETWEEN ATTRIBUTE VECTORS ARE CORRELATIONS, SO WE CAN EASILY SEE HOW ATTRIBUTES ARE RELATED IN DISTINGUISHING AMONG THE BRANDS.

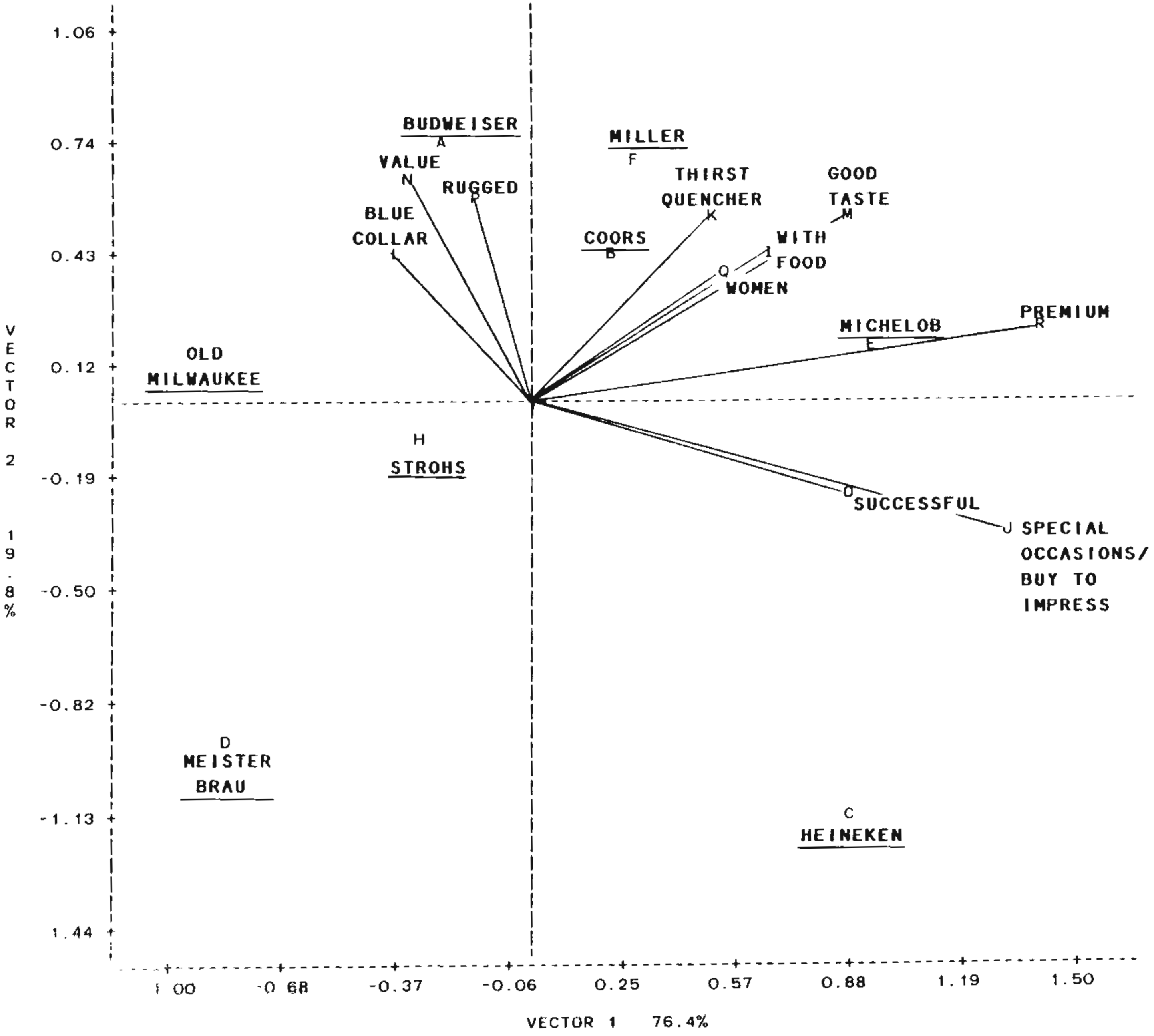
  positive correlation = small (<90°) angle
  negative correlation = large (>90°) angle
  zero correlation = 90° angle

BIPLOT WITH STANDARDIZED BRAND POINTS: (continued)

- BRAND POINTS ARE "STANDARDIZED," SO THAT THE HORIZONTAL AND VERTICAL SPREAD ARE THE SAME. THIS IS ANALOGOUS TO CALCULATING STANDARDIZED "FACTOR SCORES," AND IS NECESSARY TO INSURE THAT PROJECTIONS OF BRANDS ONTO ATTRIBUTE VECTORS ARE VALID.

- SINCE THE FIT IS GOOD, THE RELATIVE POSITIONS OF THE BRANDS ALONG ANY ATTRIBUTE VECTOR WILL CLOSELY MATCH ACTUAL BRAND POSITIONS IN THE ORIGINAL MATRIX.

GRAPHIC DISPLAY OF BRAND DIFFERENCES
BIPLOT WITH STANDARDIZED BRAND POINTS



BIPLOT WITH STANDARDIZED VARIABLE POINTS:

- THE RELATIVE DISPERSION OF THE TWO DIMENSIONS IS REFLECTED IN THE POSITIONS OF THE BRAND POINTS --

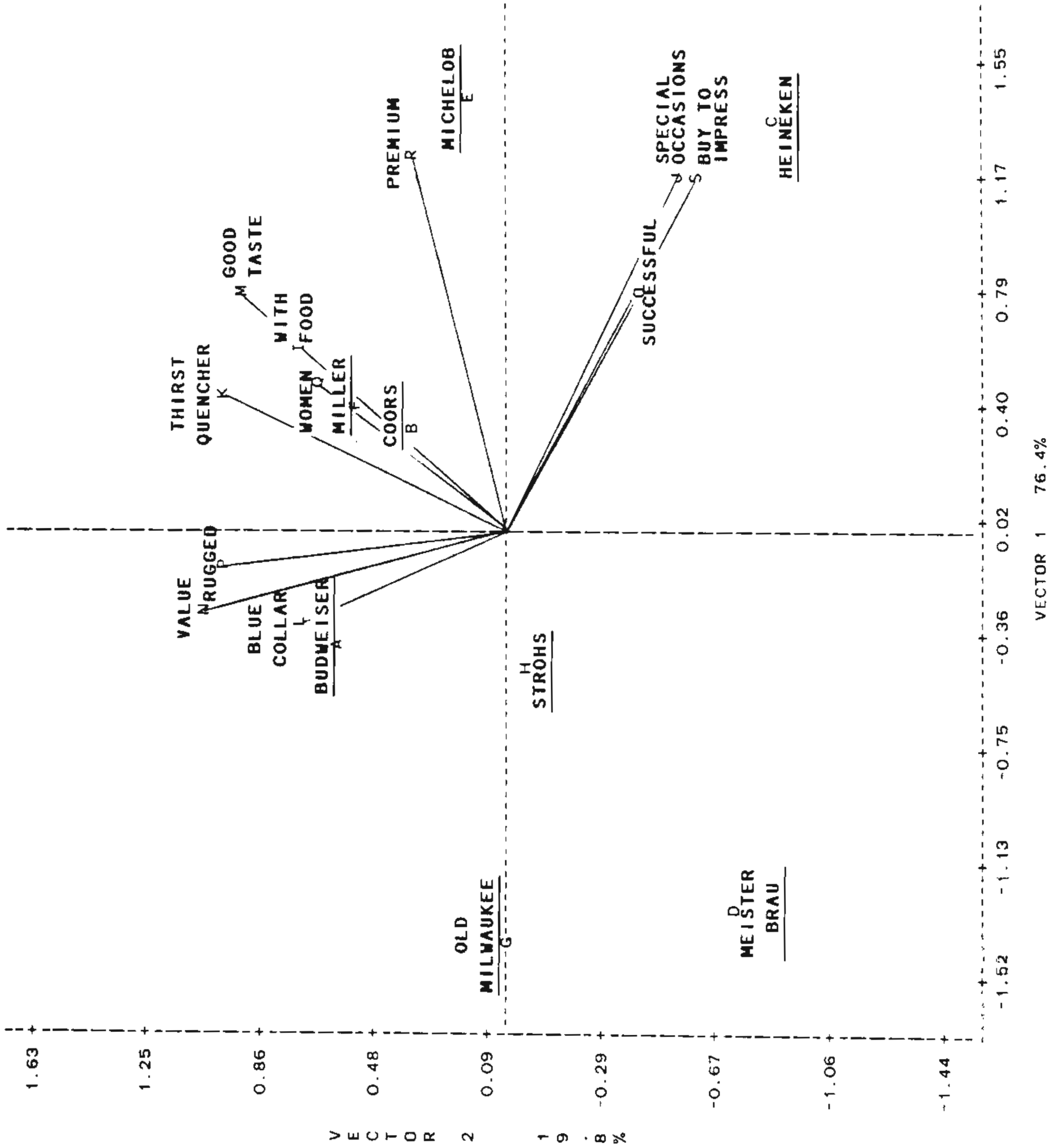
  (MORE HORIZONTAL THAN VERTICAL SPREAD)

- DISTANCES BETWEEN BRAND POINTS CLOSELY REFLECT "RAW" (RATHER THAN STANDARDIZED) DISTANCES BETWEEN BRANDS.

- ATTRIBUTE VECTORS ARE POSITIONED SO THAT PROJECTIONS OF BRAND POINTS ONTO THESE VECTORS ARE VALID (AND WILL APPROXIMATE THE ELEMENTS OF THE MATRIX DISPLAYED).

- BUT LENGTHS OF ATTRIBUTE VECTORS ARE NOT STANDARD DEVIATIONS AND ANGLES ARE NOT CORRELATIONS.

GRAPHIC DISPLAY OF BRAND DIFFERENCES
BIPLOT WITH STANDARDIZED VARIABLE POINTS



THE TWO BIPLOTS ARE VERY SIMILAR, EVEN THOUGH THE FIRST DIMENSION ACCOUNTS FOR ABOUT 4 TIMES AS MUCH VARIATION AS THE SECOND (76% vs. 20%).

THE EYE, THOUGH, SEES THE RELATIVE STANDARD DEVIATIONS, A RATIO OF ABOUT 2:1.

UNLESS THE FIRST DIMENSION IS VERY DOMINANT -- INDICATING A UNI-DIMENSIONAL MATRIX, THESE TWO TYPES OF BIPLOTS WILL APPEAR VERY SIMILAR.

THE CHOICE BETWEEN THEM HINGES ON WHETHER YOU ARE MORE INTERESTED IN DEPICTING "RAW" DISTANCES AMONG BRANDS OR STANDARD DEVIATIONS OF AND CORRELATIONS AMONG ATTRIBUTES.

FINALLY, OTHER BIPLOT SCALINGS ARE POSSIBLE, ALTHOUGH THE TWO PRESENTED HERE SEEM TO BE THE MOST USEFUL IN TERMS OF REPRESENTING DISTANCES, CORRELATIONS, ETC.

REFERENCES

Bradu, D. and Gabriel, K.R., "The Biplot as a Diagnostic Tool for Models of Two-Way Tables," Technometrics, 20(1), 1978, 47-68.

Gabriel, K.R., "The Biplot Graphic Display of Matrices with Application to Principal Component Analysis," Biometrika, 58(3), 453-467.

Gabriel, K.R., "Biplot Display of Multivariate Matrices for Inspection of Data and Diagnosis," Paper presented at Symposium, "Looking at Multivariate Data," Sheffield, England, 1980.

Goodman, L.R., "Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-Linear Models Approach to the Analysis of Contingency Tables," Paper presented at 45th Session of the International Statistical Institute, Amsterdam, Netherlands, 1985.

OTHER SOURCES OF INFORMATION CONCERNING THE SVD, DUAL/OPTIMAL SCALING, CORRESPONDENCE ANALYSIS AND BIPLOTS:

Greenacre, M.J., Theory and Applications of Correspondence Analysis. London: Academic Press Inc., 1984.

Nishisato, S., Analysis of Categorical Data: Dual Scaling and Its Applications. Toronto: University of Toronto Press, 1980.

CORRELATIONS

FOOD	SPECIAL	THIRST	BLUE COL	TASTE	VALUE	
1.00	.69	.96	-.17	1.00	.05	GOOD WITH FOOD
.69	1.00	.48	-.76	.69	-.64	FOR SPECIAL OCCASIONS
.96	.48	1.00	.11	.96	.30	GOOD THIRST QUENCHER
-.17	-.76	.11	1.00	-.17	.93	FOR BLUE COLLAR TYPES
1.00	.69	.96	-.17	1.00	.03	GOOD TASTING
.05	-.64	.30	.93	.03	1.00	GOOD VALUE FOR THE MONEY
.67	.99	.46	-.74	.67	-.64	FOR SUCCESSFUL PEOPLE
.25	-.46	.49	.77	.27	.76	FOR RUGGED, OUTDOOR MEN
.87	.57	.86	-.01	.84	.19	FOR WOMEN
.91	.90	.76	-.54	.91	-.36	PREMIUM, HIGH QUALITY
.68	1.00	.47	-.75	.68	-.63	BUY TO IMPRESS

SUCCESS	RUGGED	WOMEN	PREMIUM	IMPRESS	
.67	.25	.87	.91	.68	GOOD WITH FOOD
.99	-.46	.57	.90	1.00	FOR SPECIAL OCCASIONS
.46	.49	.86	.76	.47	GOOD THIRST QUENCHER
-.74	.77	-.01	-.54	-.75	FOR BLUE COLLAR TYPES
.67	.27	.84	.91	.68	GOOD TASTING
-.64	.76	.19	-.36	-.63	GOOD VALUE FOR THE MONEY
1.00	-.45	.54	.88	.99	FOR SUCCESSFUL PEOPLE
-.45	1.00	.14	-.07	-.47	FOR RUGGED, OUTDOOR MEN
.54	.14	1.00	.69	.58	FOR WOMEN
.88	-.07	.69	1.00	.88	PREMIUM, HIGH QUALITY
.99	-.47	.58	.88	1.00	BUY TO IMPRESS

```
                              DISTANCES BETWEEN BRANDS

DISTANCE        BUDWEISER  COORS  HEINEKEN  MEISTER  MICHELOB  MILLER  OLD MILW  STROHS

BUDWEISER       0.000      0.878  2.317     1.673    1.981     1.028   1.261     0.800
COORS           0.878      0.000  1.629     1.972    1.156     0.384   1.796     0.942
HEINEKEN        2.317      1.629  0.000     2.684    1.099     1.786   2.939     2.040
MEISTER BRAU    1.673      1.972  2.684     0.000    2.878     2.149   0.815     1.105
MICHELOB        1.981      1.156  1.099     2.878    0.000     1.128   2.848     1.944
MILLER          1.028      0.384  1.786     2.149    1.128     0.000   1.910     1.152
OLD MILWAUKEE   1.261      1.796  2.939     0.815    2.848     1.910   0.000     0.966
STROHS          0.800      0.942  2.040     1.105    1.944     1.152   0.966     0.000
```

PREPARING DATA FOR MAPPING

Roger Buldain
Burke Marketing Research

Mapping algorithms generally require the data from multiple respondents' evaluations of multiple events on multiple attributes. Regardless of the form these three components may take:

- o Respondents = Cases, consumers, voters, subjects, etc.
- o Events = Objects, brands, nations, candidates, etc.
- o Attributes = Similarity ratings, agreement ratings, "brand personality" ratings, characteristic checklists, etc.

The point of the mapping exercise is to reduce the number of dimensions used to represent the data into a smaller number that lends itself to graphic representation. The goal of mapping is to achieve the parsimony of "one picture is worth a thousand words."

A number of difficulties may be encountered in achieving this goal, however. Perhaps one of the greatest encountered by lay researchers is the influence of "learned opinion." The old saw "There are as many ways to calculate statistical relationships as there are ways to skin a cat," holds true for mapping as well as for the Analysis of Variance. Research is a market just like any other, and new mapping products that offer a point of differentiation are always much in demand.

Automobiles and political candidates are often used in examples or demonstrations of various mapping techniques, and accordingly, I'll try to incorporate them in this brief presentation. Let's think of perceptual maps as automobiles for a moment. As finished products, autos have different body styles, engine displacements, carburetion, electrical systems, power trains, interiors, and so on. Construction methods for cars may vary just as they may for maps and the consequences will define the end product. (For example, one may think of Adaptive Perceptual Mapping (APM) as a mapping innovation comparable to unibody construction in autos.) Design approaches may vary as well, calling for welds here, rivets there, the dimensions resulting from factor analysis in one shop (e.g., Burke's MIRROR), discriminant function analysis in another (e.g., Sawtooth's APM), and simple matrix reduction techniques in a third (e.g., MDPREF).

With so many sexy alternatives available, some of us have been forced to step back from the "how" of mapping in order to differentiate ourselves on the "what." The subject of this presentation is "Preparing Data for Mapping" and its matter is akin to the metallurgy that defines the composition of materials before they are cast into the shapes that define the various components of a car.

A data cube is often used (c.f., Rummell, 1970) to represent the elements of dimensional analysis (Figure 1). One obvious problem with this cube for mapping is that it is three dimensional, while maps have a tendency to be two-dimensional at least. A second problem lies in the variabilility of each of the elements of each dimensions. Respondents vary in their evaluations, objects vary in how they are evaluated, and characteristics vary in that they may be evaluated differently depending on respondent and object.

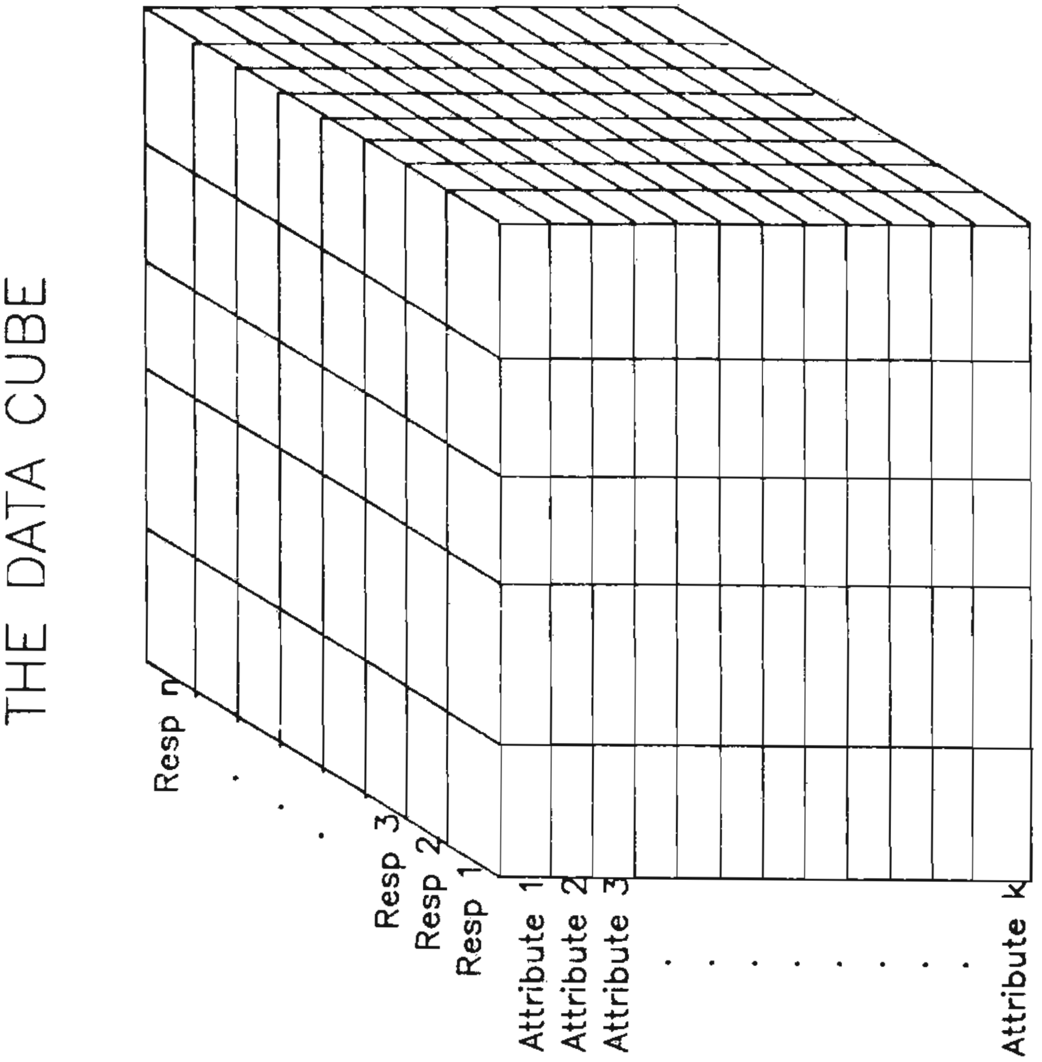

THE DATA CUBE
Resp n
Resp 3
Resp 2
Resp 1
Attribute 1
Attribute 2
Attribute 3
Attribute k
Object 1
Object 2
...
...
Object j

FIGURE 1

Let's take one cell out of this cube and look at some of the components of variability for a single respondent's evaluation of one object on one attribute (Figure 2). The common variance described is applicable to all respondents, objects, and attributes. The specific is tied to how this respondent perceives this object on this attribute. Errors are tied to random influences on how measures are obtained (such as lightning striking respondents while they are completing questionnaires), or to systematic biases (such as shocking respondents with an electrical cattle prod in order to keep questionnaires under 25 minutes).

Each of these sources of variance will have some impact on basic characteristics of measures:

- Elevation - the level of rating, often represented by the mean
- Scatter - fluctuations in rating, e.g., standard deviation
- Shape - density of ratings at a value, often estimated with measures of skewness and kurtosis (Dillon, Mulani and Fredrick, 1984).

After the data are collected (or during the collection process), three basic problems in data preparation confront the researcher cum cartographer:

- How to handle missing data.
- How to make the individual variance distributions of each component of the cube match the requirements of the mapping algorithm and the research issue (Distributional Transformations).
- How to make all the variances in the data matrix match requirements and assumptions (Matrix transformations).

Very briefly, I want to deal with all three, but dwell just a little longer on one of the possible matrix transformations that we have found useful in our mapping work.

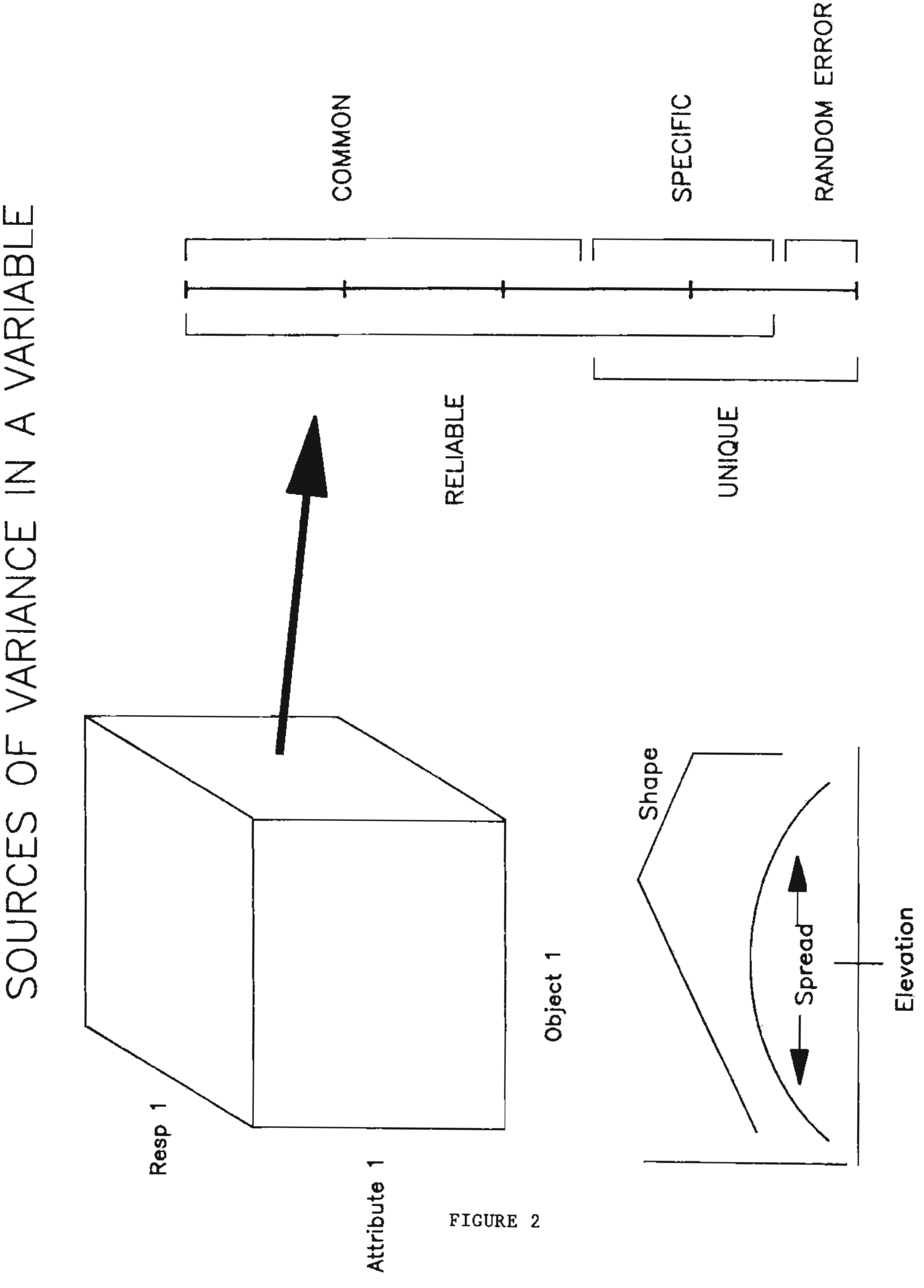

SOURCES OF VARIANCE IN A VARIABLE
Resp 1
Attribute 1
Object 1
COMMON
SPECIFIC
RANDOM ERROR
RELIABLE
UNIQUE
Shape
Spread
Elevation

FIGURE 2

MISSING VALUES

Again, following Rummel (1970), four basic alternatives are available to researchers who encounter missing values in their data:

- Data matrix left as is; transformation or analysis compensates
- Estimate some missing values for which high confidence exists
- Estimate all missing values
- Variables or cases may be omitted.

Each of these options has relatively large or small consequences for resulting maps depending on the size of the problem and the nature of the mapping algorithm.

- For mapping algorithms that employ eigen vector analysis (e.g., factor analysis), leaving missing values in the data and calculating correlations on a pairwise case deletion basis can theoretically result in overestimation of the contribution of individual variables (as a result of negative eigen values).
- Replacing missing values for some variables allows the variables that are evaluated to contribute to the generality of the results, but can have the effect of reducing the actual variance estimates for the variable. Consequent covariance estimates or correlations may be affected (possibly by attenuation), and the resulting relationships among dimensions may be off.
- Eliminating variables or cases may increase the reliability of what remains, but may have some impact on the generality of results.
- Moving to a lower level of measurement and using a different mapping algorithm (e.g., correspondence analysis) may sacrifice precision.

DISTRIBUTIONAL TRANSFORMATIONS

In Rummel's (1970) words, the point a distributional transformations is:

- "...to match the range of potential values of the concepts perationalized in the data, to reduce the effect of extreme data values on the analysis, and to linearize relationships between the variables (170)."

The transformations are undertaken in two phases, and they should be done in a manner that will meet specific criteria.

In the first phase, univariate and bivariate distributions are inspected for distortions that may have some impact on subsequent dimensional analysis. Three of the nastier things that can happen to a distribution or two are:

- <u>Underlying Distribution</u>. Misspecification of the phenomena to be measured can lead to distortions of results. For example:

  Desired Construct to Be Measured: MORALITY

  Operational Measurement:

  Perceived Number of Affairs With a Person
  Named After a Commodity Staple

	#
Ted Kennedy	4
George Bush	2
Robert Dole	2
Herbie Porter	0

  Is "4" twice as many as "2"? Do I have the heart to tell you that once is enough? Perhaps a more appropriate approach would be to use a lower (nominal) level of measurement.

- <u>Extreme Frequencies</u>. When frequencies of events do not accurately reflect the phenomena of interest, some question arises as to their worth in detecting differences among rated objects, for example:

  Would Let Him Marry Your Sister

	Yes	No	%
Robert Dole	0	50	0
George Bush	0	50	0
Herbie Porter	1	48	2

- <u>Restriction in Range</u>. This may occur when distributions are skewed in the same or opposite directions. The net effect is to reduce the maximum range of correlation from between +/- 1.0 to something smaller. For example:

How Likely to Vote for Candidate

	George	Herbie
Definitely Would Not	8	4
Probably Would Not	12	10
Might/Might Not	15	20
Probably Would	10	8
Definitely Would	5	8

Maximum P-M correlation = .93 (After Carroll, 1961).

There are five basic guidelines to follow when selecting a distributional transformation (demonstrated in Figure 3):

1) Monotonicity - the relative rank positions of objects that are evaluated should stay pretty much the same, and if possible, information about the distances between objects should be preserved.

2) Error - the resulting distances between respondents should not be greater or smaller disproportionately relative to the size of the error band in measurement.

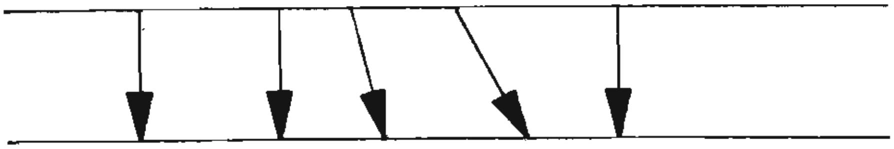
3) Empirical Distribution - what does the raw distribution look like? Which transformations would be most appropriate to achieving other criteria?

4) Normality - transforming to a normal distribution will take care of problems of restriction in range while assuring that resulting bivariate relationships will be somewhat linear, additive and homoscedastic.

5) Outliers - an extreme case of extreme values. A few observations including outliers can seriously alter relationships, e.g.,

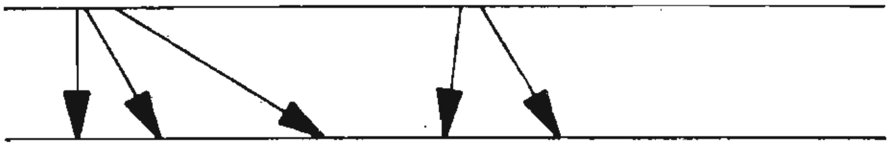
Criteria for Distributional Transformations

Montonicity: Rank positioning of respondents remains the same before and after transformation



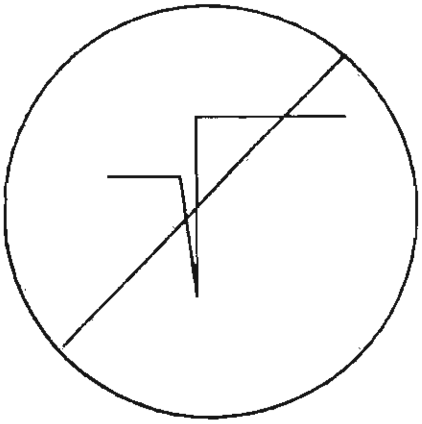
Error:

Transformations are undesirable to the extent that differences within sampling error become meaningful.



Empirical Distribution:

One transformation may not be appropriate for all variables.



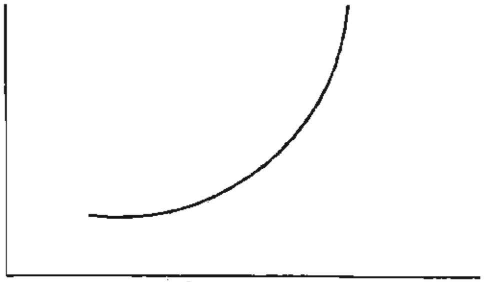


FIGURE 3

Agreement That Candidate Is Wholesome

Respondent	George	Herbie
1	7	2
2	6	3
3	5	4
4	4	5
5	3	6
6	9	9

where, without Respondent #6 the correlation between George and Herbie would be r = -1.0, but with #6 included, the correlation jumps to r = +.25. Iteratively standardizing distributions with outliers may substantially reduce the problem.

A number of transformations may be applied once the univariate and bivariate empirical distributions have been identified. Examples of such distributions are shown in Figure 4.

Criteria for Distributional Transformations

Normality

Univariate and bivariate:

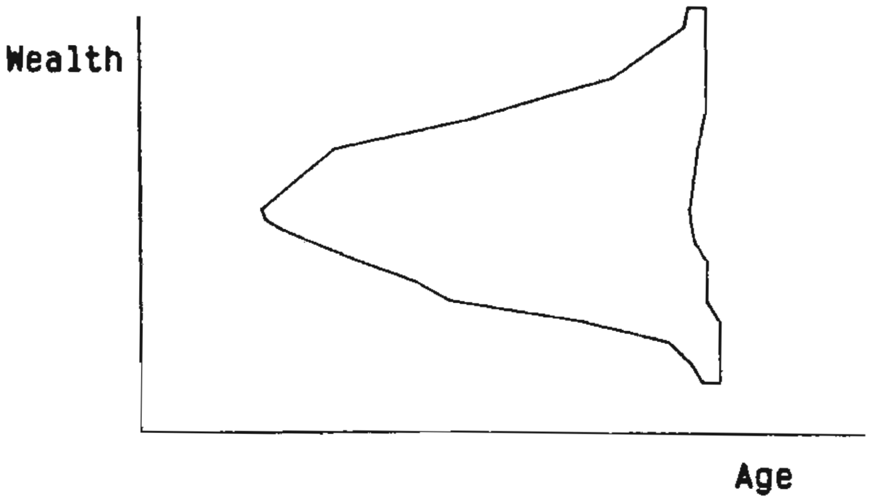


FIGURE 4

MATRIX TRANSFORMATIONS

Many mapping algorithms (e.g., Factor Analysis, MDPref, etc.) employ matrices to determine the locations of attribute dimensions and object points relative to an eigenvector. Thinking back to the components of variance, matrix transformations have the potential for

- Removing problems of non-comparable variances (heteroscedasticity), and
- Isolating particular types of variance that may be of more use than others for representing a perceptual space.

Four of the more popular types of transformations that involving matrix scaling are:

- Standardization - generally undertaken when dealing with non-comparable variances to bring each variable into a unit-normal distribution so that mean and deviation differences do not impinge on covariance calculations. For those mapping algorithms that employ correlations as inputs, standardization within columns of the data cube occur in the calculations of the correlations.

- Double Standardization - standardizing within a row of the data cube as well as within columns. This reduces the influence of the mean and standard deviation of responses across attributes so that the only variation remaining lies between objects that are evaluated within a respondent. The variability in evaluations can be studied and mapped, but the contribution of each variable to positioning is fixed in unit-normal space.

- Centering - subtracting the average evaluation for a respondent (calculated across objects and attributes) from each rating of an object on an attribute. This results in data that retain the variability of objects and attributes within a respondent while removing the influence of the individual respondent.

- Double Centering or Centering-by-Row-Standardizing-by-Column - (partial ipsatization) which will more or less retain the variability of attributes and entities while reducing effects of non-comparable variances between attributes.

The last three matrix transformations are used to deal with problems of "halo effects" (Thorndike, 1920) or "response bias," the tendencies for groups (a source of systematic error) or single individuals (a source of random error) to rate objects homogeneously across attributes. This homogeneity in rating generally may take the form of elevation differences (error) in evaluations of objects and attributes (Cronbach and Gleser, 1953; Dillon, Mulani and Fredrick, 1984). Each of the transformations that reduces the row evaluations of the data cube to a deviation (unit or raw) from a respondent's mean eliminates elevation effects and thereby improves the accuracy of true object positionings by reducing error. Selection of a particular one will depend on the purpose of the mapping.

SOME USEFUL TRANSFORMATIONS TO NORMALITY

DISTRIBUTION CHARACTERISTICS	TRANSFORMATION
1. Right skewed, platykurtic	$\hat{X}_j = \sqrt{X_j}$
2. Right skewed, leptokurtic	$\hat{X}_j = \log X_j$
3a. Extreme left skew Platykurtic, no skew	$\hat{X}_j = 2 * \text{Arcsin} \sqrt{X}$
b. Extreme left skew	$\hat{X}_j = \log \frac{X_j}{1 - X_j}$
4. Nearly J Platykurtic, no skew	$\hat{X}_j = 1/2 * \log \frac{1 + X_j}{1 - X_j}$

FIGURE 5

My bias is to use a matrix transformation that will yield values that may be interpreted relatively across attributes and objects. By this I mean, clients generally want to know about the relative contribution of attributes to determining consumer behavior ("What's the most important thing about my brand?"). Accordingly, since ipsatizing (Cattell, 1944) the data through double standardization modifies all attribute variances to unit length, my preferences lie towards partial ipsatization. Since restriction in range may attenuate correlations, row centering, and then factoring correlations computed from the centered data seems preferable (since standardization may reduce the effects of restrictions in range).

Let's take a look at the consequences of row centering and then factoring standardized column data versus simply factoring the standardized columns.

As indicated earlier, no discussion of mapping is complete without an example involving automobiles or politicians. Unfortunately, the first data set I found that had data readily available in both raw and centered form involved food additives. In the interest of protecting my client's brand, I thought it best to change a few names, and since it is de rigueur to include a political example, the major candidates had to serve as replacements. Since they obviously do not fit appropriately into this space, a substantive interpretation of the relationships among attributes and objects is not warranted.

First let's examine the relationship between the means before and after centering the data. Figure 6 presents the means on each attribute for each of the candidates. Centered values are plotted on the top ordinate, while raw means are plotted on the bottom. This is one excellent check to see that the centering process has been performed properly, i.e., no differences will be found in mean profile (or shape) across respondents for attributes and objects before and after centering.

ATTRIBUTE RATINGS

Raw and Centered Political Candidates

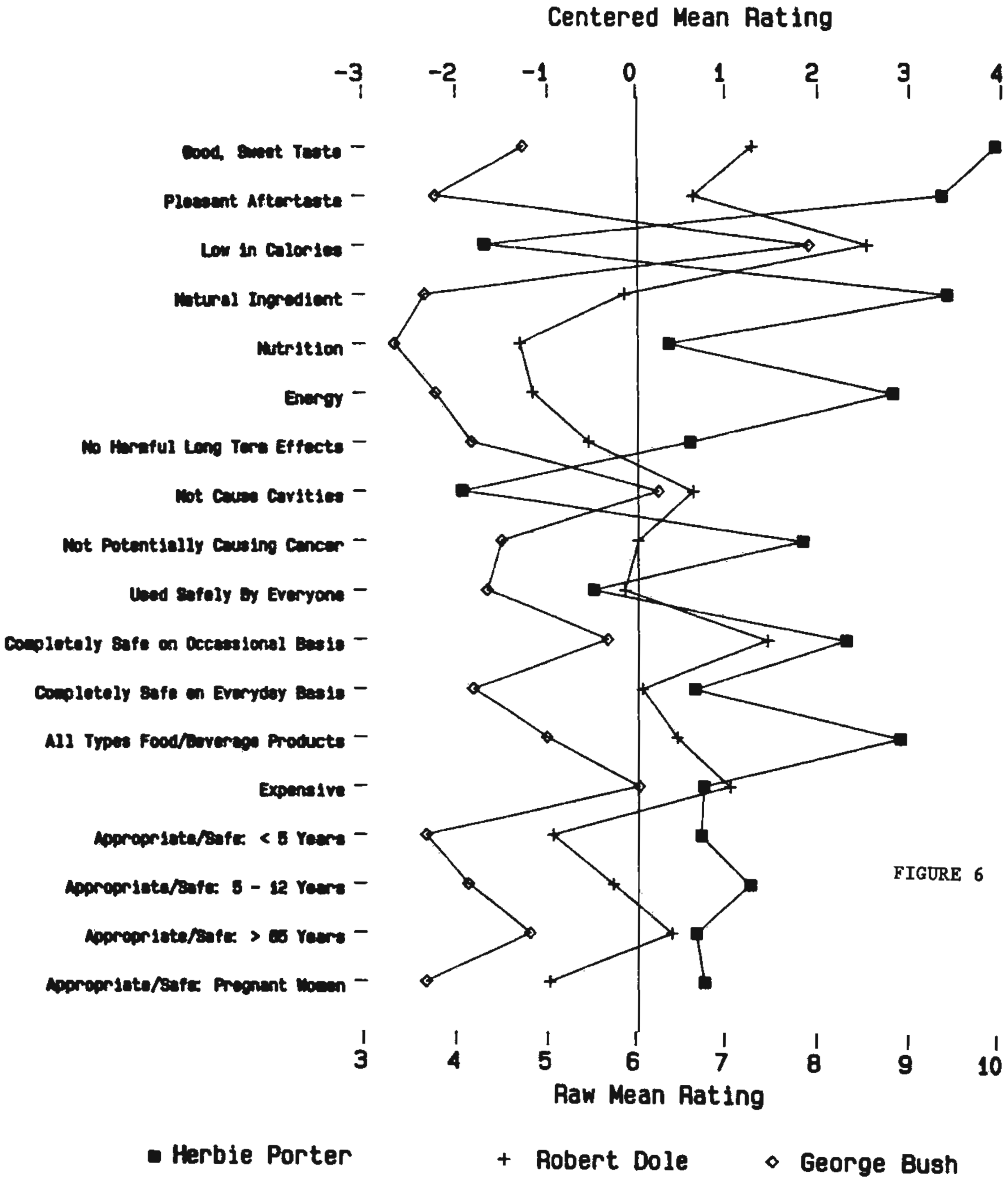


FIGURE 6

Next, let's look at the relationships among the attributes. Tables 1 and 2 present the data from factoring the correlations among the attributes when the raw and centered data are used. The same factor model specification was used in both cases (principal components factor analysis, mineignen = 1.0, varimax rotation). In the raw data solution, three factors are extracted from the data, which we call "Safety," "Calories/Cavities," and "Taste." The centered solution is more articulated as a result of removing response bias; five factors are extracted rather than three, and their meanings are more specific.

Generally, in factor analysis, the first factor uncovered accounts for the greatest amount of variance in the original attributes. In the analysis of the raw data, the first factor is Safety (accounting for about 26% of the variance), while in the centered data, the first factor is Nutrition/Taste (accounting for about 19% of the variance).

A comparison of the relations among attributes and candidates is offered in Figures 7 - 12. Fortunately, the perceptual maps for both factor solutions are in agreement with respect to the ordinal positioning of each of the candidates on each of the factors and attributes. From either solution, the prescription for each candidate with respect to their ideal positioning is clear.

What may not be clear, however, from the raw data factor solution is just how far each of the laggards should go in improving their image as being "Appropriate/Safe" for particular segments relative to some of the other "Safety" attributes. A comparison of regressions of the factor scores against respondents' evaluations of their likelihood to vote for each candidate (bold vector on each map) does not help when the raw data are examined. When the centered factor scores are regressed, separate answers will be given for what are clearly two separate perceptual dimensions of evaluation and a candidate could apportion time, energy, and dollars against each issue in respective proportion to their worth.

Pursuing this idea of regressing factor scores against a criterion measure to infer the relative importance of the factors, the raw data factor solution indicates that "Safety" is more important than "Taste." The centered factor solution indicates that "Taste/Nutrition" is more important that either "Safety" dimension separately or combined. The choice of which of these factor analyses to use also will depend on the extent to which one believes that it is socially desirable to elevate one's evaluations of safety concerns relative to more hedonistic ones. If the position is taken that people will behave in the marketplace or the voting booth in a socially desirable manner, the raw data factor analysis is preferable. Alternatively, if one takes a more hedonistic approach, and believes (perhaps cynically) that we behave more with our stomachs than our minds, then the centered data factor analysis is preferable.

TABLE 1

Factor Analysis of Raw Data

	Safety	Calories/ Cavities	Taste	
Appropriate/Safe: Children 5 - 12 Years	0.85	-0.02	0.30	
Appropriate/Safe: Children < 5 Years	0.83	-0.08	0.27	
Appropriate/Safe: Pregnant Women	0.82	-0.04	0.29	
Appropriate/Safe: People > 65 Years	0.81	0.17	0.16	
Completely Safe on Everyday Basis	0.59	0.32	0.41	
Used Safely By Everyone	0.52	0.41	0.29	
No Harmful Long Term Effects	0.50	0.31	0.44	
Low in Calories	0.07	0.75	-0.27	
Not Cause Cavities	0.12	0.75	-0.11	
Pleasant Aftertaste	0.23	-0.06	0.80	
Good, Sweet Taste	0.24	-0.09	0.76	
Natural Sweetener	0.30	-0.11	0.74	
Good Energy Source	0.32	-0.16	0.69	
All Types Foods/Beverages	0.33	0.14	0.59	
Not Cause Cancer	0.42	0.09	0.52	
Good Nutritional Value	0.45	0.02	0.50	
Completely Safe on Occasional Basis	0.45	0.32	0.46	
Expensive	-0.14	0.44	0.38	
% Variance in Correlation Accounted For:	25.7	10.6	23.7	
% Explained Variance of Voting Likelihood	52.7	9.8	37.5	
% Voting Likelihood Explained				31.2

TABLE 2

Factor Analysis of Centered Data

Price

	Taste/ Nutrition	Segment Safe	Universal Safety	Calories/ Cavities	
Good Nutritional Value	0.79	0.25	-0.01	0.02	-0.10
Good Energy Source	0.74	0.17	0.09	-0.34	0.07
Natural Sweetener	0.70	0.12	0.29	-0.27	0.14
No Harmful Long Term Effects	0.57	0.27	0.27	0.20	-0.07
Pleasant Aftertaste	0.56	0.11	0.39	-0.33	0.31
Good, Sweet Taste	0.53	0.15	0.37	-0.33	0.33
Not Cause Cancer	0.48	0.19	0.29	-0.09	0.09
Appropriate/Safe:					
Children 5 - 12 Years	0.26	0.83	0.21	-0.12	0.07
Children < 5 Years	0.28	0.83	0.10	-0.17	0.04
Pregnant Women	0.35	0.75	0.19	-0.11	-0.07
People > 65 Years	0.00	0.74	0.36	0.06	-0.02
Occasionally Safe	0.09	0.19	0.81	-0.02	0.04
Safe Every Day	0.23	0.28	0.74	0.08	-0.10
All Types Foods/Beverages	0.25	0.19	0.60	-0.26	0.23
Used Safely By Everyone	0.23	0.17	0.54	0.30	-0.15
Not Cause Cavities	-0.06	-0.08	0.03	0.78	0.01
Low in Calories	-0.28	-0.12	0.01	0.71	0.16
Expensive	0.07	-0.02	-0.04	0.14	0.88
% Variance in Correlation Accounted For:	18.6	16.4	14.5	10.1	6.4
% Variance of Voting Likelihood	44.6	9.4	22.8	11.1	12.1
% Voting Likelihood Explained					40.2

SAFETY VS CALORIES/CAVITIES

Raw Political Candidates

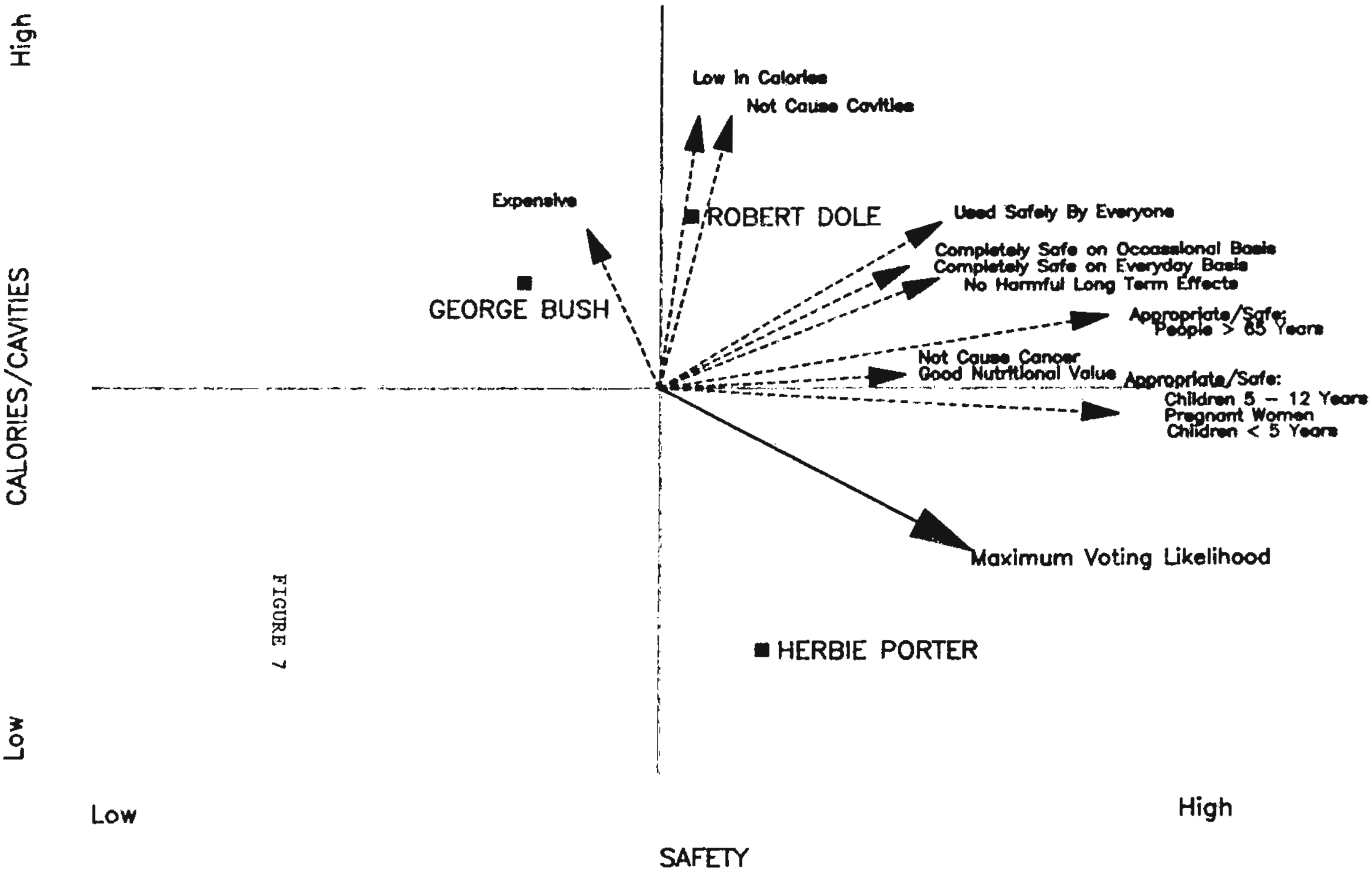


FIGURE 7

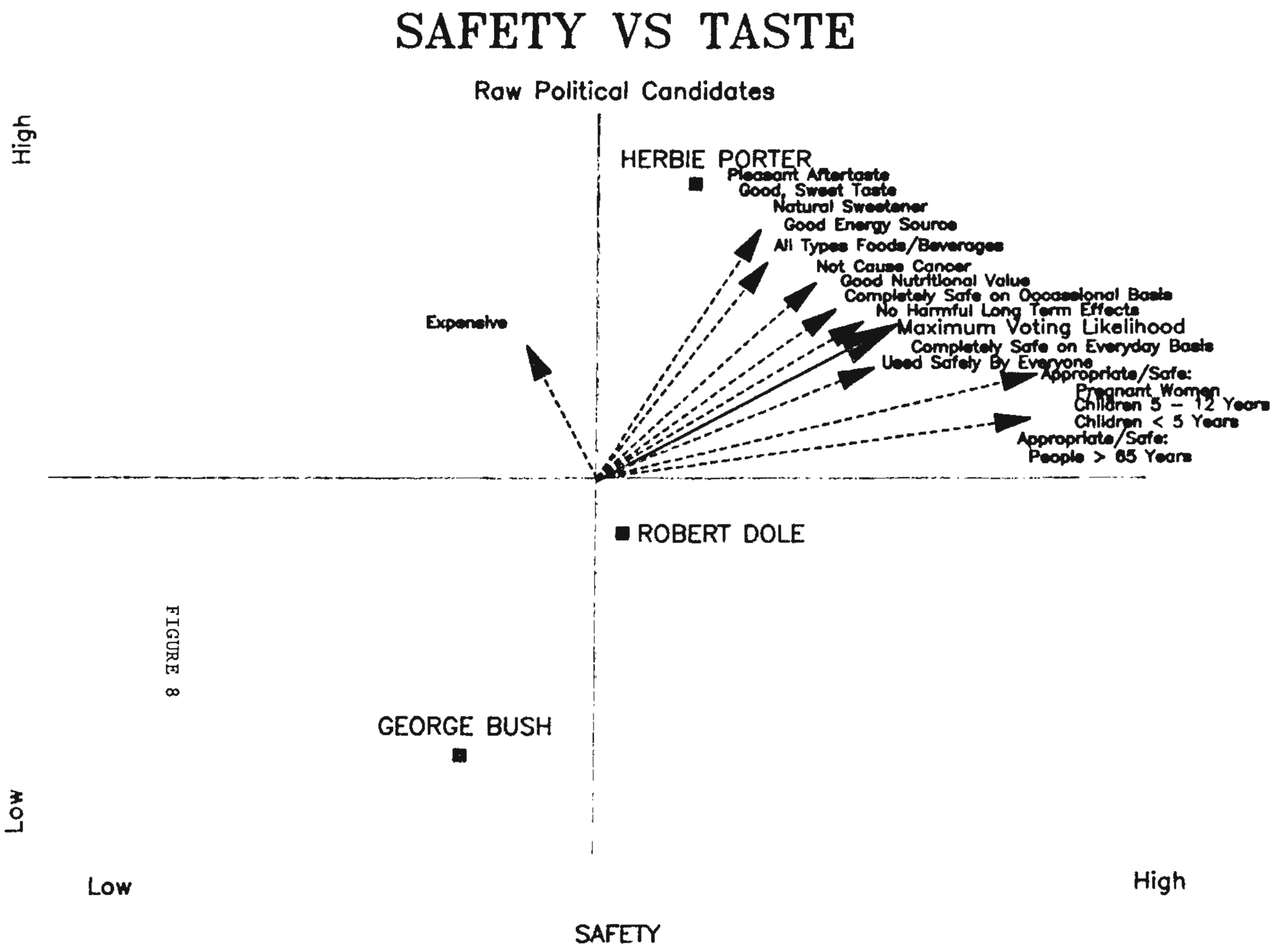

SAFETY VS TASTE
Raw Political Candidates
High
TASTE
Low
Low
SAFETY
High
HERBIE PORTER
ROBERT DOLE
GEORGE BUSH
Expensive
Pleasant Aftertaste
Good, Sweet Taste
Natural Sweetener
Good Energy Source
All Types Foods/Beverages
Not Cause Cancer
Good Nutritional Value
Completely Safe on Occasional Basis
No Harmful Long Term Effects
Maximum Voting Likelihood
Completely Safe on Everyday Basis
Used Safely By Everyone
Appropriate/Safe:
Pregnant Women
Children 5 – 12 Years
Children < 5 Years
Appropriate/Safe:
People > 65 Years

FIGURE 8

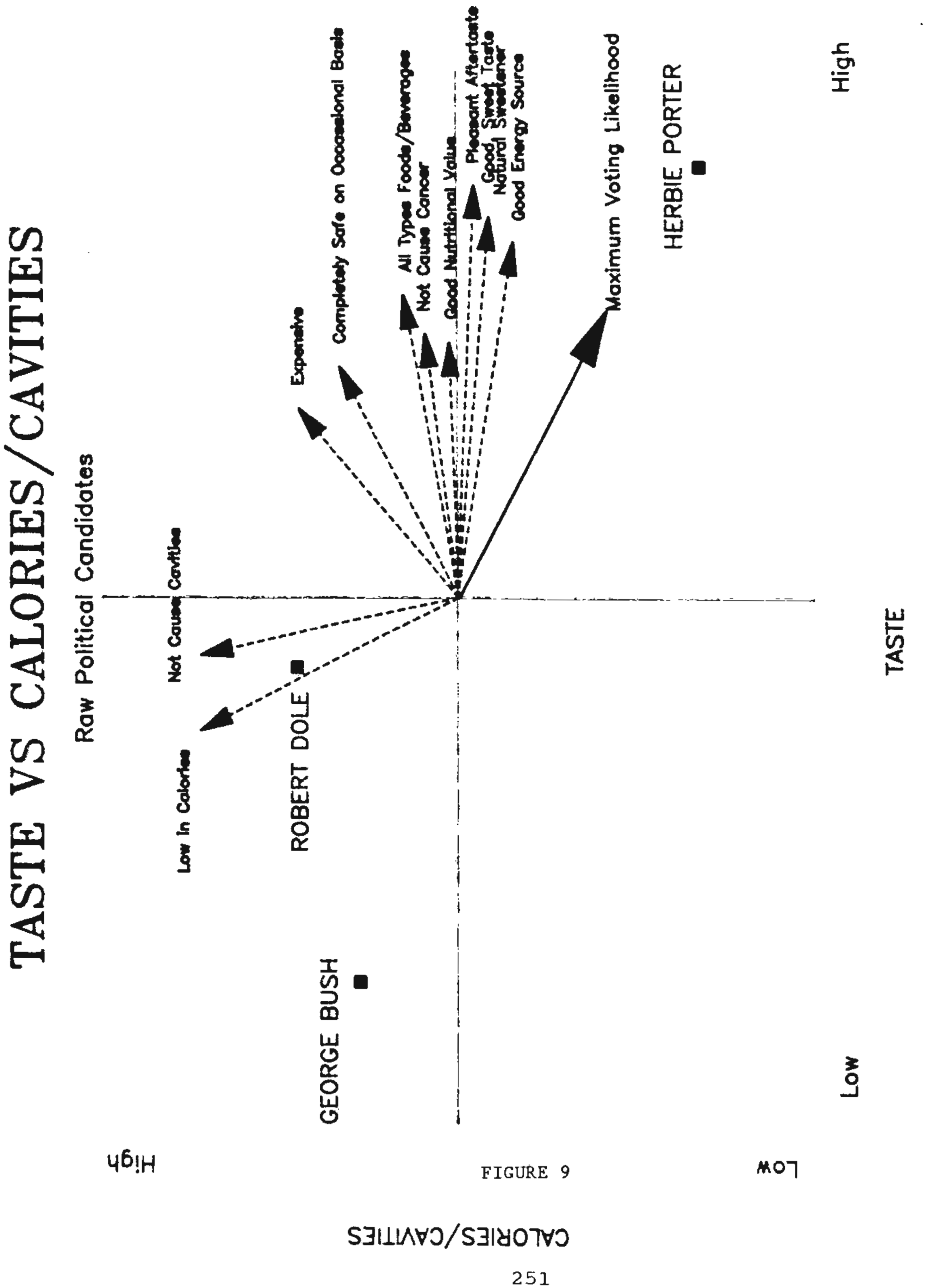

TASTE VS CALORIES/CAVITIES
Raw Political Candidates
Expensive
Completely Safe on Occasional Basis
All Types Foods/Beverages
Not Cause Cancer
Good Nutritional Value
Pleasant Aftertaste
Good Sweet Taste
Natural Sweetener
Good Energy Source
Maximum Voting Likelihood
HERBIE PORTER
Not Cause Cavities
Low In Calories
ROBERT DOLE
GEORGE BUSH
TASTE
High
Low
CALORIES/CAVITIES
High
Low

FIGURE 9

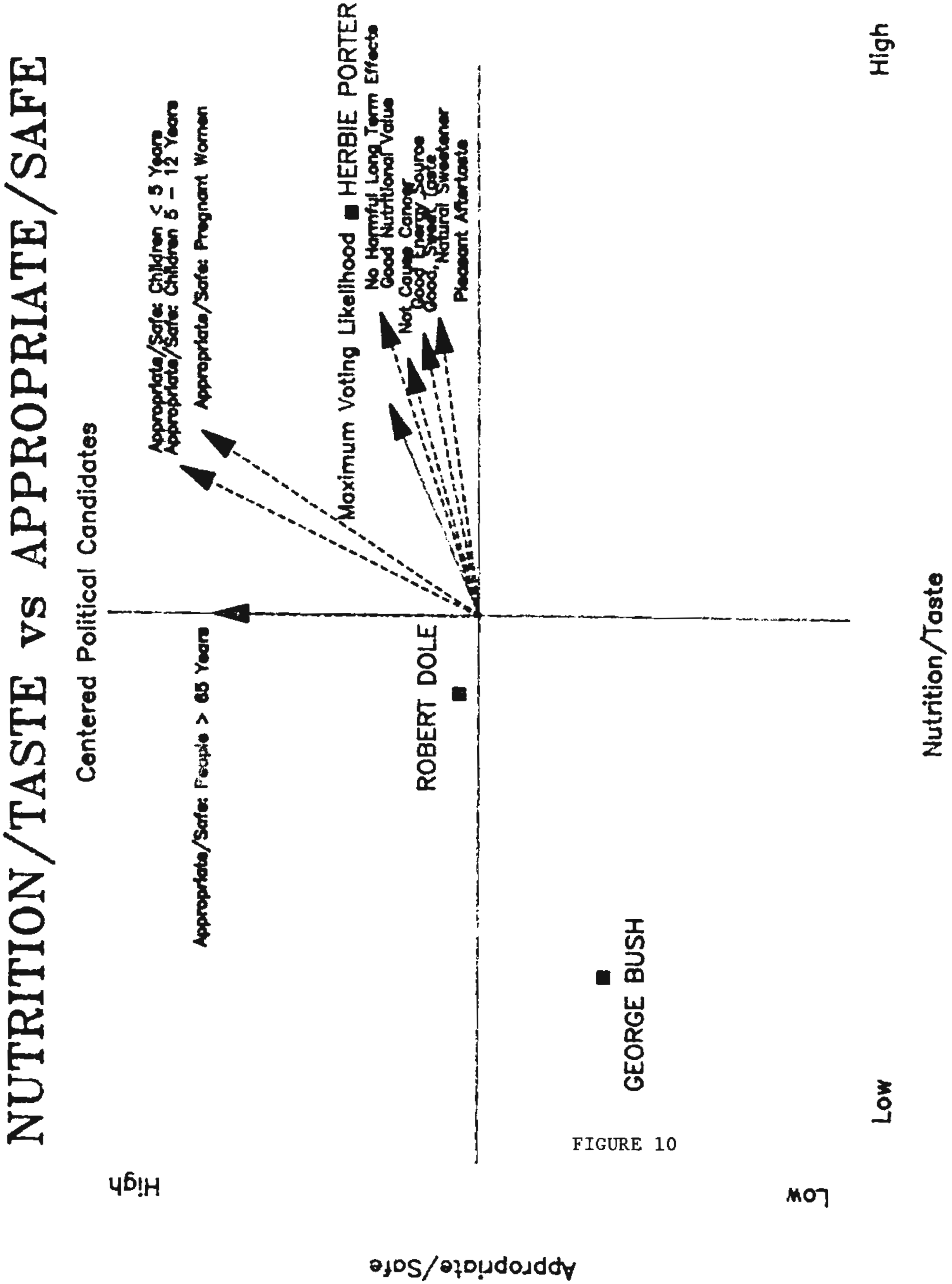

NUTRITION/TASTE vs APPROPRIATE/SAFE
Centered Political Candidates
Appropriate/Safe: Children < 5 Years
Appropriate/Safe: Children 5 – 12 Years
Appropriate/Safe: Pregnant Women
Appropriate/Safe: People > 65 Years
Maximum Voting Likelihood
HERBIE PORTER
No Harmful Long Term Effects
Good Nutritional Value
Not Cause Cancer
Good Energy Source
Good, Sweet Taste
Natural Sweetener
Pleasant Aftertaste
ROBERT DOLE
GEORGE BUSH
Nutrition/Taste
Low
High
Appropriate/Safe
High
Low

FIGURE 10

NUTRITION/TASTE vs UNIVERSAL SAFETY

Centered Political Candidates

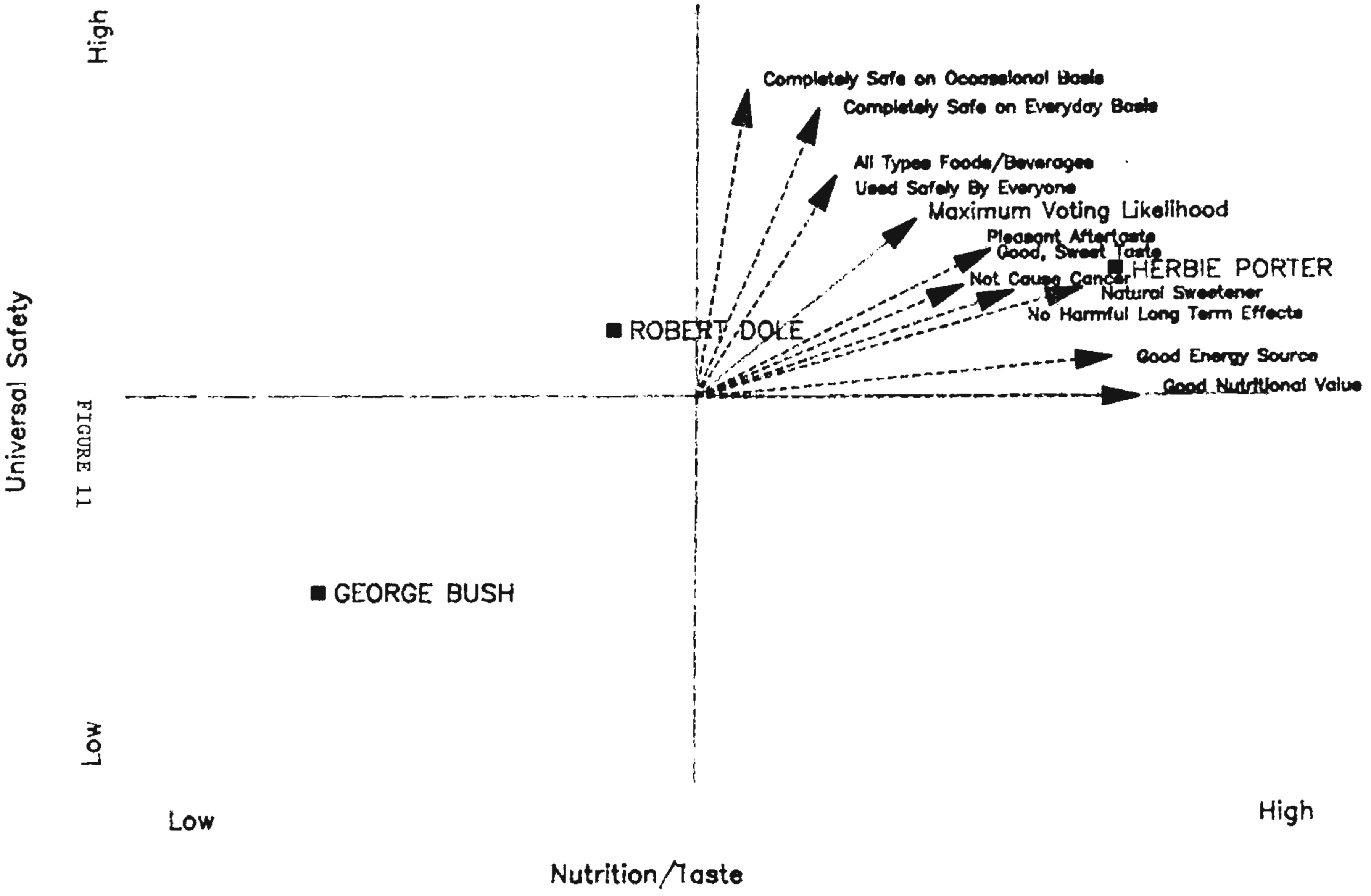


FIGURE 11

UNIVERSAL SAFETY vs APPROPRIATE/SAFE

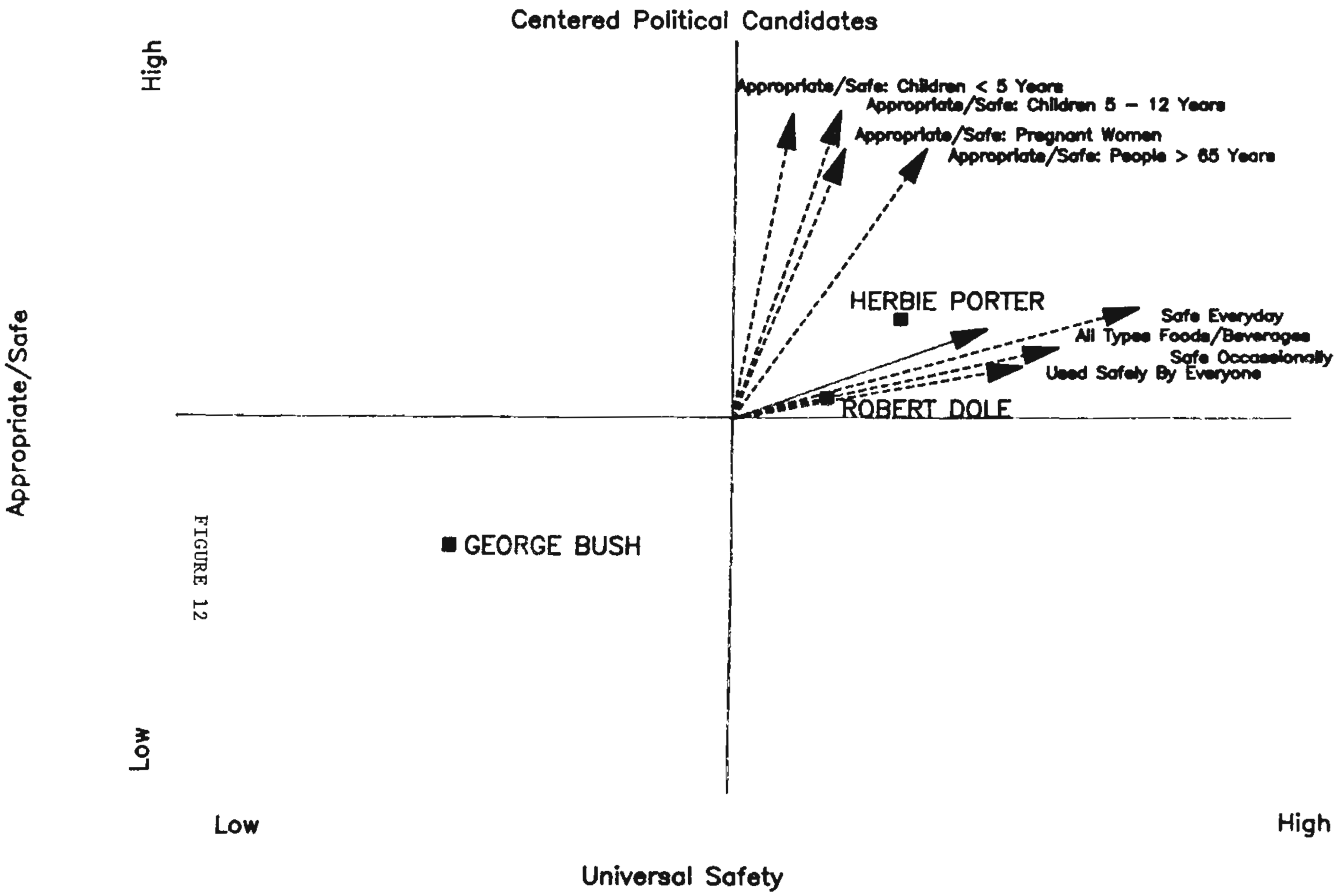


FIGURE 12

THE TROUBLE WITH TRANSFORMATIONS

Here is a simple formula for centering raw data responses when incomplete evaluations are obtained across objects, e.g., when some objects aren't evaluated because the respondent has no familiarity with them. The algorithm is slightly more complex because centering data for which a different number of evaluations are obtained per respondent can remove object and attribute variance as well as respondent halo:

$$\hat{X}_{ijk} = (X_{ijk} - \bar{X}_{i..}) + (\bar{X}_{.j.} - \bar{X}_{...}) + (\bar{X}_{..k} - \bar{X}_{...})$$

where:

i is the ith respondent
j is the jth attribute
k is the kth object
X is an evaluation
$\bar{X}$ is a mean
. indicates summation across a subscript.]

Coupled with a few distributional transformations and a missing data algorithm or two, the process of preparing data for mapping algorithms can be exhaustive, time consuming, error prone itself, and consequently costly. Training and experience are required, and sometimes I wonder in today's market if the consumers of research would even know the difference if data weren't prepared properly. Perhaps they do find out eventually, but the consequences may be to throw up their hands in frustration with research in general. If one takes a position that a quality product is to be offered in research, then training and experience should be offered to, and acquired by, most researchers.

Short of that, if a researcher can appreciate the role of the components of variance and error in mapping and wishes to minimize their impact on dimensional representations, the best advice to be given would be to make sure that:

- Data are complete in the cube; make it difficult for a respondent to avoid providing values or evaluations of an object.
- Data distributions be kept at the lowest level of measurement that can be effectively handled -- correspondence techniques are fairly robust and require only nominal levels of measurement.
- Standard algorithms be employed (such as APM) where the intervention of the researcher is kept at a minimum.
- Demand more and better algorithms for mapping from those who provide standard methods.

In summary, a number of sources of variation are present in the data cube. Some cautions are necessary when dealing with the data, particularly as they relate to missing values and distributional problems. Beyond that, more sophisticated approaches may be taken to the variance in the cube to tease out those sources that are less error laden. Serious thought should be given by the researcher as to the approach that will be given to data preparation.

REFERENCES

Carroll, J.B. 1961. "The Nature of Data, or How to Choose a Correlation Coefficient." Psychometrika 26: 347-72.

Cattell, R.B. 1944. "Psychological Measurement: Normative, Ipsative, Interactive." Psychological Review 51: 292-303.

Cronbach L.J. & Gleser G.C. 1953. "Assessing Similarity Between Profiles." Psychological Bulletin 90: 218-244.

Dillon, W.R., Mulani N. & Fredrick D.G. 1984. "Removing Perceptual Distortions in Product Space Analysis." Journal of Marketing Research 21: 184-193.

Rummel, R.J. 1970. Applied Factor Analysis. Northwestern University Press.

Thorndike, E.L. 1920. "A Consistent Error in Psychological Ratings." Journal of Applied Psychology 9: 188-195.

DECISION CRITERIA FOR SELECTING MAPPING TECHNIQUES

Gordon A. Wyner
Senior Vice President
M/A/R/C, Inc.

The purpose of this paper is to consider, from the perspective of the practitioner, some of the key research design issues that arise in selecting a perceptual mapping technique. Successfully dealing with each of these issues will contribute to a better understanding of the uses and limitations of mapping. The intent is not, therefore, to produce a set of prescriptions for exactly "how to do it." To the contrary, one of the few generalizations made in this paper is that no one approach is likely to work for all situations, nor does any one situation always call for one particular approach. The goal here is to draw attention to design issues that cut across many other types of research, and to apply them to mapping.

The frame of reference for this paper is consumer marketing research as conducted by the major packaged goods, durables, industrial products, and services companies in the U.S. The "Sample" of experience that is drawn on is therefore fairly representative of mapping experience today.

Specifically, the following issues will be addressed:

- Objectives
- Sampling
- Data Collection
- Analysis
- Presentation

After outlining how these concerns each impact the mapping technique, some implications of each for mapping practice will be discussed.

1. CLARIFY OBJECTIVES

The reasons for using mapping techniques vary considerably across applications. Sometimes the reasons for mapping are critical and essential to the research being done, and other times mapping is conducted more as an afterthought. Sometimes there are fairly specific research hypotheses to be tested through mapping, while other times the reasons for mapping are primarily methodological. In order to better understand what can and cannot be accomplished with mapping, it is useful to establish what the objectives really are.

Identify Market Structure

One of the classic textbook reasons for using maps is to understand the structure of the market. This typically refers to the way in which consumers perceive brands to relate to one another. Are they similar or substitutable? Do consumers organize their thinking about the available brands the same way the client's marketing department thinks? For example, does the consumer "sort" the brands by form, flavor, packaging, price or some other variable? By understanding the structure in this sense, the client can be assured that marketing actions are taken against the right set of product attributes and competitors.

Evaluate Product Position

Either an existing or new product is backed by a particular positioning strategy. The goal of mapping is to assess the performance of that strategy by measuring the extent to which the product is perceived to have each of the relevant attributes, against the context of the other products in the market. One of the necessary steps in this analysis is the determination of which attributes are most important, so that the less important variables can be suppressed. The fewer attributes presented in the map, the simpler the map will be to understand. Visual representation of this kind of analysis through maps is preferable to a series of (perhaps many) attribute profiles.

A variation on this type of map is the importance/performance grid. A common way to display data is to select one brand per map and array its average scores on the attributes on one dimension and the average importance weights of those same attributes on the other dimension. One of the greatest benefits of this type of map is its simplicity. It is easily understood by non-researchers, and leads quite directly to marketing action. For example, low performance on high importance attributes suggests a problem area and a potential solution. Although the same kinds of conclusions can be derived from the analytically more sophisticated maps (including "gap" analysis, the search for marketing opportunities, etc.), it's valuable to keep in mind the humble roots of the map in marketing. The ultimate goal, after all, is communication of results, and two-dimensional arrays can facilitate understanding of many types of data.

Reduce the Data to an Interpretable Form

Often the analyst is presented with a massive data matrix of brands by attributes by respondents and asked for help in pulling out what is meaningful. While there may be other research hypotheses, the goal here is essentially methodological. Find some way to reduce stacks of printout to the key relationships, i.e., the subset of variables that is most important and the interrelationships among brands. In this sense, mapping is a descriptive tool and a display device. Looking at maps saves considerable labor when compared to sifting through many tabulations.

Track Changes in Positioning Over Time

There is a belief that brand positions should change over time, in response to marketing actions and expenditures. Typically, a tracking study is already under way, the selection of brands and attributes having been determined long ago. Again, the goal is in large part methodological. Find a way to demonstrate and display whatever change is occurring. A perceptual map that plots each brand's position at each point in time is developed to monitor changes.

2. CONSIDER THE SAMPLING IMPLICATIONS

In those situations where mapping is central to the study, there is an opportunity to design the sample in ways that materially improve the results. If the sampling design is a given, and mapping a lesser consideration in the study, then a review of the sampling properties of the study can reveal whether the design is at least acceptable for mapping purposes.

In general, sampling relates to the selection of people, brands, and the occasions or situations for which the brands are to be rated. In an ideal world, extremely large samples of people and brands would take care of most of the potential problems. The cost and burden on respondents rating many brands introduce constraints on what gets rated and potential biases in the results. For example, is the sample of brands representative of some "universe" of available brands? How are the variations in availability by region to be addressed? If the client's brand is a less familiar brand in the category, is it to be over-sampled (and with what consequences)?

Is the sample of people representative, especially with respect to the brands being rated? Since priority schemes are often used to assign people to brands to rate, are the people representative of those aware of the brand? If low incidence quotas are always filled first, then the higher incidence brands may never be rated by people who are aware of low incidence brands.

In many product categories there are reasons to "sample" different situations for administering a mapping task. In fast food, for example, attribute ratings are often fundamentally different in relation to lunch, snack, or dinner occasions. Much of the mapping research in restaurant markets is occasion-specific. The general question is whether a consumer's evaluation of a brand or product is constant across situations. To the extent that the ratings have to do with company image and other company-specific issues, the situation can probably be ignored. In other cases, where product evaluations are usage-specific, it may be inappropriate to ask the respondent to mentally average his impressions across occasions.

A number of other methodological issues in mapping were identified in a recent paper by Srinivasan et. al (1987). One of the most relevant to sampling is the potential for ecological or environmental correlation between brands when no true casual relationship exists. For example, if all or most high performance cars are of European origin then "performance" and "European" may be unduly forced together in perceptual maps.

More generally, these authors point out the importance of the assumption of "homogeneity of structure" across brands. If the map structure is not the same for all brands then the analysis can be misleading.

3. DATA COLLECTION: CONSIDER THE RESPONDENT BURDEN

Rarely does the respondent voice a concern about the rating task he is being asked to perform (although it has been known to happen). Unfortunately the consequences of overloading the respondent may show up only at the analysis stage. For this reason it is worth considering these potential problems at the design stage.

Repetitiveness

Often, respondents are asked to pass through the same, long (20 plus items) list of attributes for many brands (e.g. more than two). Boredom may set in and the answers may not be s "thoughtful" as they ought to be. Errors in response may be common. More specifically, there may be carry-over effects, in which the act of rating the previous brand may influence ratings on the next one. There may be a context effect, in which the ratings may be influenced by the particular subset of brands rated by a respondent. The same considerations that cause concern in product testing (e.g., in sequential monadic designs) ought to be addressed here. Perhaps some of the same solutions can be used more in mapping applications. For example, balanced incomplete block designs for allocation of brands to people could be used. Analysis of variance can be used to explore carry-over effects.

Difficult Rating Tasks

In our experience, respondents generally have more trouble with similarity scales and open-ended sort tasks than with brand attribute ratings. This conclusion is somewhat clouded by the fact that many of the tasks incorporate extremely large numbers of ratings as well: for example, over 100 paired magnitude estimation ratings, over the telephone. Perhaps respondents can handle paired similarity judgments in relatively small numbers. The analogy with product testing may be relevant again, consumers don't seem to have trouble doing overall preference ratings and a small number of paired preference attribute ratings (e.g. less than 10).

4. ANALYSIS: CONSIDER MORE THAN ONE

It's unlikely that any one technique can ever be demonstrated to be universally more effective, if for no other reasons than the difficulty in defining rigorously what is effective. Interpretability is, almost be definition, extremely important. A multiple methods approach is desirable, in which the convergence of results from several analytical methods enhances credibility. Obviously it is more feasible to combine some methods together than others. Factor and discriminant analysis can work off the same raw data. Multidimensional scaling and other dissimilarity matrix-based methods are more "compatible." Ideally, in the future it would be useful to know the conditions under which the alternative methods yield comparable results so that multiple methods wouldn't be needed. One could safely choose the analysis technique based upon other important criteria.

5. PRESENTATION: A DESIGN ISSUE

There are limits on the number of dimensions, attributes, and brands that the audience can comprehend and assimilate. In principle this is a researchable issue. At what point does the number of dimensions and objects displayed degrade rather than improve the communication value of the results? In essence the map is a presentation device so consideration ought to be given to how complex the final output will be during the design phase of the study. If the map is likely to be too complicated, then other techniques ought to be contemplated. For example, for problems with 5 or more dimensions, hierarchical clustering is an alternative. It is relatively easy to understand five branches of a tree diagram, but cumbersome to present pairs of all five dimensions.

6. IMPLICATIONS

As promised there is no specific prescription for how to do maps. However, it is possible to extend some of the points made into suggestions for mapping practice.

Objectives

Clearly, when the objectives relate to market structure and positioning evaluation then mapping has an important role to play. But what about the more methodological uses of maps? Why is it that there is such a great need for data reduction? Are we asking too many questions that don't need to be asked? Use of prior studies with factor and discriminant analysis can achieve data reduction in the design stage of the next study. It is important that we act on this information in order to keep costs of research under control and to keep respondents willing to participate.

In the area of tracking we ought to ask whether it is realistic to expect change in the relative position of a brand on a frequent, e.g., monthly, basis? By the nature of the map itself, change is difficult to observe. It captures a whole constellation of fairly enduring images, which are embedded in a net of many variables. The movement of any one brand is therefore constrained. Perhaps the emphasis in tracking research should be placed more upon the recall of recent advertising and other variables that have the potential to move quickly in response to current marketing actions. The first step is to understand how these factors translate into an overall brand image. Once these are well understood, then hypotheses about the movement of images themselves can be established and better measured.

Methodology

The primary implication is to assess the mapping issues early in the design phase of the study. The goals of establishing the appropriate sampling design and minimizing respondent burden are easy to agree with, but not always easy to implement due to competing objectives. If the analyst gets involved early in the process, there is at least the prospect of being heard.

In many research situations there is an inherent choice between gathering a large amount of information from a few people, versus gathering small amounts of information from many people. The design considerations raised here suggest that the latter option is preferable. In the extreme, a very large representative sample could be drawn and each respondent be randomly assigned to rate just one brand. This is preferable to having fewer people rate many brands for several reasons. First, any environmental correlations are natural and truly representative of the current situation. If excessive use is made of over-sampling, quotas, and priorities, then environmental correlations are natural and truly representative of the current situation. If excessive use is made of over-sampling, quotas, and priorities, then environmental correlations will be design driven, and not necessarily representative. Second, more refined analysis, e.g. to examine the impact of familiarity with brands can be accomplished by subsetting the data, rather than by elaborate weighting schemes. Third, the respondent burden problems are minimized, if not eliminated. Fourth, the concerns about "stacking" the data (e.g. Dillon et. al., 1985) go away. For example, if there is only one set of ratings per respondent, then there is no halo effect from brand to brand.

It might be argued that this approach will increase research costs dramatically. While possible, this is not necessarily so, if questionnaire designers are sufficiently disciplined to incorporate only the most important variables. If marketing research departments would adopt some degree of standardization across studies, then there would be ample data to consult, in advance, to determine what needs to be included.

Reporting

There is an opportunity to do more on the communication of results with the increasingly sophisticated technology available at relatively low cost. Software programs for the PC's allow for dynamic display of results; for example, rotating axes and viewing three dimensional solutions. The numeric data can be retrieved on the same screen to further elaborate the characteristics of individual data points (for example, market share data, media spending, product life cycle information). This is more intensive (compared to extensive) analysis than we are used to doing. Once the basic map has been produced, based upon hundreds or thousands of respondents, the aggregate results can be dissected more fully to understand patterns. One of the key advantages of this approach is that the analysis itself, on the screen, becomes the presentation. The challenge for us is to make the analysis compatible with the way the marketing decision maker thinks. The "ingredients" for doing this, in terms of methodology and computer resources, are available. To make it more useful, it must be packaged as an active part in a decision support system.

References

Dillon, William R., Donald Frederick and Vanchai Tangpanichdee, "Decision Issues in Building Perceptual Product Spaces with Multi-Attribute Rating Data," Journal of Consumer Research, vol. 12, June, 1985.

Srinivasan, V., P. Vanden Abeele, and I. Butaye, "The factor Structure of Multidimensional Response to Marketing Stimuli: A Comparison of Two Approaches," Unpublished manuscript, December, 1987.

PERCEPTUAL MAPPING

A Comparison of APM with Paper and Pencil Data

Herb Hupfer
Elrick & Lavidge

BACKGROUND

In designing a marketing research study, one of the issues that needs to be resolved is how the data are to be collected. In essence, there are three options available: in person, by telephone, or through the mail.

In some cases, there is not much choice in the matter. For example, if it is necessary to prepare and serve a food product to consumers in order to have them evaluate it, a personal interview at a central location makes the most sense. In many other situations, however, there is some leeway regarding how the information might be collected.

When there is a choice, the researcher must weigh the advantages and disadvantages of the options that are available. Some of the criteria that might be considered are as follows:

- Is one procedure less expensive?
- Is one technique easier to administer, leading to
  - easier training of interviewers?
  - fewer mistakes in the field?
- Will one method deliver the results more quickly?
- Is one procedure better able to gather more information, possibly because it is more interesting to respondents?
- Does one procedure have less bias associated with it than the others?
- Will one technique enable the researcher to gather data which are of higher quality?

In addition to the questions above, there is the question of whether the procedures under consideration will produce similar results. Naturally, researchers hope that the outcome of a survey is not heavily dependent on the procedure used to secure the information.

ISSUE TO BE RESEARCHED

The issue of concern in this paper is perceptual mapping. In particular, how does the manner in which the data are collected for a perceptual map impact, if at all, the map itself?

There are a large number of ways of developing perceptual maps; some are very sophisticated and some are just a cut above "gut feel." This paper focuses on two methodologies for producing perceptual maps, Sawtooth Software's Adaptive Perceptual Mapping (APM) which collects its data using a personal computer, and perceptual mapping done via a paper-and-pencil interview.

APM utilizes multiple discriminant analysis as an analytic procedure. Therefore, in the interest of parity, the data collected with paper-and-pencil also will be analyzed using multiple discriminant analysis.

Before delving into the issue of similarity, the prudent researcher should recognize the advantages and disadvantages of each procedure. Even if the two procedures produce identical results, these other factors still need to be considered before a choice between techniques can be made.

On the positive side, APM, being a computerized data collection technique, is interesting to respondents. Most respondents appear to like the idea of "playing with the computer." Maybe it's the idea they do not know what will appear next on the screen or they can somehow "beat" the machine that intrigues them. In any case, very few respondents get up and walk away once the interview has begun.

Clients also like the notion of collecting information via the computer. Some clients feel that any data collected using a computer are better data. To some extent, their argument is based on the computer's ability to randomize the order in which items (such as products to be evaluated) are presented to study participants. The computer also is capable of remembering answers to previously asked questions (in order to insert the answer in questions that follow) and automatically handles skip patterns. While the quality of the data gathered is a debatable issue, research suppliers are not likely to encounter many clients who think that computerized interviewing is a poor way to collect the needed information.

APM is a self-contained perceptual mapping procedure. The questionnaire portion of the study and the data analysis part are combined in a neat package. To some organizations, this enables them to avoid subcontracting the analysis portion of the study to an outside vendor.

Collecting the information via paper-and-pencil and analyzing it using a standard discriminant analysis package also has its advantages. One advantage is flexibility. The data can be gathered using a telephone, personal, or mail survey whereas APM needs to be administered in person. In most cases, the need to use a personal interview causes the interviewing to be clustered, which for some, would raise sampling questions. Moreover, personal interviews are likely to be more expensive than other forms of data collection.

Paper-and-pencil questionnaires also have the advantage of being easily amended in the field. Unintelligible instructions, missing words or troublesome questions can be corrected quickly at the interviewing site by writing on the questionnaires. Few field firms, however, are capable of making on-site corrections to a computerized survey.

Problems, if any, are easier to spot when paper-and-pencil interviews are conducted. This is particularly true for certain aspects of the questionnaire. For example, if the product selection process in APM is not properly programmed, respondents may be asked to rate products with which they are not familiar. This problem may be particularly difficult to detect if the respondents are familiar with <u>some</u> of the products they are asked to rate. The respondent assumes this is just part of the program and the interviewer has no way of knowing how the products to be rated are selected.

Paper-and-pencil interviews allow the researcher to use data analysis packages they currently have on hand (and for which they already have paid). This enables them to avoid the expense of buying a computer package which may not be used frequently.

Virtually any field organization can administer a paper-and-pencil questionnaire. APM requires the field firm to have a computer <u>and</u> have a computer which is compatible with its programming. Additionally, some of the interviewers at field firms are timid about using computers. Generally speaking, younger interviewers seem to have fewer difficulties in working with the computers. This may be the result of an increased level of computer literacy among younger people.

If the researcher plans to produce a perceptual map for different sub-groups of the sample, no special preparation is required above making sure the variables of interest are contained within the questionnaire. With APM, however, a Ci2 (Computer Interactive Interviewing) questionnaire is needed in order to collect the information that will be used to divide the sample into sub-groups. For some companies, Ci2 is a "second language." Therefore, it is not efficient to use. Moreover, because a computer interview is necessary, the trouble and expense of programming and debugging the questionnaire become possible problems.

COMPARING THE TWO PROCEDURES

In order to investigate whether APM and paper-and-pencil interviews produce equivalent perceptual maps, a "mini" study was conducted by the Chicago office of Elrick and Lavidge. In an effort to hold down the costs associated with conducting the study, all of the interviewing was done in Chicago.

In deciding on the products to map, a number of categories were considered: automobiles, watches, perfumes, and supermarkets to name a few. In order to avoid screening for buyers (users) of hard-to-find brands, the decision was made to map supermarkets.

The proposed comparison between procedures presented a number of other design issues. One possibility was to collect information from two matched groups of respondents, one group using APM and the other utilizing the paper-and-pencil approach. The maps resulting from the data collected from each group would then be compared to one another.

A second potential approach was to have respondents go through both APM and the paper-and-pencil procedures. This alternative essentially eliminates the chance that differences in the maps produced by the two procedure are due to real differences in perceptions between the two groups.

The two-sample method has merit from the standpoint it keeps the interview short, eliminating the need for any incentive to induce respondents to participate in the interview. It also avoids the chance of one procedure having an impact on the other.

The single group approach excludes the possibility that the maps might show real perceptual differences between two groups of respondents, even though they have been carefully matched with respect to a number of key characteristics. It also precludes the chance that the groups were not matched on the appropriate set of characteristics. For this particular study, the decision was made to go with the one group approach.

The data were collected at a mall intercept interviewing facility located in the Ford City Shopping Center, a large indoor shopping mall located in the southwestern part of Chicago. Supermarkets within a reasonable driving distance of the mall were selected as the stimuli to be mapped.

As it turned out, there were five large-size supermarkets in the immediate area of the mall. The specific stores selected were as follows:

- Aldi's
- Butera
- Cub Foods
- Dominick's
- Jewel

Both males and females were included in the study. Participants in the survey were screened to meet the following qualifications:

- Primary food shopper

- No critical industry employment

- 18 years of age and older

In addition to the above screening criteria, participants in the study also had to indicate that they did most of their grocery shopping in one of the supermarkets targeted for mapping.

A total of 200 interviews were completed. In order to minimize the likelihood of the two procedures biasing one another, the order in which respondents were exposed to the techniques was rotated.

Participants in the study were asked to rate the supermarket at which they do most of their shopping on a number of criteria which are likely to be important to consumers when choosing a food store. The criteria used to evaluate the supermarkets were as follows:

- Has competitive prices
- Is interested in me as a shopper
- Is conveniently located
- Has ample parking facilities
- Has convenient store hours
- Has the largest selection of stock
- Is the first with new products
- Has a floor plan that is well laid out
- Offers pharmaceutical products
- Offers produce/meats that are of the highest quality
- Offers a check-cashing service
- Has quick service in the check-out line
- Offers deli products
- Offers bakery products
- Has courteous/friendly employees
- Accepts manufacturers' coupons
- Has a clean store
- Has wide store aisles
- Is well lit
- Has good sales on the products in which I am interested
- Offers good promotional products (dishes, luggage, etc.)

In addition to rating the supermarket at which they shop most often, respondents also were asked to rate some of the other stores (maximum of two) at which they had shopped for groceries.

RESULTS OF THE TEST

Given that the two procedures go about collecting the data they need for perceptual mapping in entirely different ways, the assumption was made that it would be very unlikely that the two techniques would produce precisely the same results. The key point to keep in mind, however, is whether the researcher would arrive at the same conclusion using either procedure. Naturally, it would be disconcerting to learn that the manner in which the data are collected can change the outcome of the study.

The perceptual map produced by the computerized approach (Sawtooth Software's APM) is shown in figure 1. The map developed from the paper and pencil data using discriminant analysis is depicted in figure 2.

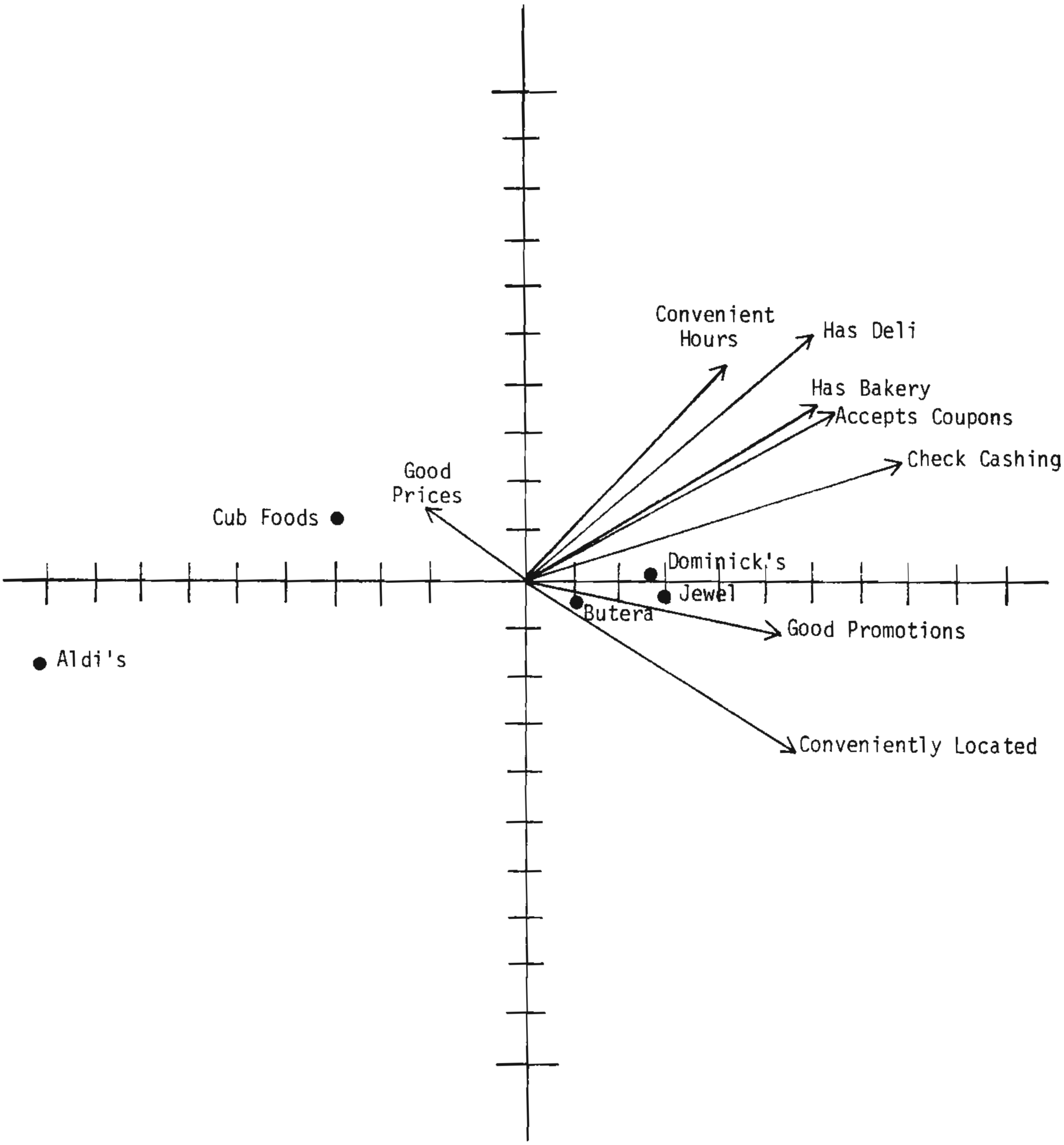

Convenient Hours
Has Deli
Has Bakery
Accepts Coupons
Check Cashing
Good Prices
Cub Foods
Dominick's
Jewel
Butera
Good Promotions
Aldi's
Conveniently Located

FIGURE 1

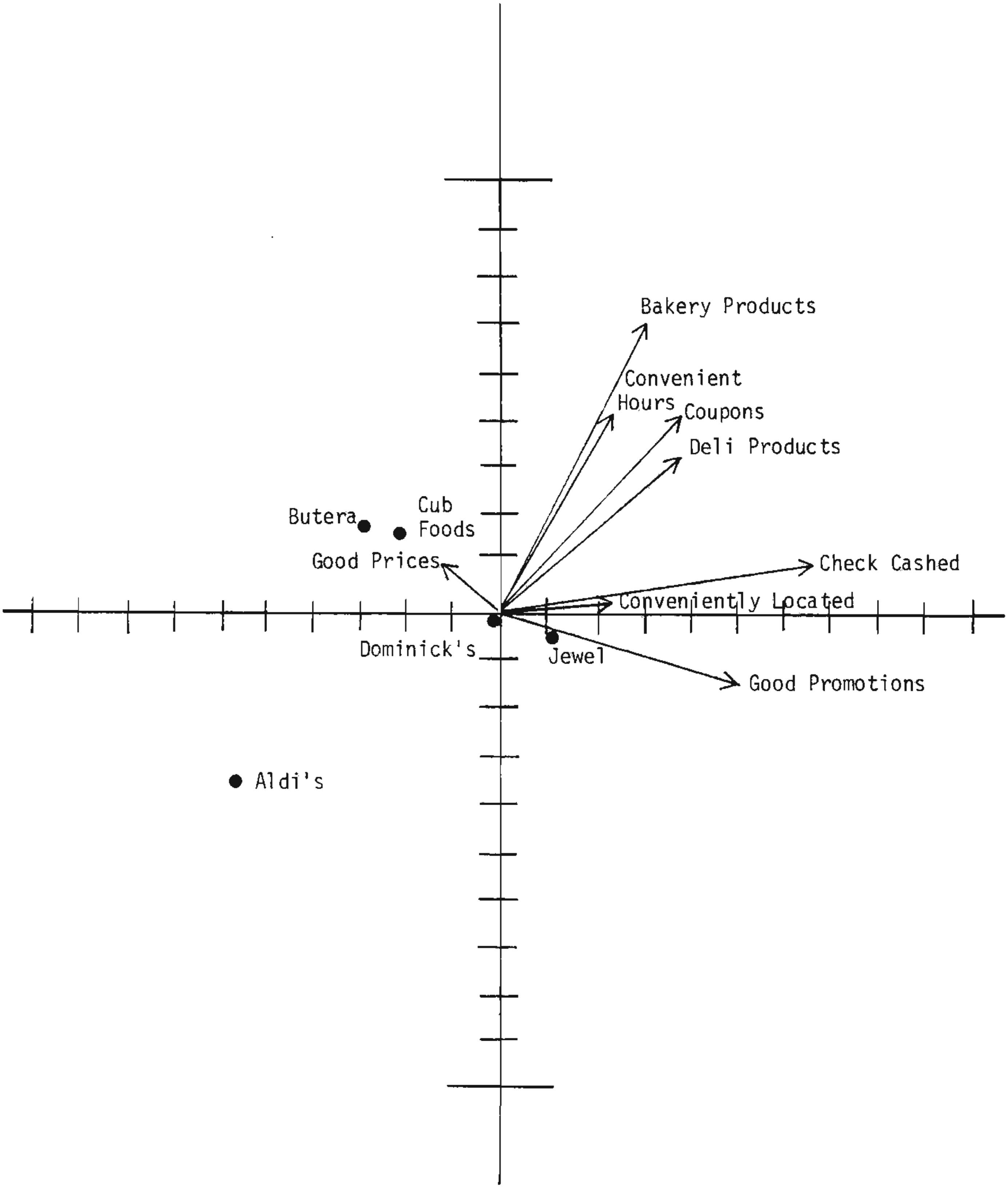

Bakery Products
Convenient Hours
Coupons
Deli Products
Butera
Cub Foods
Good Prices
Check Cashed
Conveniently Located
Dominick's
Jewel
Good Promotions
Aldi's

FIGURE 2

In general, the two maps clearly are more similar than they are different. Looking only at where the supermarkets are plotted in relation to one another, Dominick's and Jewel are viewed as being very similar, whereas Aldi's is viewed as being unique among the stores that were mapped. Cub Foods also is located in the same quadrant of both maps.

The paper-and-pencil map positioned Butera as being similar to Cub Foods whereas the APM map positioned Butera as being more like Jewel and Dominick's. At least part of the difference is explained by the relatively limited number of respondents who reported shopping at Butera most often.

The APM program indicated that there were three significant dimensions on which the supermarkets were differentiated, at the 95% level of confidence. These three dimensions accounted for 53% of the total variance. The discriminant analysis on the paper-and-pencil data suggested that there were only two significant dimensions differentiating the supermarkets. The two discriminant functions accounted for 81% of the total variance in the data.

The vectors plotted on the two maps are very similar with respect to their location on the maps. Virtually all of the vectors are in the same position. This being the case, the two main axes of the map would be similarly labeled. The horizontal axis appears to represent a dimension which could be labeled as "full service" while the other dimension might be titled the "price" axis.

There are some differences in the length of the vectors between the two maps. In most cases, however, even though the vectors would differ somewhat if rank ordered, the researcher viewing the maps is likely to arrive at similar interpretations regardless of which map is used.

CONCLUSIONS

The results obtained from APM and perceptual mapping based on paper-and-pencil data were very similar in this study. Whether these results can be generalized to the mapping of other stimuli is not established at this time.

Both procedures have merit under certain conditions. It is comforting to believe, at least for the moment, that the conclusions drawn using these procedures are not likely to be influenced heavily by the manner in which the data are collected.

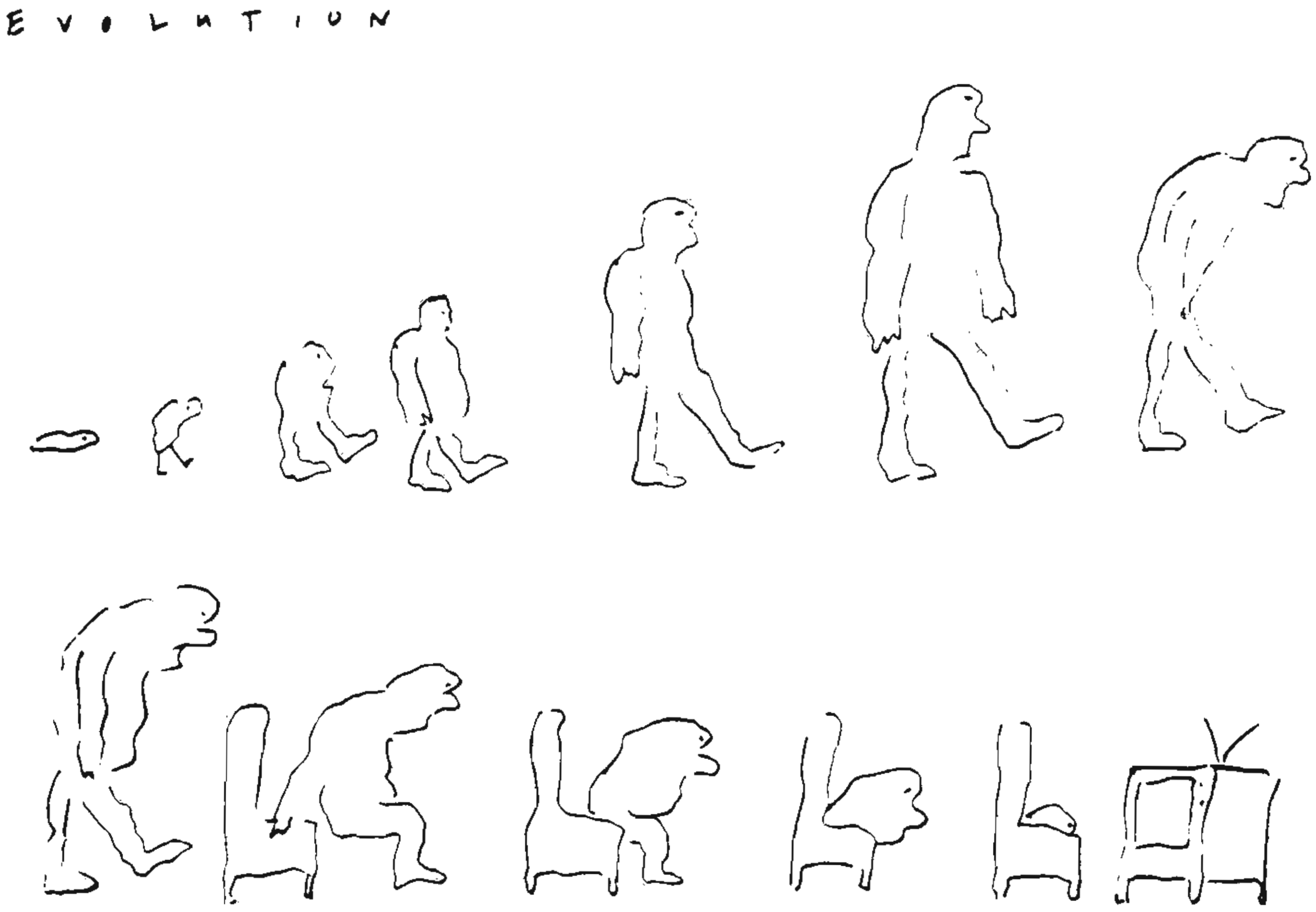
EMERGING TECHNOLOGY AND ITS IMPACT ON DATA COLLECTION

Vincent P. Vaccarelli
Xerox Corporation

INTRODUCTION

EVOLUTION

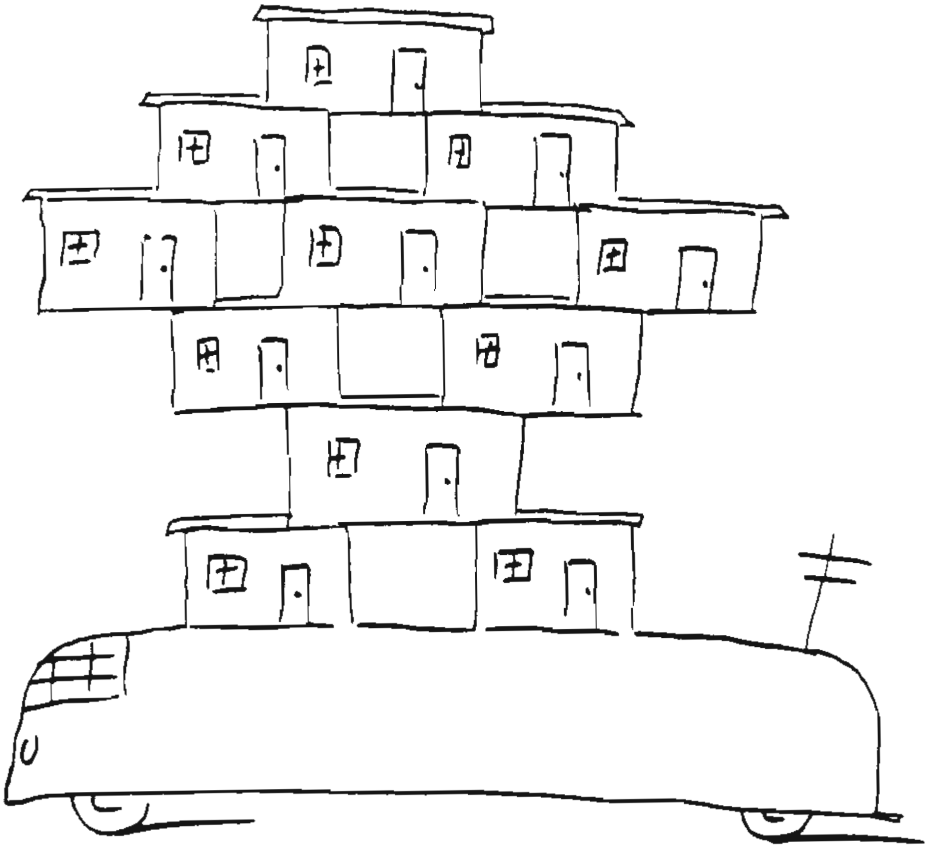
Discussing Emerging Technology and Its Impact on Data Collection should respect the notion of evolution, where we have been and where we are.



THINGS TO COME

It is only with some perspective can we discuss Things to Come.

THINGS TO COME: THE MOTORCONDO



In this discussion I'd like to regard some of our problems in data collection and what we have done so far to solve these problems, including, of course, what Sawtooth Software has done.

Then I'd like to stimulate your thinking regarding things to come.

CURRENT INTERVIEWING LIMITATIONS

Current interviewing limitations that may be solved by the application of newer technologies include the two broad categories of Respondent Limitations and Measurement Limitations.

RESPONDENT LIMITATIONS

Respondent Limitations include the areas of Comprehension, Retention, and Evaluation.

COMPREHENSION

The limitations in respondent capabilities include the respondent's human inability to really consider ALL attributes of interest, simultaneously. Of course, ACA's (Adaptive Conjoint Analysis) use of systematic paired comparisons tends to avoid the problem of a respondent having to comprehend several product profiles simultaneously.

RETENTION

Another limitation is found in the respondent's inability to retain the value expressed for attribute levels presented earlier RELATIVE to other attribute levels presented later. ACA seems to solve this problem by computing and retaining values for the respondent, while asking respondents only for intensity of choice.

EVALUATION

During a one hour interview, it is difficult for a respondent to be as studied in his choices as over the typical period of actual DELIBERATION. Again, ACA's paired comparison approach seems to allow far closer and more realistic study than whole-impression measurements offer.

MEASUREMENT LIMITATIONS

Measurement Limitations include problems in wide Differences in Respondent Interests, Interviewer Influences, and MisMatches between Maps and Territories.

DIFFERENCES IN RESPONDENT INTERESTS

Not all respondents are interested in same attribute SUBSET. ACA handles such differences through a "Simalto-like" approach which focuses attribute questions on each respondent's own interests.

INTERVIEWER INFLUENCES

The personality or mood of the interviewer can be a major extraneous CONTAMINANT on measurements. ACA, of course, controls for this contaminant by interfacing the respondent with the interview, not the interviewer. (We prefer to utilize a highly CONTROLLED random selection of respondents with small group administration though laptop computers; followed by a group discussion to reveal possibly EXPLANATORY variables.)

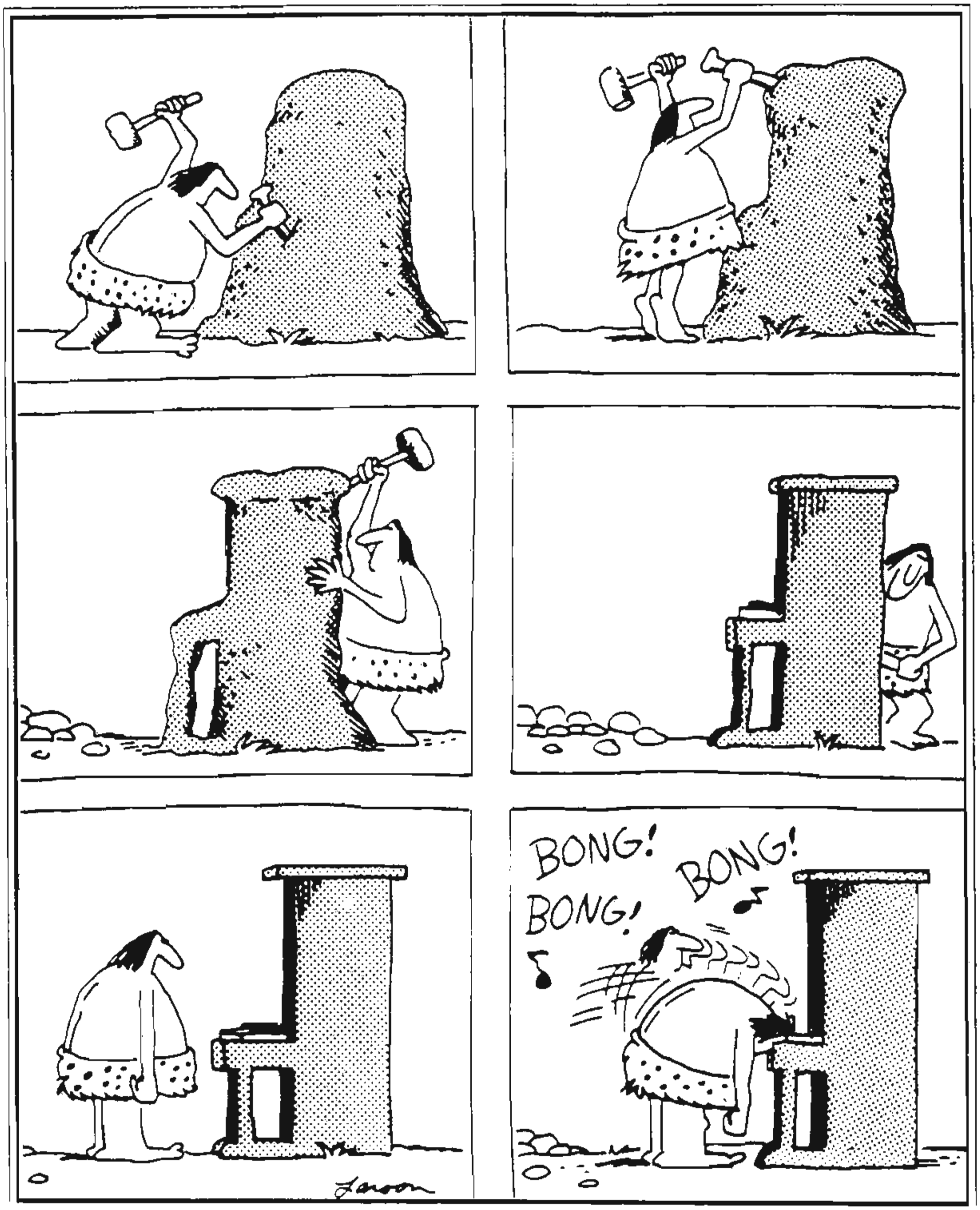
MIS-MATCHES BETWEEN MAPS AND TERRITORIES

There may be a significant difference between verbal description and sensory experience of the objects to be evaluated, such as restaurant menu descriptions and the actual dishes.

This is the area of Measurement Limitations that I think ACA, as yet, does not offer a solution for, but potentially it holds the key to the solution. The solution for which ACA may hold the key is in sensory simulation, rather than verbal description of choice objects and conditions, given the use of new technology to provide such simulation.

I foresee the use of ACA-like computer interviewing to control devices that will CUE sensory simulations of choice objects and conditions in order to make more accurate predictions of choice behavior.

In effect, technology should be used to reduce the difference between mind-sets and real choices.



Neanderthal creativity

NEW TECHNOLOGY APPLICATIONS

Applications of technology in marketing research data collection, analysis notwithstanding, may be categorized as in the area of Measurement or Stimulus.

MEASUREMENT

Given our technical breakthroughs in medical and other scientific measurements, it is reasonable to expect marketing research to have a variety of future DATA CAPTURE elements.

These data capture elements run the gamut from paperand-pencil, video recording and other behavior monitors and scanners, the keyboard, mouse, and touch screen, to voice analyzers, eye movement analyzers, and other physiological monitoring elements from the galvanometer all the way to EKG, EEG, etc., and perhaps even to someday include brain image recorders such as suggested in the movie "Brainstorm."

STIMULUS

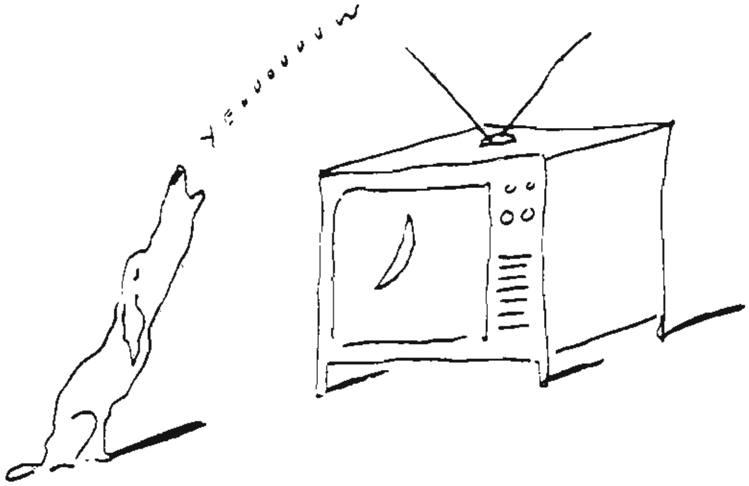
On the other hand, while sophisticated measurement elements may be common, sophisticated sensory simulation elements may not be so widely anticipated.

Simulations of choice objects and conditions may involve a wide variety of technology applications, involving such enabling system elements as slide projectors, audio tape recorders, VCRs, video-graphics screens, voice synthesizers, olfactory simulators, and holograms, along with other devices that may be more directly connected to the nervous system to simulate sensory input.

It may be possible, then, to foresee computer-based interviewing, such as ACA, being the key controlling element for devices that will cure sensory simulations of choice objects and conditions in order to more accurately predict choice behavior.

THESIS

My thesis, then, is that foreseeable applications of new technology in marketing research will focus on sensory simulation cueing as much as on data collection. Further, I believe that this application of technology will add the validity we require to the precision we now have. That is to say, now that we have the ability to accurately measure the Response, we need the ability to accurately present the Stimulus. I believe, therefore, that our next use of technology should be to help us ask our questions as well as get our answers.


Yeououuuw

COMPUTER INPUT INTERFACES

Technology applications can be viewed as solutions for otherwise unsolvable problems.

PROBLEM: RESPONDENT REACTION ERRORS

Errors may be introduced by the use of devices to capture reactions.

INCONSISTENCY

There may be disconnection between seeing a concept choice on a screen, and pressing a key on a keyboard.

IMPRECISION

The selection of a number may not precisely express an intensity of behavior-indicative preference.

FATIGUE

Boredom or stress may result from repetitive demands for routinized responses.

EXAMPLE: KEYBOARD ADMINISTERED INTERVIEWS

For many respondents the use of keyboards may be similar to using a foreign language, requiring constant mental translations and leading to frustration and errors.

SOLUTION: ANALOG MEASUREMENTS

The application of analog measurements may reduce some of the reaction errors caused by digital input devices.

CONSISTENCY

The response reaction may be more connected to the object choice behavior.

PRECISION

Intensity may be gauged through correlated reactions rather than digital translations.

STIMULATION

Boredom and stress may be reduced by variety and actions.

EXAMPLE: USE OF PRESSURE-SENSITIVE TOUCH SCREENS

The use of pressure-sensitive touch screens can reduce errors by making an otherwise tedious task more "natural" and fun, while measuring intensity of preference through intensity of touch.

COMPUTER CUEING INTERFACES

Computer cueing interfaces should also involve the application of technology to address problems requiring solutions.

PROBLEM: ATTRIBUTE DESCRIPTION

Some product attributes may be impossible to describe verbally within reasonable limits.

COMPLEXITY

The attribute may be far too complex to describe effectively.

UNFAMILIARITY

The attribute may be insufficiently relatable to the respondent's previous experience.

DISTORTION

Verbal description may not convey the sensory qualities of the attribute.

EXAMPLE: COPIER JAM CLEARANCE

Descriptions of alternative methods to clear a jammed sheet of paper from a copier paper flow path are prone to complexity, unfamiliarity, and distortion.

SOLUTION: OPTICAL DISK VIDEO

The application of optical disk video technology may allow a more accurate portrayal of the attribute levels.

A PICTURE IS WORTH A 10,000 WORDS

If a respondent may opt to see a video portrayal of any attribute level asked about, complexity is less likely to be a obstacle.

RECOGNITION

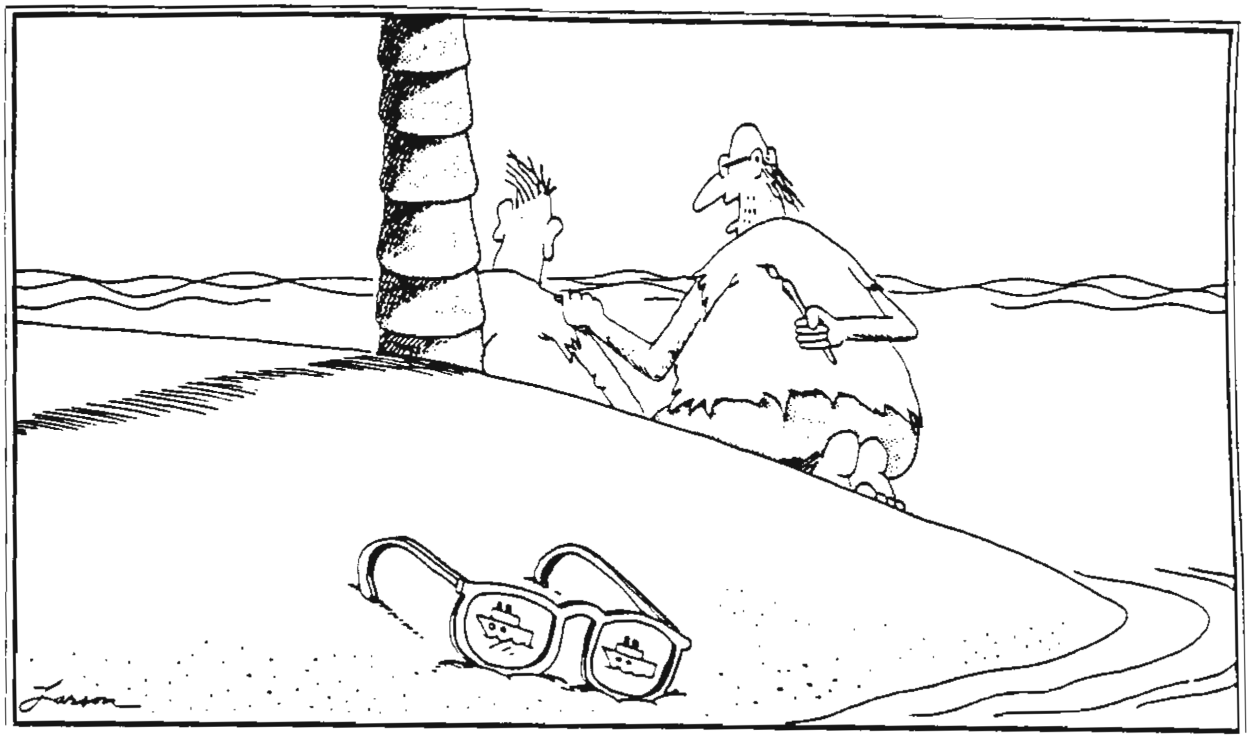
The power of video to prompt recognition and understanding can compensate for even a complete lack of familiarity with some attributes.

CLARITY

The illustrative capability of video is likely to be the form of communication most free of distortion.

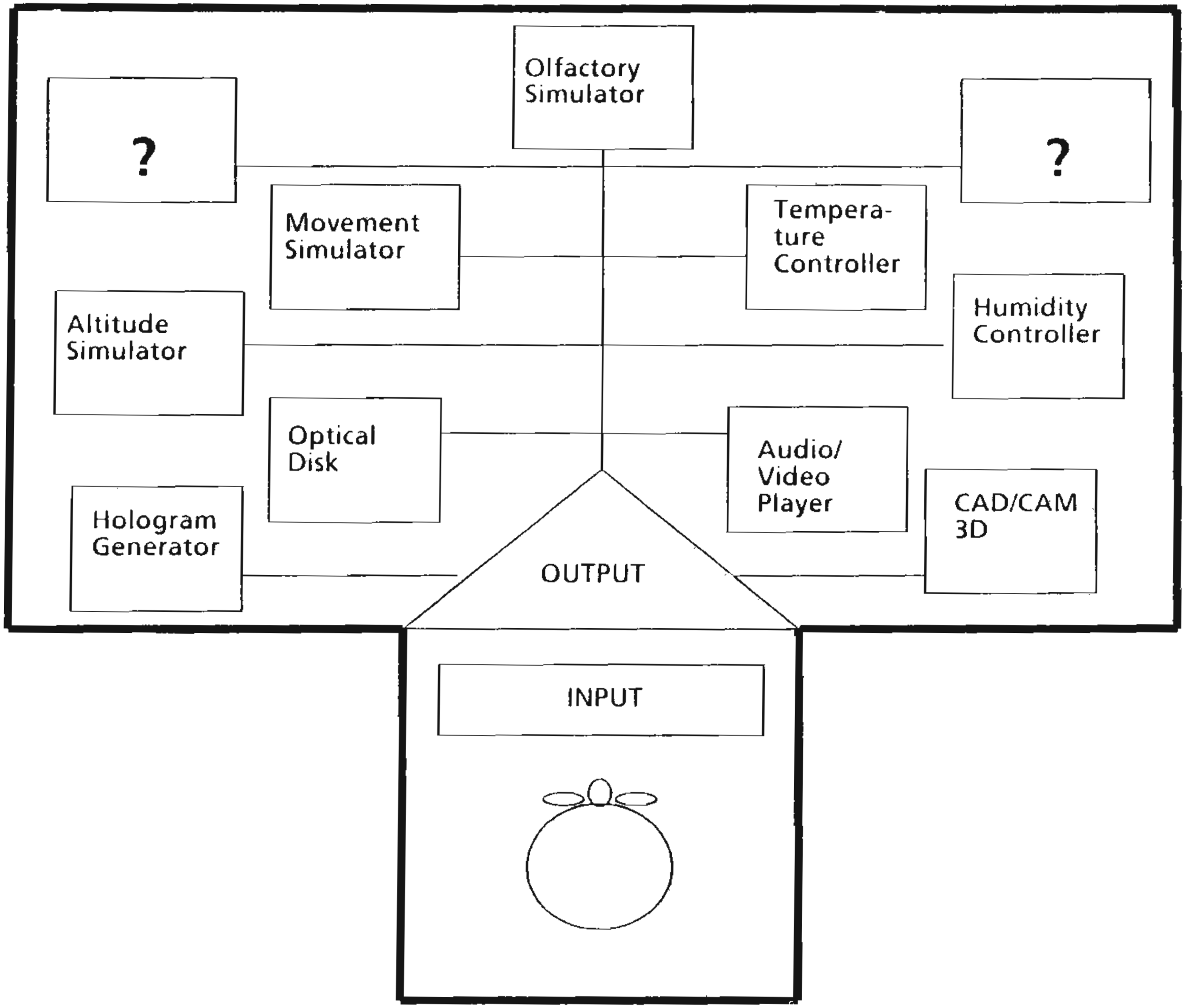
EXAMPLE: VIDEO SEGMENTS OF JAM CLEARANCE ALTERNATIVES

Accessible video segments, each portraying an alternative paper jam clearance method, can effectively and validly ask for respondent choice.



"Bob! Wake up! Bob! A ship! I think I see a ship! ... Where are your glasses?"

CHOICE SIMULATION SYSTEM



EXPECTATIONS FOR OPTICAL MASS STORAGE

Optical mass storage is likely to be a major enabling technology, especially as the ability to both read and write becomes more affordable.

VIDEO DESCRIPTIONS

The ability to scan to a computer-controlled video segment to convey a choice object is most efficient with optical disk.

TEXT/PROGRAM CAPACITY

Far more extensive programs may be permitted, and/or many more choice variations may be accessible.

AUDITORY INSTRUCTIONS

Auditory instructions and guidance may be used to support respondent tasks.

DATA STORAGE

Many more interviews and algorithms can be contained for PC simulations or data files, including graphics.

DISK REPORTS

Reports on optical disks containing video presentations as well as data bases and simulation programs, scanned-in industry literature, related government statistics, etc. - accessible through a table of contents menu - may round out the initial uses of optical disk technology in marketing research

INTERACTIVE RESPONDENT NETWORKS

Interactive respondent networks may evolve to offer unique advantages.

DETERMINING AREAS OF SHARED AGREEMENT FOR PREDICTIONS

Interactive interviewing networks may be used for efficient reiteration in Delphi Studies or Prophesy Panels.

MEASURING THE INFLUENCE OF SHARED INFORMATION

The influence of shared information may help reveal word-of-mouth patterns and product evaluation strategies, with built-in forms of content analysis.

BRAINSTORMING

Customer or executive brainstorming, utilizing expert systems modeling forms of Synetic, Osbourne, or Morphological Structuring techniques, perhaps utilizing artificial intelligence.

"Hey! Look what Zog do!"

HUMAN FACTORS CONSIDERATIONS

In the adoption of new technology for marketing research we must respect the human neurology of information perception, processing, learning, and expression.

NEURO-LINGUISTIC PROGRAMMING [TM]

We perceive, process, learn, and express ourselves using strategic sequences of our basic senses. Individual differences in the use and sequences of these senses account for individual differences in performance. If we are to understand and predict performance, we should utilize technology in accordance with these human patterns.

PRODUCT SELECTION STRATEGIES

Product selection strategies may vary by type of consumer and type of product.

SEE-LISTEN-FEEL

Some consumers for some products (like office products) may SEE an advertisement, TALK/LISTEN to friends, and then get a FEEL for the product through a trial.

LISTEN-FEEL-SEE

Other consumer (perhaps for services) may LISTEN to friends, get a FEEL for the product, and then be convinced by SEEing effective advertising.

FEEL-SEE-LISTEN

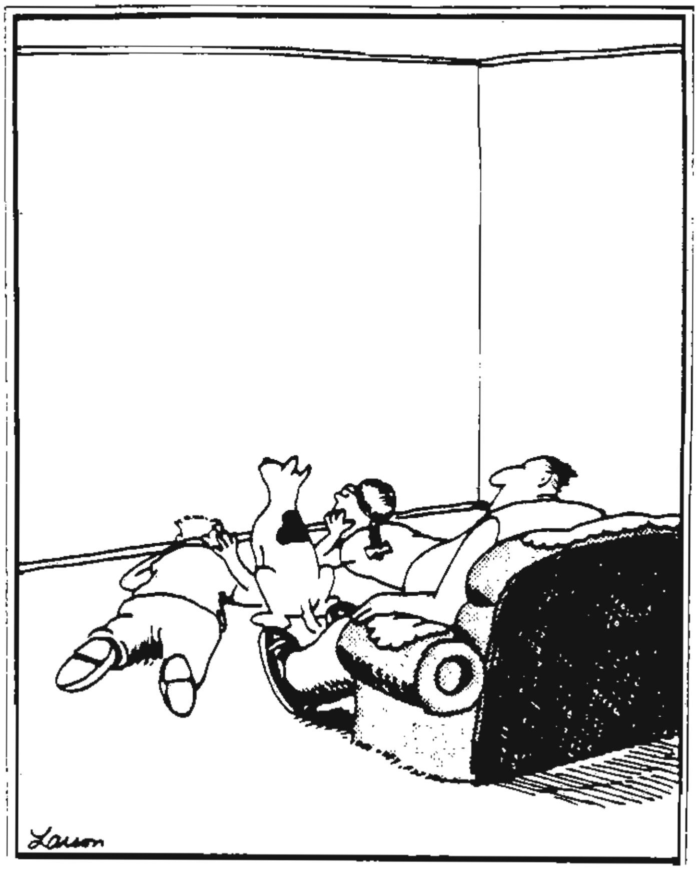
Still others (perhaps for furniture) first get a FEEL for the product, SEE its advertising, and then LISTEN to friends.

ETC., ETC. (ALL POSSIBLE COMBINATIONS WITH TWO OR MORE)

TECHNOLOGY IN RESEARCH SHOULD RESPECT HUMAN PROCESSES

If we are to simulate choice, we should simulate choice objects and conditions in an attempt to attain validity. Whether through cueing or gauging, we should try to use research technology in correspondence with the way consumers might select products, not only in content, but also in sequence.

COMMERCIAL-DISCUSSION-DEMONSTRATION
DISCUSSION-DEMONSTRATION-COMMERCIAL
DEMONSTRATION-COMMERCIAL-DISCUSSION



In the days before television

SUMMARY

TECHNOLOGY: SOLVING OTHERWISE UNSOLVABLE PROBLEMS

As in other applications of technology, those that are most highly valued and durable solve otherwise unsolvable problems of importance.

SENSORY SIMULATION OF CHOICE OBJECTS/CONDITIONS

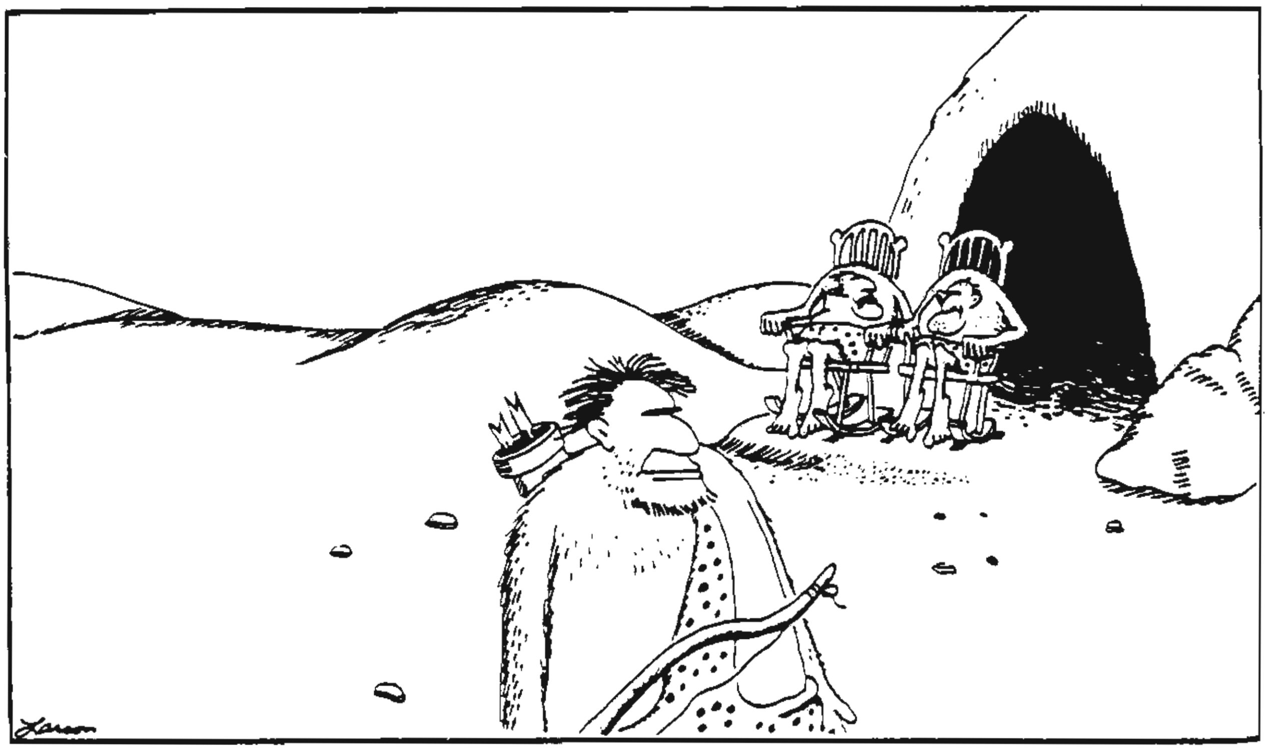
A currently unavoidable source of error in data collection may be found in the limited verbal descriptions of choice objects and/or conditions. The impact of technology on data collection may be most useful in sensory simulation of choice objects and/or conditions, rather than further improvement in measurement devices.

HUMAN FACTORS MUST BE CONSIDERED

Technology applications should follow, rather than lead, human processes. If not, these applications may only increase rather than solve problems.

RESPONDENT FEEDBACK

We must use respondent feedback to guide our current applications and future considerations of technology in marketing research.



"Look at that! . . . Give me the good old days when a man carried a club and had a brain the size of a walnut."

INCREASING THE USE OF MARKET RESEARCH AND THE STATUS OF MARKET RESEARCHERS

Marc Prensky
MicroMentor, Inc.

The idea for this presentation was suggested to me by observations I made during my six years at the Boston Consulting Group. During my first three years there, I was a strategy consultant, helping senior management of large companies decide on their future strategic directions. During my second three years I served, among other roles, as a market researcher, specializing in conjoint analysis. I therefore gained perspective both as a user and a provider of market research in the context of senior management decisions.

What I am going to focus on here are senior management needs in the area of market research, and how and why market researchers often fail to fulfill them despite their best intentions and efforts. My thesis is that if senior executives learn more about how market research can serve their needs, their use of research is likely to increase. If, at the same time, researchers can learn to better fulfill those needs, their status is likely to increase as well.

Let me begin with a story. While at BCG, I sold and performed a conjoint analysis for a division of a large international company, for which the client paid a great deal of money. The client's problem was how to price each of the two new products they were planning to sell in the second-generation competitive marketplace, so as to maintain their dominant market share. We did a lot of fairly sophisticated analysis for them, such as plotting iso-share curves under different pricing scenarios, and contrasting the value systems of different user and influencer groups. Basically, however, all the data came from ACA (Adaptive Conjoint Analysis). Just two days before our final presentation, I learned that the division's market researcher had recently purchased the ACA package. I spent a good many hours worrying about how the client would react to the enormous difference in cost between the program and the study, and tried, mostly in vain, to think up even more sophisticated analyses to support our rather straightforward conclusions. As it turned out, however, there was no problem. The study was well received by the company CEO and management committee, and the conclusions, one of which was somewhat counter-intuitive, were believed. When I mentioned their purchase of the program, the CEO's response was this: "The software may be the same, but we would never believe that guy anyway."

This story has two morals. The first is that in-house researchers generally do not have credibility with top executives unless they earn it. Too often, however, their actions cause them to lose, rather than gain credibility with top management, who go elsewhere for advice that the researcher could, in fact, provide. The second moral is that should your company ever hire an outside consultant or firm to do market research, you would do well to try to find out what software they are using. If it is commercially available, buying it and displaying it prominently on your desk probably make the consultant work a whole lot harder to give you value for your money.

The real question, however, is this: What makes one researcher believable by top executives and another not?

Clearly, many factors are involved: relationships, reputation, and previous experiences are among the most obvious factors leading to believability and trust. However, one can add to this list some equally important factors that are perhaps less obvious, and, I think frequently overlooked.

In my view, the believable researcher is the one who:

- understands the thinking and business needs of decision-making executives,

- meets those needs specifically and directly, and

- communicates the information concisely, in language the executive uses and understands.

Conversely, the researcher executives tend not to believe, to ignore, or to hire an expensive consultant to replace, is the one who:

- doesn't understand business thinking and decision making,

- delivers masses of data and random conclusions rather than concise answers to specific problems,

- gives presentations filled with jargon, technical details, unnecessary degrees of precision and arcane, non-intuitive levels of explanation.

The task of senior managers is to make decisions that, it is hoped, create value for their company. They generally do this by building up a picture of how the world behaves, by generating some hypotheses about how it might behave if they take certain actions rather than others, and by testing those hypotheses by answering a few key questions. Market research is often the best way to test an hypothesis, and as a strategy consultant, one of the first things I did when I began a project was to ask for any previous market research that had been done in the area I was studying. I generally received a number of thick books, put together by in-house people or by outside research firms, containing more data than one could ever want or use. However, I would say that nine times out of ten, when I looked for the specific data or cross-tab that was needed to support a particular strategic hypothesis, it wasn't there. Why? Possibly because things had changed since the research was done. But very often it was not there because the researchers and the executives had not communicated well enough about what the specific business hypotheses and decision support needs actually were.

Business executives are often quite unaware of modern market research techniques, and how they can add value to their decisions. Particularly in non-consumer goods environments, market research, if done at all, often tends to be extremely general, or "scatter shot" in its approach. At MicroMentor we work exclusively with senior executives and are often amazed at the things they don't know. Although they frequently have business degrees, and are usually good intuitive decision makers, we find they are often completely unaware of how many modern business tools and techniques, sometimes even such simple ones as financial ratios, deflator indices, or net present value analysis, for example, can be used to improve their decisions. Although it may seem strange, terms like regression, conjoint analysis, factor analysis, discriminant analysis, and certainly anything in statistics are, to most executives, completely meaningless, even if they may have heard them once or twice. One of the most important skills for researchers therefore, is to be able to communicate to executives precisely how and where particular research techniques can add value and support their business decisions.

For example, if one can convince an executive of the value of using conjoint analysis in new product creation, he is likely, in my experience, to use it again and again. I had one client in a large company who at the time I first worked for him was a planner in headquarters. He was looking for the cause of a particular computer division's poor performance, and had a particular problem. The division's product planners were maintaining that the cause of the problem was lack of IBM compatibility in the product. The planner was convinced this was not right, but needed evidence to back his intuition. We did a conjoint analysis for the planner, using subjects who had both bought and decided not to buy the client's product. The results showed clearly that IBM compatibility was of little interest to the users, but that another feature, not even under development by the division at the time, was <u>extremely</u> important. The planner used the evidence to convince his boss to make important changes. This planner subsequently went on to become the head, in turn, of a number of the company's divisions. So impressed was he by the power of that first conjoint, that he has done one for every subsequent product for which he has had responsibility.

The example illustrates a number of interesting points.

First, the executive learned to associate conjoint analysis with product design. This simple connection greatly increased the number of research projects he requested.

Second, the executive needed an answer to a specific question. This question was feature weightings. Market simulation was irrelevant to him in this context, and in fact, we did not even include it in the project. The message we delivered to the executive was an extremely simple one: here is what is important, here is what is not. No matter how complex a piece of research may be, or how many conclusions one may potentially draw, the only ones that will matter to the executive are the ones that are relevant to his problem at the time. This almost always reduces to one, two, or at most three key points that the executive should take away from the research.

Third, the executive realized that the value of conjoint analysis in product design, useful as it was in this case, would be even greater if the technique were used before, rather than after, a new product was actually created. The executive remembered this, and recently commissioned a conjoint on consumers' feature weightings of a product his new division is designing.

In general, researchers can gain credibility and status with executives by communicating to them three things:

- the VALUE of market research in solving executives' business problems

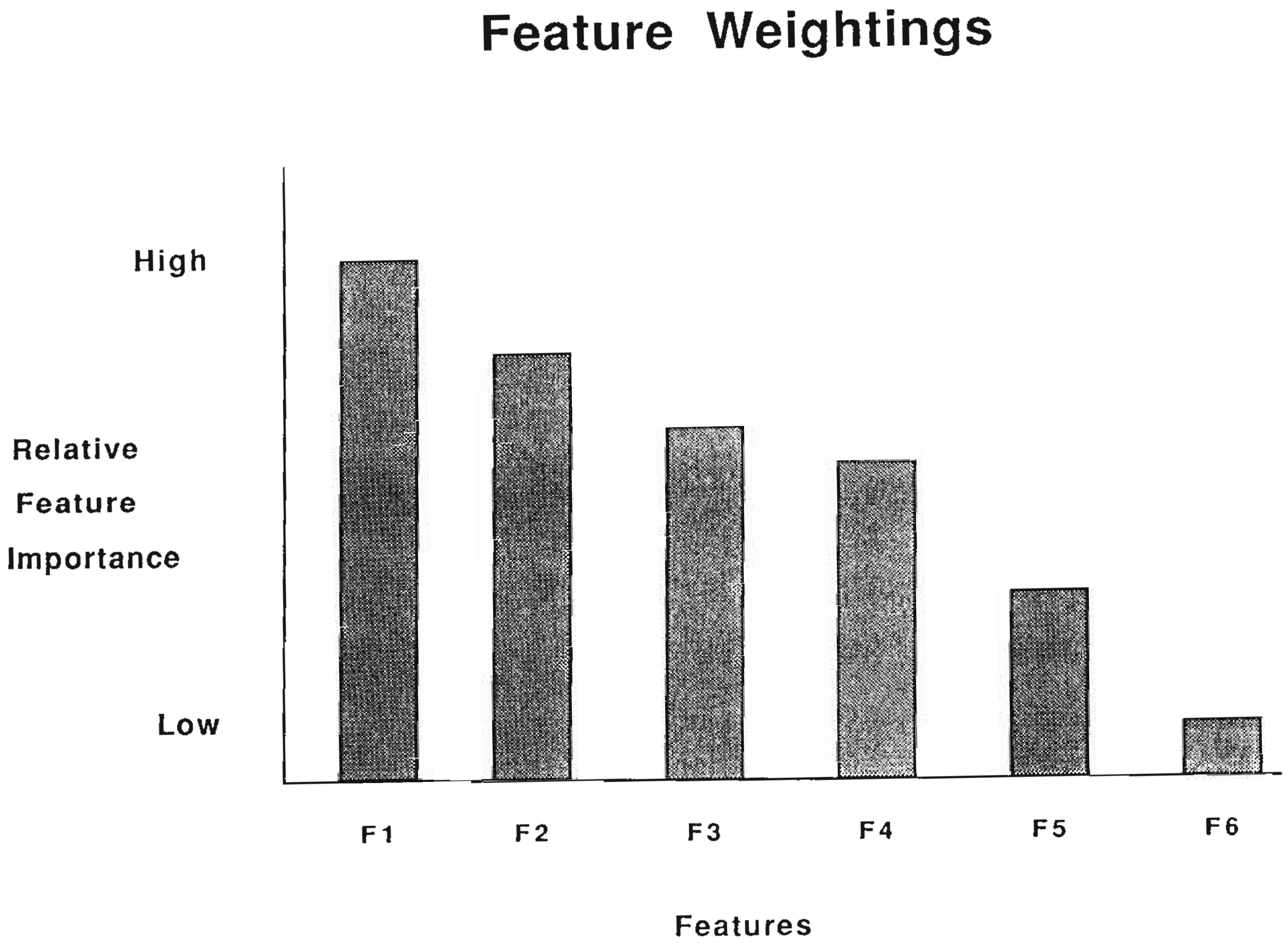
- the DECISION CONTEXT in which each of the various techniques work best

- the DIRECT, SIMPLE ANSWER to the executive's specific problem at the time.

Researchers tend to lose executives' interest quite quickly, and therefore lose credibility, when they go on at great length and in great detail about such things as sample size, statistical accuracy, validity, and so on, rather than concentrating on the executive's business problem. Of course it is possible, and even likely, that once the executive hears the direct, simple answer, he might question its validity, particularly if the answer is counter-intuitive. At that point the researcher, rather than going into obscure statistics and methodological comparisons, needs to be able to provide the executive with a concise, easy to understand, non-jargon based explanation of how the conclusion was reached. For example, to explain how weightings are derived, I often say that if every attribute were traded off against every other, one-on-one (all other things held equal), the attributes that win the most tradeoffs would get the higher weightings. While this is not strictly the way it is done, particularly in the ACA methodology, it is both close enough and intuitively understandable enough to work as a believable explanation. In a similar vein, I would chose to use an additive rather than a multiplicative simulator, even if it could be shown that the multiplicative model is slightly better, because the concept of a product's value being the sum of its attributes is much more intuitively understandable than its value being the product of the attributes.

While executives are generally looking for a specific answer, researchers can also lose credibility and status by being <u>too</u> precise in their predictions. When simulating market shares, for example, the researcher should understand that the real question on an executive's mind is usually "does this product's share appear likely to be within an acceptable range for us to make a profit, and will that profit be affected a lot or a little by particular product changes we or our competitors might make?" A more precise answer than that is not really needed, which is fortunate, because despite the apparent precision of the decimal points in the simulator models, the "range" question is the only one that conjoint can really answer. Few executives are likely to believe a prediction of, say, a 32.6 share for a product that does not even exist, and unless that result comes true, the researcher is unlikely to be called upon again. In presenting to senior executives, I always take great pains to explain that while conjoint is a very good directional predictor, it should be used as a directional predictor only. Conjoint is very useful for separating the important from the unimportant, and for identifying the things one should pay attention to and those one can afford to ignore. Conjoint is not as good, on the other hand, for discriminating between alternatives that are very close. To underscore this point, I try on my charts and graphs to use scales that are <u>not</u> numerical, (e.g. high/low), whenever possible.

FIGURE 1



Feature Weightings
High
Relative
Feature
Importance
Low
F1
F2
F3
F4
F5
F6
Features

When a researcher is called upon to present his results to senior executives, I would suggest the following rules: The higher the level of executive one is presenting to, the shorter, the more specific, and the less detailed the presentation should be.

For example, we all know that it is very important that a sample be balanced, properly stratified, and large enough to be extrapolated to the population one is studying. A final report would contain, probably in an appendix, precise descriptions of the sample, comparison of sample parameters with those of the population, statistical tests, etc. Were I presenting the study to a client's research staff, I would go through all of this material. Were I presenting it to middle management, I would probably summarize it in one or two tables. Were I presenting it to top management, I would most likely comment only that "the sample is statistically valid," and move on to the conclusions.

It is also quite important, when presenting to top executives, to have pre-thought-out, simple, easy to grasp explanations of the complex techniques one has used, (such as those I described previously for conjoint) to answer the inevitable "How does this work?" questions.

As I mentioned before, the senior executive will usually retain only one or two key conclusions from a presentation. It is extremely important, therefore, that the researcher decide which conclusions he wants those to be, and hammer them home in as many ways as possible. This involves not only understanding what the executive's decision problem actually is, but often requires pre-testing the results on some of the executive's staff, to see what strikes them as important, useful, or counter-intuitive. A very limited number of well-designed charts and word slides will often suffice in an executive presentation, leaving at least half the meeting time for discussion of the implications of the study's conclusions on the business decision, which is what the executive really wants and needs. As a BCG partner once instructed me, it is important to be "absolutely merciless" in cutting out everything that does not directly support the few points you want to get across.

FIGURE 2

Maintaining Executives' Interest: Do's and Don't's

DO	DON'T
• focus on business problem	• overemphasize statistics
• provide intuitive explanations	• revert to jargon
• provide directional answers	• add spurious precision
• emphasize major conclusions	• present everything you know
• be concise	• waste executives' time

Let me conclude by reiterating the few points I want you to take away from this talk. First, useful research is based on understanding the business problem the research is supporting. Second, the answers the research provides should be few, simple, and directly related to the problem. Finally, most, if not all, senior executives could benefit from further education as to the value of the various kinds of market research and how research can help them solve their business problems. As such education becomes more widespread, both the use of market research and the status of market researchers who use the tools properly will inevitably increase.

A SURVEY OF CURRENT PRACTICES OF COMPUTERIZED INTERVIEWING

Richard Miller
Consumer Pulse, Inc.

I'm going to discuss current practices in computerized interviewing. Although most of you share an enthusiasm for computerized interviewing, there are some cautions that come with the technology. To assist in my discussion, my firm completed two independent surveys.

First, we completed a telephone survey among a random sample of 100 members of the American Marketing Association. The sample was drawn from the AMA's Vocational Section of Private Firms. The purpose of this survey was to determine the extent that computer interviewing is being used in the marketing research community. One hundred WATS interviews were completed with principals of marketing research firms or department heads of marketing research departments of end-user organizations. We think the results of this study ARE projectionable to the entire market research community. We have analyzed the data in three groups: total sample, users of computer interviewing or those planning to use in 1988, and non-users of computer interviewing.

We also sent a computerized mail survey to the first 177 registrants to this conference to determine the current practices of people whom we perceived to be the most likely to use computerized interviewing. One hundred six interviews were returned, and 99 were completed and usable. After the cutoff date, an additional 7 were received for a total return rate of 64%. These results are NOT AT ALL projectionable to the entire user base of computerized interviewing. These data will also be presented in three groups: total sample, users of computer interviewing, and nonusers of computer interviewing.

First, let's take a look at the results from the random sample of AMA members:

Has your firm used computerized interviewing in the past year (1987)?

	TOTAL	Users/ Plan to	Non-Users
TOTAL	100	41	59
Yes	23%	56%	--
No	77%	44%	100%

Computerized interviewing is having major impact in the marketing research community. Today, 23% of the community is completing interviews using computers. Most of the research is being done on the telephone. However, a significant percent of the work is now being completed in malls and other central location test sites.

% of sales / budget for projects which utilized computerized interviewing

	TOTAL	Users/ Plan to	Non-Users
TOTAL	100	41	59
None	81%	56%	100%
Less than 20%	12%	27%	--
20 - 39%	4%	10%	--
40 - 59%	--	--	--
60 - 79%	2%	5%	--
80 - 99%	1%	2%	--
100%	--	--	--
Mean	3.68	8.93	0.03

It appears that less than 5% of all the research dollars being spent are for studies that include the use of computerized interviewing.

Also, two-thirds of marketing researchers think that computerized interviewing is changing marketing research, for the following reasons. . .

Why do you say that?
Based on respondents who indicate computer interviewing IS changing marketing research

	TOTAL	Users	Non-Users
TOTAL	90	67	23
Faster results	31%	30%	35%
More accurate data/ quality of data	22%	22%	22%
More sophisticated studies (conjoint)	11%	9%	17%
Less expensive overall/efficiency	11%	12%	9%
More complex studies/more control	9%	7%	13%
Labor shortage/less int. intensive	7%	4%	13%
Other	6%	6%	4%
No answer	20%	22%	13%

Other reasons for change include...

	TOTAL	Users	Non-Users
Easier for companies to enter field	4%	6%	-
High tech image / new technology	4%	6%	-
Less bias in coding open ends/verbatims	4%	3%	9%
More up front time / questionnaire setup	3%	-	13%
More capital investment	3%	4%	-
Don't Know	3%	4%	-
More central location interviewing	2%	3%	-
Don't computer interview	2%	-	9%
Respondent interest	2%	3%	-
Manufacturer doing own research	1%	1%	-
Need more client approval	1%	-	4%

It is also interesting to note that those who do not think computerized interviewing is changing marketing research think this way because they do not use computerized interviewing themselves or they don't see the change. It will even be more interesting to see what these people think next year.

How can we best assess the impact of computer interviewing? First, 23% of the firms are already doing computer interviewing, and an additional 18% of them intend to use CRT interviewing in 1988, which means we may expect 41% of these firms to be using computerized interviewing by the end of 1988.

Of the attendees of this conference, over 90% think the computer revolution is changing marketing research and for the following reasons. . .

--Faster results
--More accurate data / quality of data
--More sophisticated studies (conjoint)
--Less expensive overall / efficiency
--More complex studies / more control
--And they also think it helps solve the labor shortage since it's less interviewer intensive
--It's easier for companies to enter the field
--There's a high tech image / new technology
--There's greater respondent interest
--There are more manufacturers doing their own research

But they think . . .

--Researchers need more up front time for questionnaire setup
--There's more capital investment
--And researchers need more client approval

Let's change direction and consider the equipment used for marketing research.

Number of PCs in Office

	TOTAL	Users	Non-Users
TOTAL	99	72	27
None	5%	3%	11%
One	6%	6%	7%
2 - 3	17%	15%	22%
4 - 5	16%	14%	23%
6 - 9	12%	13%	11%
10 - 25	27%	30%	15%
26 and over	15%	18%	7%
Don't Know	2%	1%	4%
Mean	16.63	18.10	12.62

Most people at this conference have PCs in their offices. A typical office has an average of 17 PCs and 15% of the attendees' firms have 25 or more PCs.

Number of Laptop Battery Powered Portables

	TOTAL	Users	Non-Users
TOTAL	99	72	27
None	65%	62%	74%
One	16%	18%	11%
2 - 3	10%	13%	4%
4 - 5	2%	3%	-
6 - 9	1%	1%	-
10 - 25	3%	1%	7%
26 and over	1%	1%	-
Don't Know	2%	1%	4%
Mean	1.99	2.23	1.35

Over one third of the attendees' firms now have one or more laptop computers. Although there appears to be an increased usage of laptops, the number in each firm averages only about 2.

Number of Mini / Main Frames

	TOTAL	Users	Non-Users
TOTAL	99	72	27
None	53%	53%	52%
One	22%	22%	22%
2 - 3	10%	10%	11%
4 - 5	6%	7%	4%
6 - 9	2%	1%	4%
10 - 25	4%	6%	-
26 and over	-	-	-
Don't Know	3%	1%	7%
Mean	1.36	1.54	0.88

Almost half of the attendees have a mini- or mainframe computer that is used for marketing research purposes, and on average about 32 terminals are attached to mini/mainframes. Users of computer interviewing have significantly more terminals attached to a mini/mainframe than do the non-users, as one might think.

Color monitors and monochrome monitors appear to have about equal usage in marketing research. (Complete tables are available from Consumer Pulse.) Firms that interview with Sawtooth Software and have data collection capabilities, however, do NOT have a higher proportion of color monitors. We were surprised by this, since color monitors can enhance instructions, highlight words that need emphasis, and take away the "monotony" that interviewers experience with monochrome screens. It appears most firms with monochrome monitors are conducting CRT telephone interviewing, and are using monochrome because it is less expensive.

The 5 1/4 inch disk drives dominate the marketing research industry with nearly 100% usage. Although there is a major push by manufacturers for use of the 3 1/2 inch disks, more than half of the offices have no 3 1/2 inch disks, and nearly 75% of offices have fewer than 20% of computers with 3 1/2 inch disks. It is my opinion that 5 1/2" disks will continue to dominate computerized interviewing for several more years, particularly among firms who interchange diskettes for data collection. There are too many machines using 5 1/4 inch diskettes for the entire industry to switch.

In the past three months, about a third of the attendees have rented (not leased) PCs for marketing research purposes, and the majority of the firms who rented did so for an average of 47 rental days in the past three months.

Nearly all the firms that completed our interview say they OWN their computers, with less than 10% of them leasing any PCs.

Does your firm own or lease any software for computer interviewing?

	TOTAL	Users	Non-Users
TOTAL	99	72	27
Yes	69%	94%	-
No, but others send disks	4%	6%	-
No, and do not int. or set up int.	27%	-	100%

Over two thirds own software for computerized interviewing, with an additional 4% using the software by having firms send "field" disks to complete computerized interviews. About one-fourth of the attendees do not own any software for computerized interviewing. As mentioned before, we will examine the results among two groups -- those who use software for computerized interviewing and those who do not own any software for computer interviewing.

We then asked "Which of the following services does your firm normally complete on a market research survey?". Only the most frequently mentioned services are shown.

	TOTAL	Users	Non-Users
TOTAL	99	72	27
Prepare quest. on Word Processor	77%	81%	67%
Collapse data and prepare for process	68%	85%	22%
Tabulate data from computer interview	67%	81%	30%
Design of computer int for program/setup	66%	83%	19%
Program/setup of computer interview	63%	83%	7%
Code open ends from computer interview	52%	68%	7%
Int/DC using computers by TELEPHONE	47%	60%	15%
Trouble shoot software	46%	57%	19%
Provide verbatims from computer interview	45%	61%	4%
Troubleshoot computer equipment	45%	56%	19%
Troubleshoot programmed interview	45%	61%	4%
Daily monitor progress of computer interview	42%	56%	7%
Keypunch data from paper quest/comp. interview	40%	49%	19%
Int/DC using computers in FIELD	31%	39%	11%
Int/DC using computers in MALLS	26%	36%	-
None of these	6%	1%	19%

About two-thirds of the users provide verbatim open-ends from the computer interview, and about half as keypunch data from paper questionnaires that were part of the study.

We see that 26% of the firms have used computerized interviewing for mall interviewing, 31% have used it for field or pre-recruited interviewing, and 47% have used computerized interviewing for telephone interviewing.

So how are CRT studies completed? Are they completed internally? Are they sub-contracted? We also examined those items. Around 30% of CRT mall or field studies are completed internally among users of computer interviewing software and about 33% of the users subcontract mall and field studies, while 36% of the users don't do any mall or field-type studies. The non-users, however, perceive that more mall and field work is subcontracted more than internally completed.

On the other hand, for CRT telephone studies more than half of the users of computer interviewing software complete interviews internally, while only 27% of those respondents sub-contract work. In addition, there were 21% of the users who don't do CRT telephone surveys.

We were curious about the use of brand name diskettes as opposed to generic diskettes. Among the users, 70% of the diskettes used are brand name diskettes while only 30% are generic. What is curious about the use of generic and brand name diskettes are the failure rates.

We asked respondents what their failure rate was for per 100 diskettes and we found that among the brand name diskettes about 1 diskette per 100 fails, while with the generic diskettes, about 2 per 100 diskettes fail. This certainly appears to be a very acceptable failure rate for generic diskettes considering the tremendous cost savings associated with them.

The tabbing of computer interviewing is one concern most people have when they're trying to decide what computer interviewing software to use. We asked users of computer interviewing software what tab packages they most frequently use.

There is a wide variety of cross-tabulation packages that people use when using computerized interviewing software. Most are PC-based.

When we asked people how they'd rate the interface between computer interviews and their tab programs, more than 50% of the respondents indicated that it was excellent or very good, and 20% indicated it was fair or good.

We asked users if they had had any data processing problems with computerized interviewing data, Eleven percent indicated they'd had a major problem and 37% indicated they'd had a minor problem. More than half of the respondents, however, indicated they'd had no problem with data whatsoever.

When people described what the problems were and what could be done about them, the most prevalent problems related to user error, rather than problems inherent in the software.

We then asked if their tab programs imported stubs and formulas from the questionnaire section of the program. Thirty-two percent indicated that their current tab package imports stubs and formulas automatically.

Open-ends tend to be an issue with computerized interviewing. We asked the users what percent of the time their computer interviews had open-ended questions.

We found that more than 70% of computer interviews contain open-ends; only 6% of the users never use open-ended questions in their computer interviews. We found that a little more than a third of computer interviewing studies have open-ends written out on paper.

The coding of open-ended responses is done in a little more than half of the computer interviews, while over a third provide verbatim open-ends to the client.

We then asked where requests for computer interviewing are originating. More than 80% of users answer both "within our organization" and "from our clients." Interestingly, only about 70% of non-users answer "within our organization," and about 20% answer "from our clients."

Reasons for not completing more studies using computer interviewing

	Users
TOTAL	72
No need / not part of company	35%
Other	26%
None	18%
Inexperienced in selling computer interviewing	15%

The users' reasons for not completing more studies using computerized interviewing were that there was no need for it, or there was "no reason" for not completing more. Fifteen percent, however, indicated that they were inexperienced in selling computerized interviewing.

Other reasons for NOT completing computerized interviewing were: lack of demand/awareness/fear; do all of it already; don't have/unfamiliar with software; induced bias with lack of access to computers; and limited field capabilities/lack expertise.

Is there a surcharge for computerized interviewing? We asked respondents to compare the same study on computer interviewing and on paper. Fifty-three percent of the users indicated that there should be no difference or no additional charge for computer interviewing.

A quarter of the respondents, however, indicated there would be a surcharge and, on the average, it's about $9.00 more per interview compared to paper. These results are surprising since many of us have concluded that computer interviewing generally does not cost much more than paper interviewing.

The shipping of data to and from field sites concerns many researchers who do computerized interviewing. The most frequently used method for transmitting data is overnight delivery services, such as Federal Express. Nearly 40% of users do not send or transmit the data files; they complete the studies internally within their organization. There are very few people who regularly send data files to and from field sites via modem.

The copying of data at field sites as well as within organizations has been a controversial item for several years. Half of the users keep a copy of the data at the field site and 28 % send disks without backing-up. We also asked this question of users who regularly send the data using an overnight delivery service. Equal numbers back up and don't back up data before shipping the disks. The reasons given for back up are safety, security, and fear of failure of delivery of the data.

On the other hand, there are good reasons for not backing-up the files at the field sites. The given reasons include: "haven't had a problem," or too much risk in copying, or a lack of trust in the ability of the field sites to make back-ups.

Many times we feel that the interviewing staff will have problems converting to computerized interviewing. Nearly half of the users indicated that there was no problem in converting the interviewing staff from paper to computer interviewing. If there was any problem, it was the resistance among interviewers to learn computers. And, there are some problems in typing the open-ended responses. Other difficulties mentioned were: training; data entry errors; familiarity with keyboard; documentation/instructions; fear of lost wages; productivity; terminating interviews; and changing respondents' answers.

Research firms need data collection sites across the country for computerized interviewing. Today, over a quarter of the users think that there are more than adequate data collection sites across the country for completing phone work. Over 50% of the users don't know if there is an adequate number.

When we asked the same question for mall and field work, again, almost half of the users indicate they don't know if there is an adequate number of sites. However, only 15% of the users indicate there is a more than adequate number of data collection sites.

The quality of data is extremely important to most researchers. Eighty-one percent of users think their data has improved as a result of computerized interviewing. The reasons given: There are reduced errors, data are more accurate, it's more efficient, and, skip patterns are followed. Other reasons for the data being better are: validity of interview/interviewer bias removed/uniformity; more powerful analytical tools; and greater respondent interest.

We then asked the users and non-users if they thought that the turnaround time of the completion of tabs was faster, about the same, or slower compared to paper. Among the users, nearly 2/3 of the respondents indicated that the total turnaround time was faster. Among the non-users, however, nearly 60% of the people don't know; 30% perceived that it is faster.

What can be done to speed up the turn-around time using computerized interviewing? Most users think that more time in planning and programming, and better tab packages would improve the turnaround time. Other things mentioned were: open-end coding, interviewers who type and are better trained, fewer instructions, and a bank/library of questions.

How much of a competitive advantage can computerized interviewing provide for various types of studies? We found the best competitive advantage computerized interviewing has to offer is in the area of conjoint analysis, with 73% of the respondents indicating it provided a competitive advantage to complete conjoint studies on computer. Next, three categories: WATS, perceptual mapping, and local phone studies were viewed as having major competitive advantages when completed using computerized interviewing. Next, field studies, all intercept interviews, and medical interviews were perceived as having a major advantages when completed on computer. Finally, the advantage for mail interviews was perceived by 25% of all respondents.

About 32% of users complete CRT mall or field studies internally, about 33% of the users sub-contract mall and field studies, while 36% of the users don't do any mall or field-type studies.

On the other hand, for CRT telephone studies among users of computerized software over half of the users complete interviews internally, while only 28% of them sub-contract work. In addition, there were about 21% of the users who don't do CRT telephone surveys.

So what might be the future of computer interviewing? Over 40% overall, users and non-users, indicate it's going to expand and there will be more computerized interviewing in the future.

The future looks bright for computerized interviewing. We think computerized interviewing has arrived and, by the end of 1988, about 40% of all firms associated with marketing research could be completing computer interviews.

[Note: Complete tables are available from Consumer Pulse, Inc., 725 South Adams Road, Birmingham, MI 48011]

DOOR-TO-DOOR INTERVIEWING WITH LAPTOP COMPUTERS - A YEAR LATER

Joel Gottfried
Beth Rothschild
National Analysts Division Booz·Allen & Hamilton, Inc.

BACKGROUND

With the introduction of cost effective mini-computers in the mid 1970's, Computer Assisted Telephone Interviewing (CATI) became a reality. Now a fairly common means of conducting telephone surveys, CATI has offered the research community the opportunity for better quality data in a more timely fashion. Until very recently the use of these sophisticated data collection techniques for surveys that require either in-home or in-office personal interviews has not been possible. The dramatic breakthroughs in laptop computer technology in the last two years, however, have eliminated the last technical barriers to making door-to-door Computer Assisted Personal Interviewing (CAPI) a reality.

Slightly over a year ago National Analysts began the field phase of an extremely large and complex CAPI study for the Human Nutrition Information Service of the United States Department of Agriculture. Known as the "1987 Nationwide Food Consumption Survey" (NFCS), the purpose of the study is to assess the nutritional status of the American public and to plot trends in food consumption patterns nationwide. With approximately 9,000 households scheduled to be interviewed in a 15 month time period by a field force of nearly 200 interviewers in 120 different sites nationwide, this survey, to the best of our knowledge, is by far the largest CAPI survey ever undertaken in the U.S.

Besides its size, the NFCS is distinguished by its complexity. Each interview consists of four components:

1) a brief in-person screening to establish eligibility and cooperation

2) an in-person computerized household food use interview

3) manually completed retrospective and prospective individual food intake records for all household members

4) in-person retrieval of individually completed food intake records.

The portion of the study that was computerized consists of basic household information, including information about each member of the household, as well as detailed data about every food item that was consumed in the household during the previous seven days. It was the need to improve the accuracy and reduce the time spent coding this food use section of the survey that was the driving force behind computerizing the interview.

In addition to the NFCS, National Analysts has also conducted over a dozen other CAPI studies in the last year and a half. (See Exhibit 1 for a summary of the key features of these projects). While these other studies were much smaller than the NFCS study, each one had some "twist" that made its implementation as a CAPI study challenging.

Drawing on the collected wisdom of these experiences, particularly the NFCS, we will examine the key questions that must be answered to have a successful CAPI study :

- Will the interview situation significantly benefit from computerized administration to warrant investment in the hardware and software?
- What are the scheduling implications of computerization?
- What are the cost implications of computerization?
- What is the optimal interviewing software to use for a CAPI study?
- What is the optimal hardware to use for a CAPI study?
- How can the integrity of the interview be guaranteed?
- What are the best means of selecting, training, and computer outfitting a field force?
- How are hardware maintenance and distribution best handled?
- What is the optimal approach to on-site editing of completed interviews?
- What is the most efficient way to transmit completed interviews to the data reduction site?
- What changes in data reduction procedures are necessary to optimally handle computerized interviews?

As we answer these questions particular emphasis will be placed on the aspects of a CAPI study that are different from those faced in a CATI study. Whenever possible, specific examples drawn primarily from the NFCS will be provided.

KEY QUESTIONS

- Will the interview situation significantly benefit from computerized administration to warrant investment in the hardware and software?

  Similar to CATI, personal computerized questionnaires are particularly valuable when the skip patterns are complex and/or the tasks are not amenable to satisfactory paper/pencil application. Such issues as item or question notation, complex tradeoff tasks and interwoven household- and person-based questions are more readily accommodated via computerized administration. It has been our experience that this need for additional complexity has been the key factor in the decision to create a computerized questionnaire.

  While the tasks themselves may benefit from computerized manipulation, the sample size must be large enough to warrant the programming and equipment investments. In addition, sample size considerations should be examined in the context of repeated use. For example, if the initial effort is quite small but continuing waves are anticipated, the cost of computerization can be amortized across several iterations.

- What are the scheduling implications of computerization?

  While we have not found any significant differences in the total amount of time needed for a computerized study when compared to the paper/pencil approach there has been a dramatic shifting of the necessary time commitments. Due to the complexity of a personal interview, there is a tremendous schedule strain during the early phases of a study. Producing a thoroughly tested, well documented program in a timely fashion has proven to be a continuing challenge. The inevitable last minute suggestions from the client as well as our own continuous refinements of the questionnaire have severely complicated the programming effort. To guarantee a properly operating program, additional time must be allocated before the study is sent to the field. In particular, the time between the formal pretest of the interview (which is even more important with a computerized questionnaire) and the beginning of field is significantly longer than what is needed for a paper/pencil questionnaire. Our rule-of-thumb is to allow at least one week for the re-programming and re-testing that follows the pretest. Additional time must also be set aside for the careful preparation of the diskettes. While the diskette formatting and labeling can often be done ahead of time, the final version of the program is generally not ready until the last minute. We are currently evaluating automatic disk copying machines as a means of speeding up this procedure.

With well integrated software packages and carefully thought out procedures, most of the additional time required at the front end of a study can be reclaimed at the back end. Without any data entry and reduced data cleaning and coding, the data analysis phase can generally be expedited. It is important to note, however, that these gains are not automatic. It took us several studies to develop the necessary expertise to efficiently coordinate these activities.

o What are the cost implications of computerization?

Given the many ways that a CAPI study differs from a traditional paper/pencil survey, it is not surprising that the nature of the associated costs is also quite different. Some aspects of a CAPI study generate additional expenses while others afford the opportunity for significant savings.

The most obvious additional expense is, of course, the cost of the laptop computers. If a steady stream of CAPI studies can be anticipated, purchase of the laptops can be justified. Given the newness of the methodology we have not felt comfortable yet in making such a projection. With the exception of the NFCS, where the size and duration of the study made it economically feasible for the client to purchase the computers, we have rented laptops for each of our other CAPI studies. Besides the direct rental costs, additional expenses for shipping (often several times for one computer) and insuring the computers as well as the costs of diskette purchase and preparation and program development time must also be expected.

The elimination of any data entry effort and the significant reduction in coding and data cleaning time can amount to significant cost savings to offset the additional expenses. While it is not possible to say in general whether a study will be more or less costly when it is computerized, our experience with a dozen CAPI studies indicates that they frequently are cost competitive with the traditional paper/pencil approach. Unless they are carefully planned and managed however, this cost competitiveness easily can be lost.

o What is the optimal interviewing software to use for a CAPI study?

Similar to CATI surveys, a decision on software must be made in the context of the questionnaire and the available lead time. Such considerations will dictate whether a fully customized program is necessary or an off-the-shelf product can be used as is or with modifications. In all instances, we have found that existing software, albeit with modifications, has accommodated our CAPI needs. We have acquired a source code license for the

Ci2 (Computer Interactive Interviewing) interviewing software and have used a customized version of it for all our computerized studies.

Unlike CATI, where centralized supervision and monitoring can readily correct ad hoc problems on demand, CAPI software must be foolproof. Communicating long distance about interviewing problems with computer novices makes it especially difficult to determine whether a software problem is real or imagined. Knowing that diverse levels of interviewer computer literacy would abound, we added a number of important features to the Ci2 software to improve its generality and ease of use in complex personal interviewing situations. The most important of these are:

- Edit Mode
  This is a general non-destructive means to review and alter any previous answer.

- Sets of Repeating Questions
  Permits a series of questions to be asked as many times as needed for each interview. For example, a set of person-based questions for each member of a household.

- User Comments
  This is the equivalent of a wide margin in a paper questionnaire. It allows the user to comment on any unusual occurrences or add context to any answer.

- On Screen Grids
  Allows visual display of grids that matches the paper version and incorporates horizontal and vertical skipping and error checking.

- Special "Don't Know" and "Refused" Keys
  These keys permit but do not encourage "Don't Know" and "Refused" responses.

- Dynamic Range Checking
  The range of valid answers for any questions can be based on the answers from any number of previous questions.

In addition to these generic changes, many other modifications were made to the Ci2 software specifically for the NFCS. These changes, which included informational summary screens for the interviewer and inter-module data passing as well as customized data retrieval capabilities, were all essential for the timely administration of the interview.

In general, the ability to add to and/or modify the features of the interviewing software has proven to be a critical factor in assuring successful CAPI studies. If the resources (skilled programmers and access to the source code) are not available, extra care must be exercised when initially deciding on the appropriateness of the CAPI methodology.

o What is the optimal hardware to use for a CAPI study?

With a wide array of features to be evaluated and numerous tradeoffs to be made , selecting the "optimal" laptop computer for personal interviewing is a very difficult decision. Each computerized study has its own peculiar characteristics that force a re-evaluation of the appropriate tradeoffs. In addition, the rapidly changing state of computer technology guarantees that today's optimal laptop will be tomorrow's antique. Despite these difficulties, there are a number of criteria that must be included in any laptop selection process. Exhibit 2 is a list of the criteria and the minimum standards we used in evaluating laptops for the NFCS. Due to the length of the NFCS interview (an average of two hours per household, with a range from one to five hours), and the presence of a supplemental low income sample (from which we could not be guaranteed access to household electricity), we gave the highest priority to the laptops that had the longest battery life. Another very important feature is the laptop's weight. Our experience suggests that any machines over 12 pounds (including the battery) are not appropriate in door-to-door situations.

Our systematic review of the available technology in mid 1986 led us to select the Toshiba T1100 PLUS as the optimal laptop for NFCS 1987. Interestingly, we have continued to use the T1100 PLUS for all our other CAPI studies even though it would not always have been the optimal machine. Several of our CAPI studies have been self-administered executive interviews conducted in business offices. In these cases screen legibility was of greater concern than battery life. We felt strongly however, having worked out all of the associated compatibility and serviceability issues for the T1100 PLUS for the NFCS, that it gave us the best known and most reliable hardware environment for our other studies.

o How can the integrity of the interview be guaranteed?

In this context, we use integrity to mean those critical factors contributing to the reliability and validity of the personal interview situation. While these issues are certainly of concern in CATI situations, the decentralized, unsupervised CAPI circumstances make them that much more critical. First and foremost, extreme precautions (and testing and retesting) must be employed to guard against added respondent burden and hardware/software failure. Second, protocols must be put into place to ensure accurate entry of respondent-reported information. Although paper/pencil administration can result in transposed figures or inappropriately circled responses, it is much more likely that keying errors and reporting inaccuracies will occur in the personal computerized setting.

Fully cognizant of these issues, we make sure that we pay careful attention to the factors of a CAPI study that will ensure both reliability and validity. The NFCS is a good example. Because the interview was long to begin with, we would not tolerate an increase in interview time to accommodate computerized administration. Therefore, as modules of the questionnaire were developed, we undertook multiple rounds of pretest interviews comparing the paper/pencil and computer-aided approaches. We identified areas where computer administration was shorter and longer than its paper/pencil counterpart. We worked on customized data retrieval functions to compress the time. We elected to display previous responses at critical junctures to aid the flow and length of the interview. In fact, computer administration averages 128 minutes compared to approximately 120 minutes for the paper/pencil version.

In addition to time, we needed to be certain that the hardware and software would endure even under the most grueling situations. It was not as if interviewers could switch to an alternate computer interviewing station. Backup machines were available at the home office -- ready for 24-hour turnaround distribution if necessary. To this end, upon acquisition, we tested every machine for durability -- battery capacity, screen clarity and the like. Even the slightest irregularity was reported and either replacement machines were received or servicing was performed.

No matter how much effort is placed upon ensuring computer durability or respondent receptivity, some interviews will likely be "lost" due to system failures. Depending upon the importance and expense of each interview and the difficulty in rescheduling, it might be desirable to prepare paper/pencil documents to use as a backup to computer administration. This is exactly the approach we took for the NFCS. The use of paper/pencil documents has been quite rare (fewer than 2% of the NFCS interviews completed have used this approach). The majority of these have not been associated with computer breakdowns, but rather with the unavailability of equipment and the recalcitrant mind-set of the interviewer and/or respondent. For all of our other CAPI studies we have elected to rely exclusively on computer administration. In each of these studies we were using a data gathering technique, such as Adapative Conjoint Analysis, that only can be conducted on a computer.

While the reliability issue can be satisfactorily addressed, ensuring validity is more problematic. Since we cannot look over the shoulder of every interviewer to assure ourselves that data entry is correct, we incorporate some special measures into every CAPI study that are designed to verify the accuracy of the information reported. For the NFCS we used four techniques:

- Range checks for all numeric responses.

- Summary screens at key points in the interview. For example, since so many other parts of the questionnaire were contingent upon correctly identifying the household members, a household composition screen was displayed and either accepted or corrected early in the interview.

- Provisions for dynamic range checking were incorporated into the software; that is, answers to one or several previous questions were assessed to determine the acceptability of the answer in question.

- Interviewers were periodically queried about editing and given ample opportunities to check their entries before moving on to the next question.

- Through intensive training, interviewers were taught how to enter a response, check it, and ultimately record it.

o What are the best means of selecting, training and computer outfitting a field force?

Assembling a field force capable of administering in-home personal computerized interviews requires careful attention to the schedule of interviews, number of computers available, and the varied computer/interviewing skills of the individuals involved.

In terms of selecting staff, it may appear at first blush that familiarity and experience with computers would be of paramount importance. We have found, however, that the propensity for door-to-door interviewing is still the most vital selection component. Because in-home interviewing depends heavily upon making the right first impression, gaining cooperation, and developing rapport, good interpersonal skills are essential. While computer knowledge is a plus, hands-on questionnaire training can rectify most of these shortcomings.

Since we believe basic interviewing skills are fundamental, training for in-home personal computer-aided interviews must, by necessity, focus on the computer itself, using the computer to guide and record the interview, and the specifics of the questionnaire.

For obvious reasons, personal training is optimal. Not only does the opportunity for individualized coaching present itself, but it affords you with the ability to gauge the integrity and responsibility of the interviewer. After all, it is not as if you are entrusting the interviewer with easily replaceable, inexpensive paper documents, but rather with a PC valued at $1,200 to $1,800. Not only do you need assurances that the equipment will be properly used, but that all reasonable efforts will be undertaken to ensure its safekeeping and return. For the NFCS, we were in the position of being able to conduct personal training sessions. On other studies, we have used preprogrammed training disks successfully.

Because the supply of computers is not unlimited, continual evaluations of potential staff must be made, both in terms of productivity and competence, so that available computers are being used to their full capacity. Additionally, if you are in the position of working with independent interviewers rather than employees, provisions must be made for the protection and safe return of all equipment.

Because interviewers are not generally very computer literate we have had to develop a complete set of CAPI training materials. In addition to the customary detailed instruction manual (with special computer section) used for every study, interviewers might receive:

- Advance study manual:
  This manual familiarizes interviewers with the required computer terminology, basic computer operations and the scope of the interview. It is meant to alleviate possible sources of computer phobia by giving ample opportunity to preview features of the computer.

- Elementary training exercises:
  To give interviewers a quick sense of accomplishment, subsets of the questionnaire are extracted for demonstration of the key features of the computerized interview. These exercises allow trainers to focus on proper reading and entering conventions before more arduous tasks are presented.

- Individualized practice interviews:
  These include complete hard copy information about hypothetical respondents that allow the interviewer to conduct a mock interview using the full working version of the software. The practice interviews serve to reinforce the formal group training sessions and allow more and less proficient interviewers to move at their own pace.

To ensure computer safety, we require each interviewer to sign a stringently worded agreement acknowledging receipt and use of the computer for none other than its express purpose. The penalties for mishandling or failure to return the equipment are presented very clearly. We maintain records by serial number and to date have not encountered any computer losses, although on occasion considerable efforts have been expended to retrieve the equipment. We also remove all documentation about the computer and the operating system from the original boxes before sending the equipment to the interviewer. This discourages use of the computer for other purposes.

o How is hardware maintenance and distribution best handled?

Implementation of in-home personal computer-aided interviewing creates an array of field problems not common to paper/pencil administration or CATI situations. Most obvious among these are the need to maintain and distribute (and in many cases re-distribute) expensive laptop computers.

Because serviceability is highly varied from point to point, we have found it best to secure a centralized servicing agreement where repair time can be guaranteed. The key features of this agreement are a 24-hour turnaround on broken equipment and a dedicated person at the service outlet through whom all servicing problems can be funneled.

Interviewers are given instructions to use expedited delivery services for return of faulty machines. Because return shipment is required at the end of a study anyway, interviewers are already supplied with the original packaging materials. In a year of use by nearly 200 NFCS interviewers, only six hardware failures have been encountered. Replacement machines were distributed as necessary and the failed machines serviced without problems or delay.

Optimal distribution of computers is a more difficult issue to deal with. Besides being sure that a widely dispersed staff of individual interviewers all receive properly functioning equipment in a timely fashion you also have to worry about the continued re-distribution of the equipment. As with any large-scale, extended data collection effort, a CAPI study will likely fall prey to interviewer illness, malaise, and poor performance that will dictate reassignment of work. Unlike the paper/pencil methodology, this not only requires identification of new staff, but also redistribution of the computer hardware.

For reassignments, we have used two approaches to ensure prompt, accurate equipment relocation. Personal pickup by the newly assigned interviewer has been used on occasion if the distance between interviewers was not significant and the equipment had not been used. Expedited delivery of equipment to our central office or to specially designated on-site field personnel has been used most often because it enables us to certify that all equipment has been returned in good working order ready for the next interviewer's use.

While we have had some success at sharing computers among geographically clustered interviewers, it is difficult to coordinate and time consuming to properly manage. From the interviewer's perspective, equipment sharing places added burden on hectic schedules and impedes productivity.

- What is the optimal approach to on-site editing of completed interviews?

The needs for on-site edits change when moving from paper/pencil documents to personal computerized interviews. While it is customary and often positioned as essential for field interviewers to check their work for missed skip patterns, illogical responses and the like before sending it to the home office, such checks are no longer necessary when computer-aided interviewing is undertaken. Avoidance of such problems is built into the questionnaire software.

We believe that the interviewer needs the ability to review and revise responses while the interviewing is ongoing, knowing full well how often respondents change their minds or remember items at a later point in the interview. Once the interview is complete, however, safeguards must be in place to preserve the interview, rather than change or destroy it. As such, we generally prepare the software to limit the interviewer to using a diskette only once. Prior to termination, the interviewer is given ample chance to inspect his/her work.

Such an approach greatly diminishes the possibility of faulty file manipulation and destruction of the interview as is common when curiosity prevails. However, not all situations can be anticipated. One diligent interviewer, for example, wanted to be sure all of her diskettes worked when she received them. She therefore inserted each and every diskette into the laptop and tested them. The result -- she had no usable diskettes available for her interviews, and we learned the hard way to be more explicit in our instructions.

- What is the most efficient way to transmit completed interviews to the data reduction site?

Initially, a decision must be reached regarding telephonic communications versus expedited mail transmission. Once the preferred option is selected, safeguards must be instituted to preserve data accuracy and delivery speed.

We have carefully examined both alternatives and adopted the mail or express delivery channel for reasons of technology and quality control on all our CAPI studies. We have no reason to believe, even two years later, that this decision does not continue to be the optimal approach.

Our nation's public network of low-speed (less than 9.2 Kbps) and analog lines are not suited for high-speed data communications (nor was it cost-effective to consider a private data network). More importantly, data communications from many of our 120 interviewing locations nationwide, especially those served by non-Bell operating companies, would be subject to significant transmission error if we relied on regular telephone service to send questionnaire information to our Philadelphia office. In addition, separating transmission of the computer-aided interview from other paper/pencil documents would have diminished rather than improved control of the survey data, particularly since links and cross-evaluation of all survey information together are usually imperative.

Having made the decision to use customary delivery channels (e.g., USPS, United Parcel Service, expedited carriers), we had to decide whether or not we needed a backup copy of each completed interview to guard against loss or damage during shipment. While we have experienced an occasional shipment problem, the jury is still out on this matter. For the NFCS we designed the interviewing software so that every interview is not only entered onto its own diskette, but simultaneously recorded onto a backup diskette. For our other CAPI studies, where the length and expense of each interview was not as great as in the NFCS, we have elected to ship without a backup.

- What changes in data reduction procedures are necessary to optimally handle computerized interviews?

As with all phases of a CAPI study, the data reduction phase requires careful planning to run smoothly. Unless a well integrated set of procedures and software products is used, many of the potentially significant advantages of computerizing the study may not be realized.

For the NFCS we have developed a series of check-in, editing, coding, and cleaning programs that are tailored to its particular needs. By using this suite of programs we have been able to achieve considerable savings in time compared to the handling of paper documents. In particular, we have been able to save large amounts of time previously spent coding open-ended food consumption questions. For example, we were able to change an open-ended question such as, "In what size container was that food purchased?," into a closed-ended question through the use of advanced database access techniques that provide the interviewer a list of all the common units for the food being described. Consider the magnitude of this savings. With each of 9,600 households reporting an average of 45 different food items, there are hundreds of thousands fewer open-ended questions to be coded.

For our other CAPI studies we have developed a general purpose "check-in" program. Besides allowing us to monitor a study's progress in the field and to ensure sampling integrity, this program also provides a comprehensive means of checking the quality of the interview. Unlike a paper questionnaire, where a quick visual check of the document is often sufficient to provide a basic understanding of the interview's integrity, a program such as this is necessary for even the simplest edit of the interview. Since inexperienced interviewers might easily return incorrect, incomplete, or in some cases non-existent interviews on the diskettes, this program is mandatory. Despite the additional time and expense associated with developing and testing it, we have found that through its use we have been able to significantly enhance our initial editing capabilities. Instead of just checking if all sections of the interview have been completed, as might normally be done with a printed questionnaire, this program conducts fairly sophisticated checks of the interview data and reports inconsistencies and summary statistics to the editor. The program even flags which completed interviews for each interviewer must be validated and prints the validation form with the validations responses already filled in. Since all of these analyses are conducted immediately upon receipt of the diskette, problems that typically are not discovered until much later in the data analysis phase can often be immediately resolved by calling the interviewer while the interview is still fresh in his/her mind.

CONCLUSIONS

With the technology to conduct high quality computer-assisted personal interviewing already here, and with advances in hardware and software systems being announced every day, CAPI is a research methodology whose time has clearly come. As with any technological breakthrough however, carefully thought out procedures and policies to accompany it must be developed, tested, and refined to realize the true potential of the technology. Successful use of these procedures can come only with experience. If our initial experiences are at all indicative of what lies ahead, CAPI surveys will be able to significantly extend the range of complexity that can be successfully accommodated in a personal interview. While it is too early to tell whether comparable improvements in study cost and time will also be achieved, it is very likely that less expensive hardware, more capable software, and better tested procedures will yield significant improvements.

Exhibit 1: National Analysts' CAPI Studies

Client	Subject	Sample Size	Type of Respondent	Type/Number of Interviewing Locations
USDA	Food Consumption Patterns	9000+	Household Meal Planner/Preparers -- General Public and Low-Income Supplement	In-home (n = 120)
USDA	Food Consumption Patterns	75+	Household Meal Planner/Preparers	In-home (n = 1)
Hotel Chain	Weekend Leisure Lodging	500	Frequent Travellers	Central locations (n = 5)
Hotel Chain	Hotel Preferences	600	Business Travellers	Central locations (n = 16)
National Analysts*	Computer vs. Paper-Pencil Conjoint Methods — Banking Preferences	175+	Commercial Banking Customers	In-office/central locations (n = 2)
Telecommunications Firm	Telecommunications Needs of Large Business Customers	225	Telecommunications Managers	In-office/in-home (n = 20)
Telecommunications Firm	Future Telecommunications Needs	800	Telecommunications Managers	In-office (n = 25)
EPRI	Energy Usage and Management	150+	Residential Energy Users	Central locations (n = 4)

---

*"Computerized vs. Paper and Pencil Conjoint Methods: A Comparison Study" presented at The Association for Consumer Research Annual Conference, October 17, 1986 in Toronto, Canada, by Carl T. Finkbeiner.

Exhibit 1: National Analysts' CAPI Studies
(Continued)

Client	Subject	Sample Size	Type of Respondent	Type/Number of Interviewing Locations
Conglomerate	Elderly/Retirement Housing	300	Elderly Household Heads	Central locations (n = 4)
Health Insurance Carrier	New Health Insurance Products	500	Employees	In-office (n = 3)
Telecommunications Firm	Hotels' Use of Telecommunications Services	500	Telecommunications Managers	In-office (n = 80)
Appliance Manufacturer	Appliance Product Design	600	Purchase Decision-Makers	Central locations (n = 8)
Telecommunications Firm	Telecommunications System Design and Development	100	Telecommunications Managers	In-office (n = 20)

Exhibit 2
Laptop Computer Evaluation Criteria

FEATURES	REQUIREMENTS
Memory	At least 512K
Disk type	3.5" preferred
Number of disks	1 is sufficient
Processor	Equivalent to Intel 8088
Screen size and legibility	25 rows by 80 columns
Keyboard	Full size
Dimensions	
Weight	Less than 12 pounds
Power source	Must be battery
Power duration	At least 5 hours
Availability	Must be an existing product
Service-ability	Readily available
Carry-ability	Comfortable shape and handle
Ship-ability	Sturdy and reusable box
Reliability	
Durability	Able to withstand repeated use
Ci2" compatability	Essential

PC-BASED RESEARCH : EUROPE VERSUS THE U.S.A.

Dirk Huisman
SKIM Market and Policy Research

INTRODUCTION

Comparing past, present, and future of PC-based research in the USA and in Europe means that we have to compare:

- the environment for PC-based research and the structure and development of the market for market research;
- the process and function of research;
- the function of the PC in research and marketing in the USA and in Europe.

First I would like to point out that the continents to be compared are not independent. At all levels, economical, technological, methodological and philosophical, there's a great deal of interaction. Major clients of market-research, such as, in the USA, Procter & Gamble, Xerox, Coca-Cola and IBM, and, in Europe, Unilever, Olivetti, Philips and Shell Oil, are active on both continents. The introduction of a new computer, for instance, occurs simultaneously in the USA and in Europe. Leading professional magazines consulted by European managers and researchers in order to read up on new developments are, e.g., Harvard Business Review, the Journal of Marketing and the Journal of Marketing Research.

Although resemblances exceed divergences, we are still left with some differences that have to be accounted for. It is through the differences that we can learn from each other.

ENVIRONMENT, STRUCTURE AND DEVELOPMENT OF THE MARKET FOR MARKET RESEARCH

Europe and the USA are comparable entities as far as population and the market for market research are concerned:

	Europe	USA
population x 1 million	379	225
expenditure market-research x 1 million US$ (1986)	1.350	1.800

In 1987 the divergence in expenditure has decreased owing to the drop of the dollar rate. Changes in the exchange rates and inflation disparities make any further comparison of expenditures a risky business. Besides, an equal level of expenditure does not guarantee that the efforts are on the same level. Prices differ too much to warrant such a comparison. A quadrennial price comparison commissioned by ESOMAR (European Society of Opinion and Marketing Research) shows that an identical attitude survey (face-to-face, N=1000, T=45) is three times as expensive in Switzerland as in Greece. This comparison also illustrates that the differences within Europe are often much greater than the differences between Europe and the USA.

"Communication" is a keyword for market researchers. Communication implies that those who are communicating speak the same language. Americans have a common language. Europe has to cope with sixteen different languages; speaking five of them will enable you to get by reasonably well. Besides this babel we have to cope with borders, maybe only up till 1993, which is the target date for a United Europe. You may realize what this implies if you try to imagine a truck driving from New York to Los Angeles at an average of 45 miles an hour. In Europe the same truck driving from Copenhagen to Rome will make an average of 15 miles an hour, being delayed time and again at some border. So for the time being Europe, as compared to the USA, is a heterogeneous collection of national markets, each with a culture, a language, and a legislation of its own.

What are the implications of all this for marketing and for marketing research? In order to accomplish growth the home market will soon be too small and one will have to turn to other countries, where a different language is spoken and where other customs prevail. The marketer will have to adapt; he is forced to familiarize himself with other countries and other cultures. The USA too is a melting pot of nationalities, possibly even more so than Europe. There is, however, an important difference: in the USA this motley collection has been molten down to one nation with some dominant norms and one common language. The USA as a whole is the home market. To the European researcher, operating in a heterogeneous collection of national markets implies that one has to be much more sensitive both to differences in the market situation and to cultural differences. Simultaneously, the researcher and the research company must cooperate more closely with researchers in other countries (questionnaires have to be adjusted to local situations and usage).

The following scheme illustrates the importance of cross-national research. The number of foreign companies that commissioned research in the countries cited exceeds the number of domestic companies associated with ESOMAR.

	number of firms carrying out or commissioning research in the country cited	number of firms in the country cited associated with ESOMAR	population x 1million	expenditure on market-research x 1million
Germany	283	96	61	325
France	262	74	55	215
UK	223	110	57	345
Italy	183	57	57	150
Netherlands	163	45	14	82
USA	176	-	225	1800

(fig.1: the importance of cross-national research)

Given this high intensity of cross-national research it is easy to understand that Europe faces an increasing integration and linking-up of research companies. For the time being, however, an integral European market for market research is out of the question. Integration and linking-up entail complex networks, ranging from fully-owned agencies and partly-owned agencies to agencies working together on a voluntary basis and agencies co-operating only in certain defined projects, all scattered in various European countries. Obscurity is enhanced by the fact that companies linked up with one chain also work with companies not linked up with the chain, if the client wishes.

Integration and linking-up will automatically lead up to a further concentration on the market's supply side. At the moment there is a certain amount of concentration within the respective European countries and there is a tendency towards an oligopolistical market. In this regard there is little difference with the USA. For Europe as a whole the degree of concentration will be somewhat lower than in the USA.

	75% of expenditure on market research realized by	
	number of firms	percentage for firms associated with ESOMAR/CASRO
		%
Germany	9	10
France	8	11
UK	10	10
Italy	5	08
Netherlands	6	15
USA	38	-

(fig.2: degree of concentration)

As far as the top of the market is concerned (where international integration is highest), differences between Europe and the USA are negligible. They are much more striking at the base. Fieldwork is farmed out to freelance interviewers who form part of the marketing companies' own interviewing corps. The importance of fieldwork agencies is minor. Only a few countries have fieldwork agencies operating on a national level. We do not know of any fieldwork agencies forming a network and covering the whole of Europe.

THE PRACTICE OF MARKET RESEARCH: PROCESS AND FUNCTION

Comparing the world's top ten research companies, we find that the USA-based firms are principally engaged in market measurement and database business, while the Europe-based firms are much more active in the ad hoc or custom area of research. In addition, the European companies dominate the area of international research (1). Attention to this difference between the European and the American companies was also drawn by Asselbergs (2):

"From the early beginning the major American companies have presented themselves as suppliers of information, setting up databases on a very broad scale; because of this little attention was paid to traditional research."

Another difference to which Asselbergs draws attention concerns the fact that in the USA commitment to research is on a much higher level in the organization. Besides, in the USA clients have always commanded large market research divisions, designing research and providing interpretation on their own. Things are different in Europe. Very often we find the market researcher on a lower level in the organization. The researcher's less elevated position is also mirrored in the major threats to marketing research of which Piercy (3) has taken stock (UK):

- low status

- irrational reasons for farming out research

- the researchers' lack of understanding of managerial problems

- want of a clear link-up of commercial success with marketing research.

Correlated with the above-mentioned facts we find a difference in background and training of the researcher. Comparative research by Baker (4) has shown that American qualitative researchers more often have a marketing background than do Europeans. It also showed that, in the UK, researchers on the supply side (research companies) had less experience, or no experience at all, as clients. If they had been employed previously, then they had been employed, in most cases, by another research supplier or at the university. Another difference, correlated with labor-mobility, is that American marketers are more mobile than European ones: they switch over to another job more easily. This greater mobility contributes, in the USA, to the routine production of a continuous stream of smooth data.

In Europe the purchase of standard data is not as high as in the USA. Market research and marketing companies have a different status there. Most research companies are keen on supplying "tailor-made" or custom research, featured by research design, analysis, and interpretation. A company's creativity and full-service quality is its *sine qua non*. Competition forces researchers to develop new methods and techniques and to put them into practice, which, of course, entails taking risks. Regarding qualitative research: Malcolm Baker (4) mentions a number of subtle but striking differences in design and implementation of research.

	USA	UK
setting	client oriented	respondent oriented
goal	goal oriented	understanding the consumer in general
researcher	a moderator	an expert/ a consultant
	a team	an individual
philosophy	marketing oriented	more philosophical and taking into account environmental influences
interpretation	less willing to go beyond the data	a stronger interpretative stand

The above-mentioned differences between the USA and Europe are mirrored in the criteria applied by clients when evaluating research companies. In the USA these are:

- time-planning and keeping to deadlines
- the supply of normative consistent data
- the meeting of all specifications (often put down in a formal contract)
- the supply of absolutely clean databases
- the supply of matrices that enable the projection of data into national sum totals
- budgeting within strict limits.

In Europe these criteria play an important role as well, even though they may be characterized as "conditions to be met" - research companies are assessed less stringently as far as these characteristics are concerned. Hansen has done research among managing-directors of major clients of market research (5). He wanted to know what was expected of first-class market research and came up with the following features:

- projection and identification, knowing the market
- elucidation of the results of research
- means to verify the results
- original points of view in the reports
- good recommendations, advice.

Research carried out in the Netherlands resulted in a similar list, adding "punctuality" and "reliability" as important features.

Interpreting these kinds of differences between Europe and the USA, I would like to point out that they are to be taken only as general trends because there's a vast heterogeneity within the populations. As far as Europe is concerned, I would like to recall that Europe is a heterogeneous collection of national markets. This historical situation has always forced the marketer and the researcher to empathize with the consumer (and in Europe more so than in the USA). Research has always been and still is tuned in to the local or national situation and is often tailor-made. Since many a company has failed by carrying out routinized standard research, a certain aversion to standard approaches has taken root.

A moot point might be if the American standard approach will dominate the market or if the European "tailor-made" approach will have its way. I have a simple suggestion: both are important, none of them will dominate. Baker (4) distinguished between two segments on the market for qualitative market research: "a commodity segment that provides little more than interviewing skills, and a value-added segment that, through the experience, insight and integrative skills of its practitioners, will continue to bring the consumer world alive in a manner that has utility and creates excitement." Simmons (6), researching developments in the UK market for market-research, also distinguished two segments based on the clients' requirements:

1. more attention to design, interpretation and recommendation;

2. more attention to the quality of field-work.

Developments (and possible developments) in market-research can be put into a product-process/product-range matrix. The services supplied by the market researcher vary from providing raw data to strategic advice. The production process is characterized by various phases, ranging from problem-tracing to implementation of the results. The matrix shows the stress on each service during the production process.

			product range	
production process	raw data	information	analyses/ models	strategic advice
1. problem-tracing			x	xxx
2. analysing problem/research design		x	xx	xx
3. data collecting/ data processing	xxx	xxx	xx	x
4. analysing/ interpretation		xx	xxx	xx
5. advice, implementing results		x	xx	xxx

(fig.4: production process - product range matrix)

© SKIM

For a correct interpretation of the matrix the "relative emphasis" is essential. For example, the quality of data and data-collecting are equally important for any product, considered in absolute terms. Anyone supplying strategic advice, however, will be very critically assessed as to his or her qualities in the field of problem-tracing and interpretation of results.

In the USA the emphasis has been especially on phase 3 of the production process. In Europe emphasis has been divided between phases 2 and 4. One of the weak points of the European suppliers used to be that they were often put to unnecessary great expense through orientation on phases 2 and 4, while offering products from categories I and II. They were expensive, while there was no need to be so.

Development of the market for market research clearly entails a tendency towards industrialization. Especially phase 3 lends itself to an industrial standardized approach. However, the standardized approach is not restricted to phase 3 but is gradually extending to phases 2 and 4. There is an increase in "standard and branded research" from the USA.

It is at this point that American and European developments collide. In Europe there is also a tendency towards industrialization. This tendency, however, greatly resembles the process of "flexible production automatization." An infrastructure is being built to allow for automatic production - the automated process is easily adapted to the client's "quandary" or to his requirements. Corresponding differences also turn up in marketing of products and services offered. The USA-based firms offer branded standard products (models). The Europe-based firms are "branding" themselves. In Europe there is not only a certain aversion regarding standard models, but the models themselves are often tripped up on theoretical grounds as well.

THE FUNCTION OF THE PC IN RESEARCH IN THE USA AND IN EUROPE

In the Introduction attention was drawn to the fact that the introduction of any new PC often happens simultaneously in the USA and in Europe. A PC-handicap simply does not exist, though it cannot be denied that the computer's penetration is more advanced in the USA than in Europe. The USA has .26 workstations per white-collar-worker, Europe has .19 workstations per white-collar-worker. Although it is very difficult to ascertain the penetration of MS-DOS machines or IBM-compatibles (given the multitude of brands, there are over 300), the penetration of IBM-compatibles can be put at 50%-60%. However, for the European market researchers who intend to carry out computer-aided research there is a major difference with the USA: the supply of computer rentals is heavily limited. Though it is possibile to rent a few PCs at a time, renting some dozens of PCs or Laptops is, as far as we are aware, out of the question.

We have the impression that PC-based market research has penetrated less in Europe than in the USA. It is evident, though, that there are firms carrying out PC-based research based in all major European markets. An enquiry conducted among a selective sample of market researchers (visitors of the SKIM-booth at the 1987 ESOMAR conference) shows that 62% of the researchers are conversant with one or more domestic institutions carrying out computer-aided face-to-face research. One out of three researchers has carried out such a research (or commissioned) such research. Finally, 81% expect an increase in the use of computers for face-to-face research.

Acquaintance with and use of Cati-systems is considerably higher. Eighty-one percent have personal experience with Cati, but there are big differences between the European countries. Leaders in the field of computer-aided interviewing are Sweden, the Netherlands, and Italy, followed at some distance by the United Kingdom.

We have described a number of differences in marketing research between the USA and Europe. Could there also be a difference in the use of PCs in market research? To answer this question we have to consider the use of software in the various phases of the production process. Here attention should be drawn to the fact that many research companies employ self-developed or adapted software, which implies that the picture can only be fragmentary. Besides, recall that Europe knows many different languages, which means that the software is not always fully accessible.

In the initial phases of the marketing process we have, on the one hand, the systems for artificial intelligence and decision-support helping the researcher to comprehend problems regarding policy and helping him to specify target data and decisions to be made. On the other hand we have interactive systems used in qualitative research, so that, in an interviewing situation for instance, the respondents can be confronted with possible reactions to their previous remarks. As a third alternative we have the Delphi-like systems that can be used while interviewing experts.

The collection and processing of data are being streamlined at the moment by the well-known Cati-systems that are now, thanks to the economical microcomputer, within reach of almost any company. Next to these are the computer-interactive interview systems, also known as "Capi" systems. Thanks to the computer, intricate routing and selective in-depth interviewing are now feasible. However, this is merely the tip of the potential iceberg. Potentials are especially significant in the analysis of responses given while interviewing, so that the results of the analysis can be integrated into the rest of the interview.

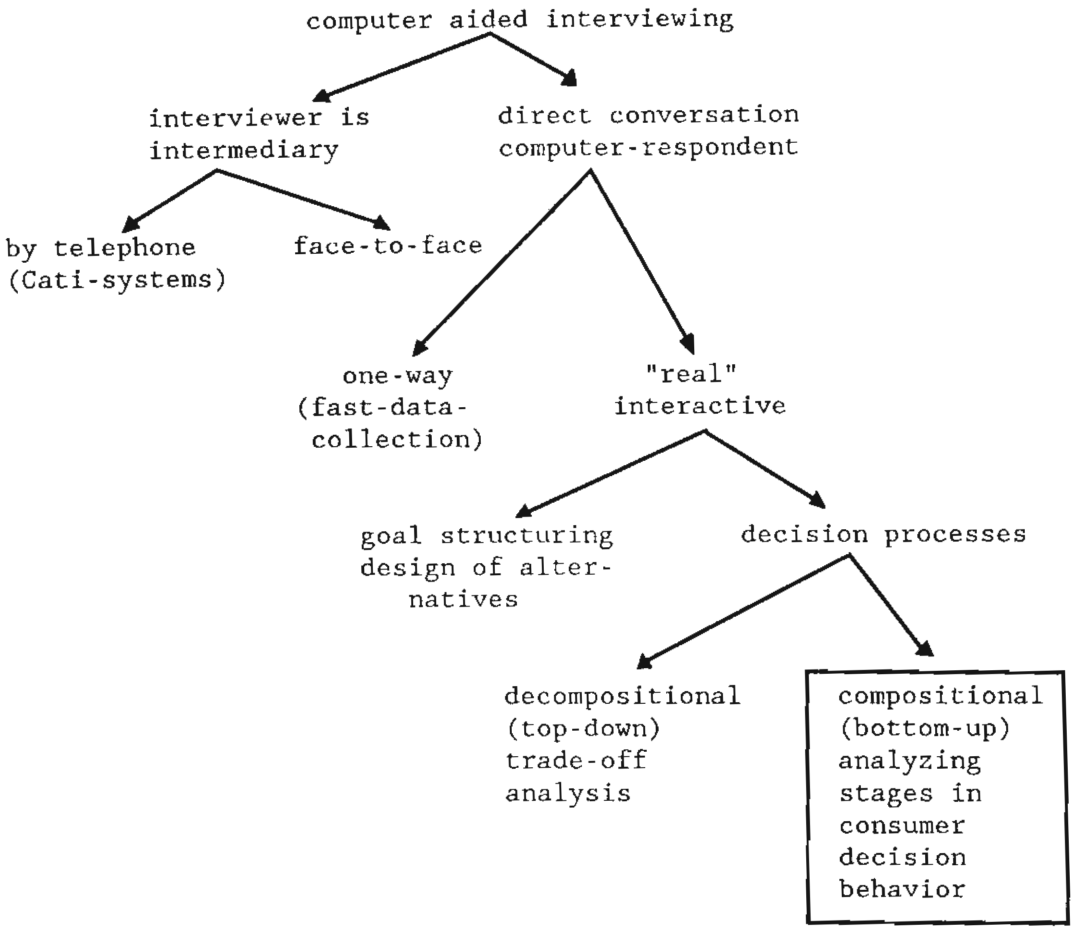
The different systems of processing (especially if these can be combined with the different enquiry-systems) enable the researcher to supply the data and the information requested instantly. Time is being gained for further analysis and various sources of error are eliminated.

Finally, several systems for analysis, interpretation and implementation have been marketed enabling the researcher to conduct very intricate and thorough analyses quickly and to test the material's reliability. Besides these, there are systems that, linked up with enquiry-systems, enable the researcher to do simulations that visualize the likely results of considered changes in policy.

The techniques used in the software are familiar to Americans - often these were developed in the USA. Researchers, especially those with a theoretical background, are often "nursed in the same cradle." However, according to Asselbergs (2) statistical techniques developed in the USA are employed much more intensely in Europe than in the USA. In order to follow up a tailor-made approach and in order to stand out as a company, command of the techniques is presupposed, isn't it? Now we are not talking only about well-known statistical systems such as Systat, SPSS, and SAS, but also about systems for correspondence analysis and conjoint analysis and about MDS systems. Unfortunately, these systems are not often characterized by traits such as "user-friendly" and "standardized." Simple user-friendly cross-tab systems are far and few between too. Software for analysis is mainly American.

Artificial Intelligence systems and decision support systems are not yet extensively used and they are not yet supplied on a large scale.

What kind of systems have been developed for computer interviewing in both qualitative and quantitative research? Besides employment of software developed in the USA we also have software developed specifically in Europe. Bronner and De Hoog (6) have schematized the different forms of computer aided interviewing as follows:



(fig.5: DIFFERENT FORMS OF COMPUTER AIDED INTERVIEWING)

Several Cati and Capi systems have been developed. Articles have appeared on some systems that may diverge from the American systems:

<u>1. Interviewer is intermediary, face-to-face</u>

The Swedish Statistical Office has created a system for statistical data collection with lap-top computers. The system was used in a consumer price index survey. The Dutch Statistical Office developed their "Blaise" system also for statistical data collecting. There were a number of reasons for developing the system, but from the publications concerning experiences with the systems it appears that validity, quality control, and built-in checking procedures were the main reasons.

These systems are not developed in isolation. The people working on them meet each other annually through the Study Group on Computers in Survey Analysis. The people involved in the Study Group are often employed by a Statistical Office or Census Office and have a background that implies much research into the quality, reliability, and potentialities of systems and methods recently developed.

2. Direct conversation computer-respondent, one-way

The University of Amsterdam in the Netherlands developed "Telepanel," now commercialized by NIPO. Telepanel can be compared to the mail surveys with diskettes popular in the USA. As the name indicates, Telepanel is a panel of 800 households. A questionnaire is regularly down loaded from the mainframe at the institute to the 800 home-computers stationed with the panel's members. They respond to the questionnaire at a time convenient to them and transfer the data by phone and modem to the institute. The design combines the opportunities of panel research and of tailor-made computer interviewing.

3. Interactive compositional analyzing stages in consumer decision behavior

The "Midas" system was designed by Bonner and De Hoog to cope with highly complex decision problems. The multi-attribute model decomposes decision makeing in a number of steps:

- identification of options
- identification of the relevant attributes for making a decision between the alternatives
- assigning values or positions to the alternatives or the attributes
- assigning the most preferred (ideal) value to each attribute
- differential weighing of the attributes
- presenting an advised preference order.

The main advantage of "Midas" is that the attributes are not fixed and do not have to be named beforehand by the researcher. In this way different respondents can use different attributes and can label the same attribute differently, consistent with their points of view. This system, which is used in qualitative studies, perfectly highlights one of the differences that Baker found between qualitative researchers in the USA and in Europe. In the USA qualitative research is rather "goal oriented" while in Europe "understanding the consumer" seems more important.

Having compared PC-based research in Europe and the USA, it will be clear that the similarities are more pronounced than the differences. The differences reflect the different approaches to market research. In Europe market researchers use the PC as a craftsman's tool that offers new opportunities. In the USA it is used rather as a sophisticated instrument in a standardized production process.

LITERATURE:

1. Cole, E., "Taking Stock: Our Progress and our Future," October 1985, European Research.
2. Asselbergs, W.J.M., "European and American Developments in Market Research," Rotterdam, 1987, Instituut voor Psychologisch Marktonderzoek (internal paper).
3. Piercy, N., "A Social Psychology of Marketing Information," April 1983, Journal of the Market Research Society.
4. Baker, M., "Qualitative Research in the US and the UK: A Contrast in Styles and Practices," Montreux, 1987, Proceedings of the Esomar Congress.
5. Hansen, J., "Vorstandsmitglieder zum Thema Marktforscher," Allensbach am Bodensee, 1976, Institut fur Demoskopie.
6. Simmons, M., "The British Market Research Industry," Brighton, 1978, Proceedings of the BMRS Congress.
7. Bronner, A.E., and Hoog, R. de, "Computer-Assisted Decisionmaking: A New Tool for Market Research," Rome, 1982, Proceedings of the Esomar Congress.

USE OF COMPUTER INTERACTIVE INTERVIEWING AT TRADE SHOWS

Jacqueline G. Labatt-Simon
Cahners Exposition Group

INTRODUCTION

Cahners Exposition Group is the world's largest producer of trade shows. As the producer of these shows, Cahners is in the unique position of being able to have its entire target audience at one location over a period of a few days. This opportunity has been heavily utilized by regularly conducting at-show research with attendees. Up to a year ago, these interviews were always personally administered by interviewers. In the past year, personal interviews have been supplanted at many shows by computer interviewing using Ci2 (Computer Interactive Interviewing) software.

WHAT TYPE OF INTERVIEWS SHOULD BE USED

The decision as to which type of interviewing to use is based on a variety of factors, but the single most important factor is the anticipated receptivity of the attendees to computer interviewing. Computer interviewing at trade shows is reminiscent of the little girl with a curl in the middle of her forehead. When it works, it works magnificently, but when it doesn't work, it fails miserably. With trade shows there is no replacing the interviews. The show ends and it's all over. Therefore, it is essential that the correct decision be made before the show begins.

Show attendees vary from owners of Mom and Pop hardware stores to cooks to computer engineers. While any of them might be willing to try a computer interview at a mall, few are willing to embarrass themselves when surrounded by peers at a show.

Basically, computer comfort level can be correlated with frequency of use of computers. Design Engineers and Management Information System Managers use computers daily and will line up and wait to take part in a project using computers to do the interviewing. Owners of small stores and cooks, on the other hand, have rarely, if ever, seen a computer up close, and are very reluctant to try a computer for the first time in front of their peers. Thus, computers are a natural for the INFO show, where instead of the quota of 500 interviews planned, we completed almost 800 interviews in a four day period. But they would most likely meet resistance at the National Hardware Show or Fancy Food Show.

In fact, our one mistake was at WestPack, a show which attracts a mix of Packaging Engineers and owners of small businesses. Packaging Engineers loved the computers, but we had to run off paper questionnaires and quickly find more interviewers to obtain the responses of the small businessmen. Obviously, we didn't know our audience as well as we thought we knew it.

Cost is a lesser factor in the decision as to whether to use computer interviewing. The savings obtained by the elimination of most interviewers and data entry are usually equal to the cost of renting computers, hooking up electrical lines, and providing guards for security during non-show hours.

You noticed, perhaps, that I said "elimination of most interviewers," not all interviewers. Interviewers are still used to screen respondents so as not to lose valuable time on the computers with ineligible respondents. Additionally, interviewers monitor the computers to deter people from walking away before their questionnaires are completed. And there are always those attendees who consider themselves too important to sit at a computer. Rather than lose the interview, an interviewer acts as a secretary and enters the responses. Finally, even at a computer show there are attendees who are afraid of computers. Again, in these cases, the interviewer will ask the questions and enter the responses.

Timing is another difference between computer interviewing and paper-and-pencil interviewing. Computer interviewing requires more upfront time for programming the questionnaire. While, with many types of projects the time is made up by elimination of data entry, a show schedule cannot be changed and so this time cannot be made up. However, the elimination of data entry and a system that provides data in hours is a definite boon for trade shows. Since shows are scheduled years in advance, plans for the next show begin as soon as this show ends. The fast availability of data enables the research to be incorporated into the plans for the next show.

ADVANTAGES OF COMPUTER INTERACTIVE INTERVIEWING

Computer interviewing has some distinct advantages over interviewer-administered questionnaires. I've already mentioned the faster availability of data. Another is that computer-interactive interviewing can be fun. The color monitors are attractive and create a carnival atmosphere. Attendees frequently stop and ask what we are doing. Sometimes they even sit down without being screened. And many take the time to tell us that they found the interview enjoyable -- something they rarely say with interviewer-administered questionnaires.

Time does not appear to be a factor. Despite attendees' rush to get back to work or to a conference session, most lose track of time as they ponder the questions themselves rather than being read to by an interviewer.

Another advantage is elimination of interviewer error. The time pressure at a trade show frequently encourages interviewers to abbreviate or skip questions, or to forget to follow rotation and skip patterns. With Ci2, questions are always asked in the same way, and rotation and skip patterns are built into the program.

We have managed to somewhat overcome one limitation of Ci2 by use of creative programming. This limitation is the difficulty of obtaining responses to open-ended questions which can be of great importance to

Cahners. The natural instinct of respondents is to try to bypass open-ends rather than take the time to type in long answers. This is similar to "no answers" with self-administered questionnaires. Our research suppliers have come up with an innovative approach to encouraging response to these questions. When a respondent tries to bypass the question, he gets a message saying "Your answer is important to us" and the screen returns to the open-ended questions. He must enter something to get to the next question. True, many see through the program and enter garbage, but the majority do give a thoughtful answer when requested to do so.

DISADVANTAGES OF COMPUTER INTERACTIVE INTERVIEWING

The loss of ability to probe unclear or superficial responses is a major drawback of computer interviewing. For example, at a recent show attendees were asked, "What will you remember most about this show?" Many typed in simply "CAD/CAM exhibits" with no qualifiers about these exhibits. Others said too many or too few CAD/CAM exhibits. Do we code those with no qualifiers as positives, negatives, or neutrals? An interviewer would have followed up with a question such as, "What about the CAD/CAM exhibits?" and obtained elucidation of the response. As it is, the responses to the question have limited value.

Computer interviewing, we recently discovered, may be skewing the sample somewhat. Let me explain. Cahners attempts to obtain a sample which is a true picture of the range of job titles and job functions at a show. Since we cannot predict the attendance profile we use random intercept sampling. Additionally, when a group of attendees is encountered, only one from each division of the same company is interviewed, similar to the one-per-family rule in consumer research. When confronted with computers, senior executives appear to delegate the job of responding to a lower level individual. Conversely, as we verified a week ago, with interviewer-administered questionnaires the senior executive is as likely to take on the responsibility of being the respondent as he is to delegate the responsibility. The reasons for this occurrence are unclear, but the pattern can easily create skewed samples.

ACCEPTANCE OF COMPUTER INTERVIEWING

When computer interactive interviewing was first suggested to me, my immediate reaction was curious but fearful. When if it doesn't work? Will people really sit down at computers and complete questionnaires themselves? Show management, my clients, had the same fears, magnified by the certainty that computer interviewing would increase costs.

Well, six projects later, I can honestly say it has worked. And it hasn't increased costs. Not only has it worked, it has created an unexpected public relations benefit. The color monitors are highly visible and are traffic-stoppers for attendees and exhibitors alike. Exhibitors are intrigued and ask questions about what we are doing. This provides a sales opportunity by allowing us the chance to promote Cahner's interest in providing exhibitors with honest information about show attendees, information which the exhibitors will receive after the show. Their curiosity has frequently resulted in an easy sale of space for the next year's show.

RESEARCH OPPORTUNITIES IN THE TRADE SHOW ENVIRONMENT

Where do we go from here with computer interviewing? I see many directions for innovation and expansion. Booths at our computer shows, such as INFO and CEPS, are filled with personal computers. Why not have exhibitors complete questionnaires at their convenience in the privacy of their own booths? Another idea is to list a market research supplier in the exhibitor manual; exhibitors could conduct their own computer interviews with prospects who stop at their booth.

I read somewhere that a problem with computer interviewing is the need for respondents to sit down at computers at a central location. Trade shows are always a central location so that is not a problem. The combination of a high-interest research method and a captive audience provides almost limitless possibilities.

COMPUTER INTERVIEWING WITH THE MOBILE VAN

Carlos J. Barroso
Procter and Gamble Co.

Overview

Computerized interviewing using Ci2 (Computer Interactive Interviewing) and Zenith 181 portable computers on a mobile testing van has proven to be feasible and to have several advantages over the traditional paper and pencil method. The main advantage to users of the van research has been the rapid turnaround of results. We usually provide tabulated results by noon the day following the completed placement of the test. In some cases, we've produced tabulated reports complete with cross tabulations within a half an hour after close of field.

Background

We contract a mobile van through an outside supplier for running small-base exploratory research. We run a variety of tests including taste tests, sniff tests, show tests, and general surveys.

The van itself is a fully functioning consumer testing facility. It has a kitchen equipped with a microwave, stove, refrigerator, and several outlets for small appliances. The three interviewing stations even have controlled lighting for doing show tests.

We set up the van at medium sized shopping malls and plazas less commonly used for consumer research work. The van stays at a site for only a day and never goes to the same site more than once in a given month. We typically recruit panelists by intercepting them as they come out of the mall or shopping center. If they qualify for the test they are asked to participate by stepping into the van.

The refusal rate for going into the van is surprisingly low - at about 10% - unless the weather is especially severe. After completing the test the panelists are given a gift (something small like a pen and typically less than $1.00 value). In general, the respondents are pleased just to have had a chance to give their opinions.

We started piloting computerized interviewing on the van in March, 1987. Prior to computerized interviewing we did all interviewing by paper and pencil. Typically, we would type the questionnaire for the client, execute the research, and return the completed questionnaires to the client for coding, data entry, and tabulation. Most surveys were, and still are, interviewer administered.

Discussion of Pilot Computer Interview Testing

Overall, it's a positive enhancement to consumer research. Reaction from the interviewers, the panelists, and the clients is very favorable. The interviewers benefit from the automation of skip patterns and the standardization of instructions. The panelists think the computers are interesting and have no problems with self-administered interviews. Clients see the advantage of being able to use more complex skip patterns and randomization of questions, but, mostly they appreciate the very quick turnaround of data.

The interview design and tabulation take advantage of compatibility between different software packages. We often start by typing the frames using MultiMate word processing software. Typically, this is done by our secretary. The next step is to convert the MultiMate document to an ASCII file and read it into Ci2. The only real programming is in providing the logic for Ci2. After the interview is complete, we convert the results to an ASCII file.

Along with the ASCII data set, Ci2 creates files containing information about where the location of the tabbed data are, and with the text from the Ci2 questions that will be used for labels and stubs.

Finally, we use Across, a PC tab package to do the tabulation. Across has a conversion program to read Ci2's files to create its stubs, labels, and formulas. The only task left to the programmer is to create the banner points and specify the desired cross tabulations.

Open-end (voluntary) responses can be handled in several ways. The preferred method varies from client to client. At one extreme, interviewers code the responses on the computer, choosing from a list of codes on the screen. This method was piloted for several hundred interviews while recording the responses on paper in parallel. The responses coded on line were compared to those written down and coded afterward and were found to be almost exactly equal. The client, in this case, was completely satisfied with the on-line coding.

At the other extreme, we do all open-end responses on paper for some clients. They like to read the comments verbatim as recorded by hand. We're flexible on how we handle open-end responses and leave it up to the client to decide. In some cases we both code on line for fast turnaround and write down responses for the client to go over separately.

The interviewing software has been able to accommodate a wide range of questionnaire designs. We've been able to reproduce any paper questionnaire on the computer. Furthermore, we've used skip patterns and randomization schemes that would be too complicated for paper questionnaires. In fact, the Ci2 software has several features we haven't had the occasion to use like continuous scales for magnitude estimation.

Ease of Programming

We're able to train people with little or no computer background to do Ci2 and Across programming. Thus, we use people already available in the van operation to do all the programming. We needed no net increase in manpower to provide computerized interviewing and tabulation. Additionally, with the ability to transfer ASCII files to Ci2 we can use our secretary, with no extra training, to type the Ci2 text. This also satisfies most of the text for the tabulation since the labels and stubs from Ci2 are converted to labels and stubs in Across.

Hardware

We use portable battery operated Zenith 181 computers. They have 640K RAM and two 720K 3-1/2" disk drives. The Zenith 181 has been a reliable battery powered PC with an exceptionally easy to read screen and is quite adequate for our needs. After over a year in the field, we haven't had any equipment failures. However, hardware is improving at a fast pace. Thus, while we were satisfied with the Zenith 181, other users may prefer different hardware. In fact, the Zenith is now available with 20 megabytes, and is still battery operated.

Expand to other consumer research applications.

Given the very favorable experience to date, we're continuing to use computerized interviewing on the van as a standard practice. In fact, we purchased several additional portable PC's for various consumer research studies handled in-house.

DEVELOPING COMPLEX COMPUTERIZED QUESTIONNAIRES

Ann Marie Weaver
American Medical Association

A lengthy questionnaire or a questionnaire with complex skip patterns can be problematic in development as well as in the field. It has been a general assumption that computerized questionnaires would eliminate these problems. While a computerized questionnaire can relieve interviewer burden especially in the case of complex skip patterns, it will not necessarily reduce and may increase the amount of time required for development.

This paper will examine a computerized questionnaire which was lengthy and contained complex skip patterns. It will discuss how Ci2 (Computer Interactive Interviewing) logic can be used to program complex skip patterns and questionnaire designs.

Background

The American Medical Association (AMA) conducted a study funded by a grant from The Robert Wood Johnson Foundation (RWJF) which examined the career paths of young physicians. There was particular interest in the number and types of positions, the hours worked, previous positions held, and factors that influenced choice of professional arrangements. The grant stipulated that a computer-assisted telephone interviewing system be used for data collection.

At that time the AMA had extensive experience conducting physician surveys both by mail and phone. Phone surveys, however, had been done exclusively using paper-and-pencil methods. Beginning in March 1986 the AMA investigated available CATI software packages for micro-computers. The AMA obtained Ci2 telephone interviewing software in June of 1986.

The RWJF Study was broken into two phases. Phase I included 8 focus groups, 25 qualitative interviews, 70 telephone interviews utilizing a paper-and-pencil version of the questionnaire followed by 212 telephone interviews using a computerized version of the questionnaire. Development and refinement of the survey instrument was ongoing throughout Phase I.

Phase II consisted of conducting 6,005 computer assisted interviews with the finalized questionnaire. Data collection was completed in November 1987 and analysis is ongoing.

The Paper and Pencil Questionnaire

The final paper and pencil questionnaire was divided into six sections and consisted of 609 questions:

Section A

Screening questions, perceptions of the medical profession as a career, and number of current practice arrangements were asked in this section. Eligible respondents were those who were born after 1946 and whose medical training was completed between July 1981 and March 1986. Respondents were asked in section A "In how many practices do you typically work at least five hours per week?" The response to this question determined the number of times that section B was repeated.

Section B

Section B obtained information about the respondent's practice arrangements. Questions included hours spent in various activities, specialty practiced, employment status, billing, revenue, and patient load. Practices were only discussed in detail if they involved patient care for at least six hours in the most recent complete week of practice. The practice name was restored from answers given in section A. Skips in section B were based on the number of hours the respondent spent in various activities, employer, and specialty. Section B was repeated for up to five current practice arrangements.

Section C

Questions in this section examined the reasons for career choices. Respondents were asked about other employment opportunities they had and reasons they chose their current arrangements.

There were a number of questions requiring data to be restored from both sections A and B. In addition, skips were based on the hours and employment status questions asked in section B.

Section D

Family background questions were asked in this section. These included marital status, spouse's earnings, number of children, parents' educational background, educational debt, and race.

Section E

The questions in this section were predominantly specialty specific, regarding the number of patients and procedures performed as well as hours spent in different activities. The section was skipped completely if the total number of hours in direct patient care was less than 20. The skips were based on specialty information obtained in the first section B.

Section F

Income, insurance costs, and practice expense questions were asked in this section.

The paper-and-pencil version of the questionnaire generally required 30 to 35 minutes to complete. Most respondents surveyed had only one practice. It was very difficult for the interviewers to follow the skip patterns and restore information collected in earlier questions.

The Ci2 version of the questionnaire ranged in length from 12 to 60 minutes. The length varied by the number of practices reported. The average time to conduct an interview was 25 minutes, and the burdens of skips and restoring text were alleviated.

Developing The Ci2 Questionnaire

Development of the Ci2 questionnaire did not begin until a final draft of the questionnaire was available. Because of the complexity and length of the questionnaire we felt that a nearfinal version would reduce the number of modifications that would need to be made.

The conversion of the paper-and-pencil questionnaire to a computerized version began in early December and a period of six calendar weeks was allowed for development. Initially, the task was split between two AMA staff. Their Ci2 experience was limited to work on one prior Ci2 questionnaire. Because of time constrains, in mid-December the questionnaire was sent to an outside firm to be developed. Staff at the firm had extensive experience using Ci2 and developing market research questionnaires. The firm put the majority of the text and logic on Ci2 in 40 hours. When the Ci2 questionnaire was returned it took an additional 120 hours to complete the questionnaire logic.

Modifications were made on the Ci2 version of the questionnaire throughout the pretest field period. Final revisions of the instrument were made at the completion of the pretest. Eight days were allowed for implementing the final changes and testing.

Although logic was continually tested as it was developed, additional testing done because of the complexity and length of the questionnaire. This testing was done by conducting mock interviews using the Ci2 questionnaire. Eighty hours of testing were done on the Ci2 questionnaire. This method identified additional logic errors.

The Ci2 System 1000 was needed because there were 779 questions. Broken down by type there were:

633	NUM Questions
5	RNG Questions
25	OPN questions
65	NOA Questions
45	CON Questions
6	ANY Questions

The questionnaire was composed predominantly of NUM (Numeric multi-digit) questions. NUM questions were used rather than RNG (Single-keystroke range) for two reasons. First we wanted a value that could be consistently used for don't know/no answer responses. Zero was the value used for don't know/no answer except where zero could be a valid answer.

NUM questions were also used to allow interviewers the opportunity to check the answers they typed. Since NUM questions were used, features available for standard questions (i.e., conditional skips, TRM (Terminate) and OTH (Other Specify)) could not be used.

Range questions were used when TRM statements were required and to save the text of an answer to a question so that it could be restored later.

Most "or" skips (i.e., if question 1 equals 2 or question 3 equals 4 go to question 5) could be programmed using Ci2's jump and skip instructions alone. Additional logic was necessary for "and" skips (i.e., if question 1 equals 2 and question 3 equals 4 then go to question 5). Geometric coding in combination with jump and skip instructions was used for these skips.

Geometric coding is a method of assigning unique numeric values to combinations of responses. Response categories for each variable are assigned a value from the geometric series (1,2,4,8,16..[2] in new questions using SEF (Set if) statements. These questions are then added together. When values within a geometric series are added together the sum is unique for each combination of values. Jumps and skips were then made based on the value of these combinations.

SET and JIF (Jump if) statements compare the value of a question to a constant. We needed to develop logic that would allow us to compare the value of two different questions. We made use of mathematical operations to compare question values. Subtraction is used to identify the larger of two questions. Division is used to identify whether a question in section B has the same value for all practices.

Most mathematical operations were stored in questions other than NOA (No answer required). Caution needed to be used however, because respondents must not skip over questions that contain math logic. If the logic is skipped, the mathematical operation is not performed. Also, if a computed value is stored in a question that is later seen by the respondent the computed value is written over. If the respondent skips the question the computed value remains and is not written over with a zero. This can result in data that is outside the valid range of values as well as other invalid data.

The RES (Restore) statement was used throughout the questionnaire to restore answers in the text of questions. Restore can be used to create answer categories for a constant sum question.

Several questions had to be changed on the Ci2 version of the questionnaire from the paper-and-pencil version because of software limitations. One limitation was the inability of basing skips on an open response. An additional question was added to the Ci2 questionnaire to code the open response and the skips were based on the coded value. The range of plus or minus 32,000 and integer arithmetic also forced changes in the questionnaire. Income figures to the nearest thousand dollar were stored rather than the actual dollar amount because of computations that needed to be performed on those figures.

Summary

It is advantageous to use computer-assisted interviewing for complex questionnaires. A computerized questionnaire will reduce the interviewer burden of restoring information and following skip patterns.

Ci2 is very easy to use and has simple logic instructions. Ci2 should not be overlooked because of its simplicity when developing complex questionnaires. Ci2 logic can be manipulated to achieve most needs. The development of complex questionnaire logic will take more time and creativity then other questionnaires. While some limitations do exist, Ci2 is very powerful when it is used to its full capacity.

Testing of complex logic is essential. A review of the logic as well as running mock interviews are the best ways of checking accuracy. Sufficient time should be allocated to test the instrument. Segmenting a lengthy questionnaire and working with a near-final version will facilitate development and testing.

COMPUTER INTERACTIVE DATA COLLECTION AT HONOLULU INTERNATIONAL AIRPORT: UNATTENDED KIOSK INTERVIEWING

Glenn M. Okimoto
State of Hawaii
Department of Transportation

BACKGROUND

The State of Hawaii Executive Budget system requires that information on the efficiency, safety, and economy of the various activities throughout the state be gathered. For the Department of Transportation, these activities are divided into distinct and separate operating units, such as Honolulu International Airport. The information is formulated into "measures of effectiveness" and presented to the legislature to justify the expenditures in the budget. In this way, decision makers are able to make wellinformed decisions and properly allocate the limited resources of the state.

In 1986, a new measure of effectiveness was included in the airports systems' budget. This measure quantified the public's satisfaction with the airport facilities. In the summer of 1986, paper-and-pencil surveys were developed and distributed at each of the six major airports in Hawaii to measure this satisfaction. On a a scale of 1 to 10, Honolulu International Airport received an overall rating of 7.21 which was above average as compared to the other airports in the state system.

In order to make the data collection easier, more efficient, and more visible, a computer interactive survey was developed. In this way, data could be gathered on a 24 hour, 7 day-a-week basis throughout the year. In early August of 1987, the first computer was installed in the central lobby of Honolulu International Airport. By the first week in September, over 5,000 responses were collected.

GENERAL DESCRIPTION

The computer-interactive survey was developed on an IBM Infowindow computer system. This system consists of a IBM PC XT which includes 20 megabyte hard disk drive and a 80286 processor. The monitor, an IBM Infowindow, utilizes a technology that allows respondents to touch the portion of the screen that corresponds to their responses. The keyboard is available but only used for programming purposes. This system was chosen because it was thought to be best suited to administer an unsupervised, unattended survey situated in a public environment. The IBM Infowindow system would prevent "hackers" from accessing the system and ruining the files.

The graphics were developed by Magnus Communications Design, a company from Vancouver, Canada, who programmed the Infowindows at the 1986 EXPO. The graphics are eye-catching, entertaining, and provided in both English and Japanese. The introductory screen shows two hula dancers, dancing to and fro, coconut trees, and the Department of Transportation logo - all in vivid color. The second screen shows a close up of a hula dancer who winks when you decide to continue the survey. The proper 'skips' and 'branches' are included as the survey asks whether you are leaving, arriving, seeing someone off, or greeting someone.

The computer, then, takes you through the various questions on the agricultural and security checkpoints and the many services at the airport, such as the directional signs, restaurants, gift shops, and overall satisfaction. For the questions that ask for a response from 0 to 10 with 10 being very satisfactory, the numbers appear as colorful pineapples across the bottom of the screen. There is some animation as automobiles drive into parking stalls, baggage moves down conveyor belts, and the Wiki Wiki shuttle buses roll along. At the end of the survey, if a respondent wishes to write comments, a keyboard appears on the screen and the comments can be typed in. All in all, it is a very clever and appealing survey.

ADVANTAGES AND DISADVANTAGES

The major advantage of the system is that it allows anyone to use it and no one is left out. In the public sector, because there are no alternatives, no choices are available; there is only one major airport on Oahu, and everyone is a critic. Many individuals want to express their opinions, but often are unable to. Public hearings may not be conveniently scheduled, visits to legislators are too time consuming, and letters to public officials never get written. The computer-interactive survey allows people the opportunity to "voice" their opinion and allows them to vent some frustrations. It is useful in minimizing major complaints.

The visibility of the system also helps to change public perceptions of public agencies. After the installation of the computer was featured on the local news and the United Press International Wire Service, many local market research firms, inquired about the system and about other possible uses. The installation of the Infowindow system at Honolulu International Airport was one of the first in the state. This indicated to the public that the airport officials were concerned individuals wanting to provide the best possible service.

User acceptance was excellent as many people wanted to answer the survey. At times, a line formed as people waited for a chance to input their opinions. Over 25% of the respondents took the time to type in additional comments on the airport. These comments were extremely useful in pinpointing problem areas and in identifying potential problems. The comments have also been useful in justifying additional airport expenditures to the state legislature. Many respondents also commented that the survey was well done.

The disadvantages arose because of the newness of this data collection method. Many problems relating to the data collection method and the statistical uncertainty of the results were encountered. Being an unsupervised, unattended survey, where anyone could answer the questions, anyone did, especially young children, who viewed the computer as an alternative to video games. Older adults hesitated to or were unable to approach the computer. There was also a tendency for the older persons to not take the survey seriously because it did look like a game. This problem showed up in the demographics of the respondents where the age distribution was highly skewed towards the younger age groups. This statistical significance question, however, may not be important because of the large sample size.

There was also the problem of the honesty of the respondents. Although studies have shown that respondents are more honest with a computer than with a human interviewer, the location of the computer may have caused some people to be dishonest. The computer was placed in a cabinet situated in highly trafficked and visible areas of the terminal. This was necessary in order to prevent vandalism and to ensure that the survey was noticeable and used. But, on many occasions, because there was an audience as the respondent was completing the survey, there is some doubt as to whether the questions were answered truthfully, especially the income question. Direct observation indicated that many people looked around to see who was watching before responding to any question. Often, an apparently unattached man would emphatically press the $75,000+ income button when women were present and watching.

RESULTS

In general, all of the ratings were lower in 1987 than 1986. The differences could be attributable to a real decrease in the quality of the airport facilities (which is unlikely because the new Diamond Head lobby was opened and several other improvements were made). It could also have been that the summer was exceptionally hot and humid, and the non-air-conditioned terminal caused lower ratings. There are too many variables involved to conclusively state that the difference in the rating was due to lower service.

The decrease in ratings may reflect technical differences between the surveys; 1986 was a paper-and-pencil survey administered by human interviewers while 1987 is by a touch screen computer. The July 1987 issue of American Demographics shows some evidence that people are more honest with a computer than with a human interviewer. There is less interviewer bias. In many instances, a person is likely to increase the rating when asked by another person and subconsciously tries to "please" the interviewer.

Another factor that could affect the results is that more residents of Hawaii are answering the questionnaire. In 1986, 20% were residents of Hawaii; in 1987, 30% of the respondents were residents. Residents may be especially critical of the airport because they observe the same problems day in and day out with no apparent improvement. This was shown in the rating of the parking facility, which dropped dramatically from 6.3 to 4.9. In 1986, with human interviewers we could target travelers and minimize resident responses. But with the unattended kiosk this was not possible and may have caused some of the differences.

Comparison with 1986

The following table shows the changes between 1986 and 1987.

Table I
USER RATING OF HONOLULU INTERNATIONAL AIRPORT
COMPARISON 1986 V. 1987

	1986	1987
Roads at the airport	6.891	6.232
Airport parking facilities	6.360	4.943
Shuttle bus between terminals	6.877	5.723
Curbside loading zones	6.605	5.387
Airport directional signs	6.726	5.565
Baggage claim areas	6.570	5.532
Airport public conveniences	7.050	6.298
Airport visitor information	7.151	6.314
Overall Satisfaction	7.297	7.033

SURVEY DEMOGRAPHICS

The differences in the ratings between 1986 and 1987 may also have been caused by the age and income distribution of the respondents of the airport survey. When compared to the demographics of the Hawaii Visitors Bureau (HVB) survey, the airport survey is highly skewed to the younger age groups, with over 10% of the respondents less than 10 years old. The HVB survey shows that this age group is only 3.1% of visitors. The age groups of 20-29 and 30-39, however, exactly match the HVB results and make up over 40% of the respondents. Also, the older age groups are under-represented as only 10% of the respondents were older than 60 years old and HVB statistics shows 16%. (See Table II)

The income distributions also showed this type of disparity. There were 26% of the respondents in the over $75,000 group. But on the other extreme, there were 13.4% of the respondents with income of less than $15,000 annually. (See Table III). It is well known that travel is a luxury that only people in the higher income brackets can afford. This discrepancy in the income distribution of the respondents to the survey once again reflect the effect of the under 10 age group on the survey responses.

There is little reason to expect, however, that the DOT and HVB demographics must match exactly; airport users and visitors are different groups of people. The Hawaii Visitors Bureau surveys actual travelers by sending a mail-out survey after the traveler arrives home. This makes the age and income distribution of the HVB and the DOT surveys difficult to compare. The DOT survey collected the responses from airport users. This included airport and airline employees, well-wishers, concessionaires, and other non-travelers.

Table II
Age Distribution of Survey Respondents
Honolulu International Airport
including under 10 age group
1987

AGE in years	% of Respondents	
	DOT	HVB
under 10	12.4	3.1
10-19	28.9	7.7
20-29	19.7	18.0
30-39	15.5	21.7
40-49	8.9	19.0
50-59	5.1	14.9
60-69	2.6	15.6 (60+)
70+	6.7	

Table III
Total Annual Household Income Distribution
Honolulu International Airport Users
Computerized Survey
1987

Household Income	% of Respondents		
	HVB	Dot	General Pop.
less than $15,000	6.0	13.4	28.5
$15-19,999	16.3	6.5	9.8
20-24,999 >		5.0	9.3
25-34,999	21.7	11.4	18.1
35-49,999	24.4	13.5	19.2
50-74,999	20.0	23.3	11.7
75+	11.8	26.6	3.3

FUTURE POTENTIAL

This a an exciting new area of data collection that shows great potential. It is an attractive and eye catching way of collecting survey data. It is appealing to the public because everyone has a chance to state an opinion.

There are, however, several problems that must be addressed before this method becomes widely accepted and used. The biases introduced by the computer-interactive method of data collection, must be accounted for. Currently, we have a joint project with the University of Hawaii, Departments of Marketing and Engineering to look further into this area and to propose solutions. I am looking forward to the future of computer interactive data collection using unattended kiosks.

DISKS-BY-MAIL: A NEW SURVEY MODALITY

A Panel Discussion Moderated By:
Marshall G. Greenberg

National Analysts Division of Booz·Allen & Hamilton Inc.

Introductory Remarks

While the mailing of computer diskettes is a relatively new survey modality, it combines several elements of other survey modalities which have been in existence for a considerable amount of time.

First, the survey instrument is an electronic questionnaire. Typically, it is user-friendly with on-screen instructions programmed on a floppy disk for use in an IBM-compatible microcomputer. The use of computer-assisted personal interviewing (CAPI) systems dates back to the 1970's, when they were typically administered in central locations or in office environments with an interviewer present.

Second, the use of the mails in accessing survey respondents, particularly in commercial and industrial markets, is certainly not new. Disks-By-Mail (D-B-M) surveys typically send an electronic questionnaire with a return mailer via either the U.S. Postal Service or one of the commercial overnight delivery services.

Finally, in many instances D-B-M surveys incorporate the use of telephone prerecruiting of respondents to screen for eligibility and to enlist a higher degree of respondent cooperation. Telephone recruiting has been an established technique in the conduct of survey research for decades.

Despite the fact that survey researchers have extensive experience with the individual components that make up D-B-M surveys, this new modality has surfaced a number of problems and issues that need to be addressed by practitioners. As an industry, we are just now beginning to move up the learning curve as we gain experience with the technique and we can all benefit by sharing our experiences, both our successes and our failures.

This session will address a number of logistical issues related to the conduct of D-B-M surveys and report on their experiences. Among the topics to be addressed are the following:

- Questionnaire Preparation
  - Faulty Diskettes
  - Respondent Instructions
  - Pretesting
  - Complexity
  - Flexibility
- Mailing
  - Cover Letters
  - Mailing Methods
  - Post Office Handling
- Respondent Cooperation
  - Incentives
  - Completion Rates
  - Curiosity
- Costs
  - D-B-M vs. Mail
  - D-B-M vs. Telephone
- Other
  - Hardware Compatibility
  - Use in Foreign Countries
  - Quality of Data
  - Data Processing
  - Interrupted Interviews
  - Timing

Finally, we will discuss the future of D-B-M as a survey modality. In particular, we will examine the likely impact of current and projected trends in the survey research industry on the use of this new modality and examine its advantages and disadvantages versus the use of mail and telephone surveys among commercial and industrial respondents.

Concluding Remarks

The survey research industry is struggling with the implementation of D-B-M as a survey modality. There are at least three major problem areas.

1. Lack of routinization and economies of scale

Because the volume of D-B-M surveys conducted by any given establishment has been relatively low to date, it has been difficult to approach the longer term potential efficiencies in costs and schedules. First, the cost of programming questionnaires is high, because relatively few professionals have either extensive experience or the opportunity for continual practice. Second, the merging of data bases from alternative survey components (e.g., Computer Interactive Interviewing (Ci2) and Adaptive Conjoint Analysis (ACA)) can still be tricky. The process requires special handling and, while it has gotten easier over the past year or so and will be further simplified over the next couple of years, it remains a nettlesome and time-consuming task. Third, we are still trying to determine the optimal methods for soliciting respondents and achieving high interview completion rates. There is a long history and an extensive literature on soliciting respondent cooperation by mail and by telephone, but we are just now beginning to develop a base of experience using D-B-M procedures.

2. Lack of standardization in respondents' facilities

As you have heard, even some respondents who are willing to participate in D-B-M surveys are unable to do so because of problems with hardware or software incompatibility. While such procedures as telephone prescreening and an 800-number "hotline" can be helpful in addressing these problems, some environments are likely to remain impenetrable for some time.

3. Inaccessibility of some potential respondents

Despite the proliferation of IBM-compatible microcomputers, it is likely that for virtually any given universe of commercial and industrial customers, some will be inaccessible -- either because the right hardware is not available or because they are averse to using a computer. It should be noted that, while such nonresponse is a potential source of respondent bias in D-B-M surveys, it may be no greater than other sources of nonresponse bias using alternative modalities. For example, D-B-M surveys may elicit cooperation from large numbers of respondents who refuse to cooperate in telephone and/or paper-and-pencil mail surveys. If so, D-B-M surveys could actually reduce the amount of bias due to nonresponse.

Having cited these three problem areas, now I'd like to offer a more optimistic outlook for D-B-M as a survey modality, since each of these problems is likely to be alleviated over time.

First, as we move up the experience curve with D-B-M surveys, as with any new product, we can expect to achieve greater routinization and to realize the resultant economies of scale, thereby lowering costs and shortening schedules.

Second, the computer industry is moving in the direction of greater standardization leading to fewer problems of hardware and software incompatibility.

Finally, with more people becoming computer literate and the ever-increasing penetration of microcomputers, respondent accessibility will be even less of a problem than it is today.

In fact, I believe that the D-B-M survey methodology will be advantaged in many applications against both telephone and mail methodologies among commercial and industrial populations. Exhibit 1 compares these alternative survey modalities on a number of critical dimensions, reflecting what are necessarily subjective judgments based on my own perceptions of industry trends.

1. Cost

On balance, D-B-M surveys should enjoy a cost advantage over telephone surveys and perhaps a slight edge over Paper-and-Pencil by Mail (P-P-M) surveys.

All other things being equal, questionnaires are most costly to prepare for D-B-M surveys and least costly for telephone (non-CATI) surveys. Data collection for both D-B-M and P-P-M are significantly more cost efficient than for the telephone mode because the latter requires an interviewer.

If telephone prerecruiting is used in a D-B-M survey, some of the cost advantage is lost, but in most cases there will still be economies in data collection over administering the entire interview by telephone. Finally, D-B-M surveys are cost advantaged in data processing because the data are already input and the need for cleaning is lessened by forced adherence to skip patterns.

2. Timing

Telephone is, and will continue to be, significantly advantaged over both D-B-M and P-P-M methods of data collection in terms of the time required to complete a survey.

D-B-M requires more time in the questionnaire development stage than either the telephone (non-CATI) or P-P-M approach. Given the time required for mail-out, return and follow-up, both D-B-M and P-P-M typically require much more time in the data collection stage than does telephone. While some of the time disadvantage associated with the D-B-M mode may be made up in the data processing stage, telephone surveys will almost certainly require less elapsed time to complete than either of the alternatives discussed here.

3. Quality

Long term, I believe that the use of D-B-M surveys has the potential for significant advantages in quality over both telephone and P-P-M approaches.

First, the D-B-M methodology provides an opportunity to break through the clutter among oversampled respondents in commercial and industrial markets, thereby achieving high completion rates and reducing sample bias. Many of these people have been inundated with requests for cooperation in P-P-M and telephone surveys and are, consequently, less willing to participate in them anymore. D-B-M can arouse the curiosity of potential survey respondents, particularly those in high technology positions (e.g., telecommunications managers). The knowledge that they can complete the questionnaires at their leisure, rather than being trapped on the telephone, increases the likelihood that respondents will cooperate in one of the mail approaches as well. Furthermore, it is virtually impossible for a secretary or administrative assistant to screen the survey instrument prior to passing it on to the targeted respondent.

Second, the quality of data collected using D-B-M should be superior to both telephone and especially P-P-M approaches. A properly programmed computer-assisted interview will ensure that proper skip patterns are followed and that responses fall within permissible ranges. The validity of responses will also be enhanced using D-B-M if respondents, as anticipated, are willing to take greater care in completing the interview because they are doing it at a time which is convenient for them. Validity can also be enhanced by the ability in electronic questionnaires to control for order effects by randomizing certain item sequences.

Finally, D-B-M offers a level of versatility in data collection that cannot even be approached with either P-P-M or telephone methodologies. The use of ACA, Adaptive Perceptual Mapping (APM) and other adaptive techniques in electronic questionnaires allows the researcher to employ procedures for multivariate data collection and analysis that embody a degree of power and efficiency unique to computer-assisted interviewing techniques.

In summary, D-B-M should provide higher quality survey data at lower cost than either telephone or P-P-M modalities, although it is unlikely ever to match the speed with which a telephone survey can be completed. Thus, if the need to conduct a survey can be anticipated and planned early enough to permit a D-B-M approach to data collection, it may well become the method of choice for surveying commercial and industrial populations.

EXHIBIT 1

Comparison of Alternative Survey
Methodologies in the Future

	Disks-by-Mail	Paper-and-Pencil by Mail (P-P-M)	Telephone
<u>Cost Factors</u>			
Questionnaire Preparation	0	◐	●
Data Collection	●	●	0
Data Processing	●	0	◐
Total Cost	●	◐	0
<u>Timing Factors</u>			
Questionnaire Preparation	0	●	●
Data Collection	0	0	●
Data Processing	●	0	◐
Total Timing	0	0	●
<u>Quality Factors</u>			
Sample Bias	●	0	◐
Data Quality			
Mechanics	●	0	◐
Validity	●	◐	0
Versatility of Data	●	0	0
Total Quality	●	0	◐

Legend: ● Superior ◐ Average 0 Inferior

DISKS-BY-MAIL: A NEW SURVEY MODALITY

PROBLEMS AND OPPORTUNITIES

Lesley A. Bahner
POPULUS, Inc.

I'd like to address the problems and opportunities presented by conducting mail surveys using diskettes for data collection. I will spend more time on the problems than on the opportunities. This is not to suggest that the opportunities aren't great, because they are, but that the problems of disk mail surveys are magnified compared to those done by other methods. When you add the complexity of disks to a mail survey, you introduce the possibility of many problems; and of course any non-mail disk survey is complex enough by itself. When the two are combined, mail and disk, the opportunity for exponential disaster exists.

PART I: PROBLEMS

Problems Are Magnified: Sites, Acceptance, Supervision

First, think about the number of interviewing sites. There could be 500, 1,000, 2,000 or more. Now envision the survey disk landing on each respondent's desk or being found among the pile of mail at home. How does that person accept your request for information? Lastly, imagine that you are personally interviewing these respondents: supervising them, guiding them through the interview.

Interviewing Sites

Unlike other survey methods, we are not dealing with ten telephone stations, one or two interviewing locations at a trade show, not even three or four computer interviewing stations at each of 10 or 15 mall facilities. We are talking about 500 or 1,000 or more interviewing locations around the country.

This means that the variety of computers and operating software is enormous. For as many interviews as you are conducting, an equal number of equipment configurations may exist. There is no standardization like there is among those of us who do computer interviewing in a controlled environment.

Compatibility. So, one of the biggest problems is computer compatibility; and this problem is going to get worse before it gets better as the market grows and new operating systems are developed. For now, and maybe for the next twelve months, we may be able to safely ignore those who have Apples, Macintoshes, 3-1/2" disk drives, OS/2, and even those who supposedly have IBM-compatible computers, particularly the AT&T and Panasonic.

We have partial solutions to the incompatibility problems. One is to make disks self-booting which prevents the Sawtooth software from interacting with any conflicting software on a respondent's computer.

But a self-booting diskette presents another dilemma: what to do about complying with your DOS licensing agreement. Legally, we would have to buy DOS for each respondent's computer. So far we have not gone to that extreme (we did, however, discuss the issue with IBM's legal department to no avail). Instead, we have put an appropriate copyright notice on each survey diskette when we have chosen to make the disk self-booting.

Variety also exists with monitors: monochrome, CGA, Hercules, Plantronics, EGA, VGA. This, fortunately, is an easy problem to deal with, thanks to Ci2's (Computer Interactive Interviewing) VID command. VID defeats the color programming of the interview for monochrome monitors and for those who have software that is incompatible with the Sawtooth systems, such as resident screen management software and an ANSI.SYS resident file.

In fact, we always program the questionnaire in color and always use the VID command. This makes it more interesting for those who have, and are used to, color.

If we make the diskette self-booting and if we use the VID command, a lot of respondents are going to successfully get into the program and be able to run the questionnaire without technical problems. But some are not. What happens to these people?

<u>One Interviewer Per Diskette</u>. Although we try to avoid them, we nevertheless compound the likelihood of technical problems by some choices we've made about how we want the questionnaire administered. In most cases, we want only one respondent per diskette: we want data only from the person to whom the questionnaire is sent. Therefore, we are going to allow only one questionnaire per diskette. We do this with NUMSTART. Or, if the questionnaire is multi-module, it <u>must</u> have NUMSTART to work. And to make it look better, we go directly into the interview, bypassing the respondent number screen. This mode has its advantages, and its disadvantages.

In most cases, as I've said, this presents no problem. But for those who do have compatibility troubles, or just don't do things exactly as you have instructed, or are the victims of a power failure (this happens!), these respondents, and you, are out of luck because the interview cannot be restarted. We've gotten some phone calls when this happens because people really do want to complete the survey, and we tell them what to do: ERASE *.DAT and COLRFILE. We can only guess at how many people don't call us.

One thing we have found to be very helpful is to encourage respondents to call our office for help. Besides being helpful to the respondent, it helps us identify problems occurring out in the field. In this way we are also likely to get more data, and it is good for public relations.

Faulty Diskettes. Of course, sometimes there is no quick and easy solution to a problem, such as when a respondent gets a blank or faulty diskette. This happens for two reasons: The only practical way to duplicate diskettes for a mail survey is to use a disk duplicating machine. While it is efficient to reproduce the survey en masse, it is often impractical and expensive to thoroughly check each one with a CHKDSK command or using Norton Utilities to do a DISK TEST of the files. You can expect some, a small percentage, not to work. One solution is to convey to respondents our willingness to send another diskette.

Then there is the Post Office. This delivery system can be hard on disks, more so than on a mailed paper survey. Again, a small percentage is going to get bent or mangled, either on its way to the respondent or as it returns to you. While I can't give you an exact percentage of how often this happens, it can be compensated for by a slight increase in the size of the mailing.

Medium Acceptance

People have to be willing to complete a survey using their computer. This acceptance of the medium is paramount in our decision to conduct a survey by mail on disk. Not every survey universe is going to be appropriate.

Appropriateness to Universe. Quite logically, those who currently use IBM-PC compatible computers, either regularly or occasionally, are prime subjects. This isn't just restricted to those who have computers at their desks; if one has access to a machine in the work environment, that person can most likely use it for the short amount of time it takes to complete a questionnaire. This universe is growing as well. Computer familiarity and literacy is expanding. More and more people have access to a computer.

Interest and Motivation. Once we have determined that our universe is appropriate, the next task is to pique interest and encourage motivation of the sample. This isn't too hard to do since the medium certainly is attention-getting. More than that, however, is that a survey on disk is easier to do than one on paper. And, in some cases, it is more relevant. For example, an engineer wrote us that he wouldn't have bothered if it hadn't been on diskette. This makes me think that computer interviewing may soon do to traditional market research what word processing has done to the typewriter.

Suspicion About Intent. While we can deal with many of the problems, another is becoming more prevalent. The fears of worms, viruses, and Trojan Horses are spreading, especially among experienced computer users. How do we go about quelling their suspicions about our quite innocent intent?

We do everything we can to look legitimate and professional. We do this by sending personalized letters on company letterhead, using printed disk labels and disk sleeves with our company name.

Reassurances can also be given that the respondent will not be copying anything onto his/her system, which is how these viruses spread. Another way is NOT to make the disk self-booting. This of course may reassure the respondent, but as I've said, a non-self-booting disk creates compatibility problems for us. One word of caution: Don't forget that one of your respondents may give you a virus. We suspect that this happened to us and now we accumulate data from disks on a computer that doesn't hold valuable programs and data.

Mistakes. One last thought on the acceptance of disk mail surveys. If we don't work to eliminate our errors and overcome the problems, we will eventually turn people off. And then this new survey modality will die and the great opportunities it provides will vanish.

Respondent Supervision

That brings me to my last problem: respondent supervision. Unlike other interviewing techniques where a respondent has access to someone for guidance or explanation, that help must be built into the cover letter and questionnaire.

Explicit Instructions. Explicit instructions that anticipate respondent problems are key. This includes things like defining the ENTER key, which some only know as "return" or by the arrow symbol ("<-"). These are in addition to making the task instructions throughout the questionnaire very clear.

Length of Interview. It is also important to create accurate expectations about the length of the interview.

PART II: OPPORTUNITIES

Now for the opportunities.

Ability to Conduct "Personal Interviews"

Of key interest to us and our clients is the ability to conduct "personal interviews." Computer assisted interviewing allows us to come close to this objective for several reasons.

Added Control. We have the least amount of control over mail surveys. Compared to personal interviews, telephone surveys, and self-administered computer assisted questionnaires at central location facilities, all of which can be customized according to each respondent's answers, paper-and-pencil mail surveys cannot. With the questionnaire on disk, however, we can control the flow of questions, have skip patterns, and select applicable questions for each person.

Complexity. We can also ask more complex questions. We can conduct more complex studies. Conjoint measurement and perceptual mapping have become available to us. In this way, we can conduct better mail surveys.

Ability to Reach the Hard to Sample. One advantage of mail surveys is the ability to reach those who are hard to sample, such as those with low incidence and wide dispersion. The disk modality, with its advantages of control, makes these samples more accessible. Not only that, but also for many of these hard-to-reach people, the method represents ease and relevance. A questionnaire on disk is the only reason some will take the time and make the effort to answer our survey.

DISKS-BY-MAIL

Richena Morrison
Morrison & Morrison, Ltd.

In the summer of 1987, the personal computer industry was positioning itself for yet another major advancement in technology. With the advent of IBM's micro-channel systems, PC users faced a challenging decision: which side of the technological fence should a company commit to?

Our client, PC WORLD MAGAZINE, wanted to determine at an early stage how corporate and individual buyers were reacting to the announcement of IBM's new systems.

For such a study, the Ci2 (Computer Interactive Interviewing) interview system was a perfect match. What better way to interview computer users about new computer systems than by using disk-by-mail interviews!

Since PC WORLD MAGAZINE is the leading personal computer publication for IBM and compatible machine users, we were assured of a receptive target for Ci2 disks-by-mail.

This study also allowed us to compare the disk-by-mail methodology against traditional mail surveys the client had previously fielded.

Presented here are several aspects of this particular study in which Ci2 was successfully utilized in a mail survey. The logistics of preparing, delivering, and accruing responses from Ci2 interviews in a mail survey should be considered carefully prior to fielding a study. We hope the following will provide you with some insights to that process.

SURVEY MATERIALS

As with other types of surveys, the importance of procuring extra materials remains a constant with disk-by-mail studies.

For this study, standard 5 1/4" floppy disks were purchased in bulk from a reputable office products supplier. The formatting procedure revealed that close to 5% of the new disks were faulty. To reduce the risk of disk failure in the field, we discarded any disks with bad sectors.

The disks were prepared with specially printed labels, which listed the return mailing address. Although a self-addressed, pre-stamped cardboard disk mailer was provided to the respondents to return the completed interview disks, several disks were received in other packaging. It is advisable, therefore, that disk-by-mail labels include the return address in case the survey materials become separated.

A cover letter in the mailing package included a direct telephone number for assistance should the respondent encounter any difficulties.

To this point, the sophistication of respondents should not be overestimated, even with a clearly "computer literate" target such as computer magazine subscribers. Several phone calls were received concerning the Ci2 interview, the most memorable being a respondent who did not know how to load a disk into the floppy drive!

RESPONSE RATE

This disk-by-mail study generated a good return rate in a short period of time.

A 68% response was received from this study. Sixty percent of the total disk mailing was returned within three weeks. (see Figure 1)

This disk-by-mail study equaled but did not exceed the client's usual response rate for standard mail surveys. The significant advantage with this disk-by-mail study was in the speed of the returns. Responses to this survey were accrued in half the time of the client's standard mail surveys.

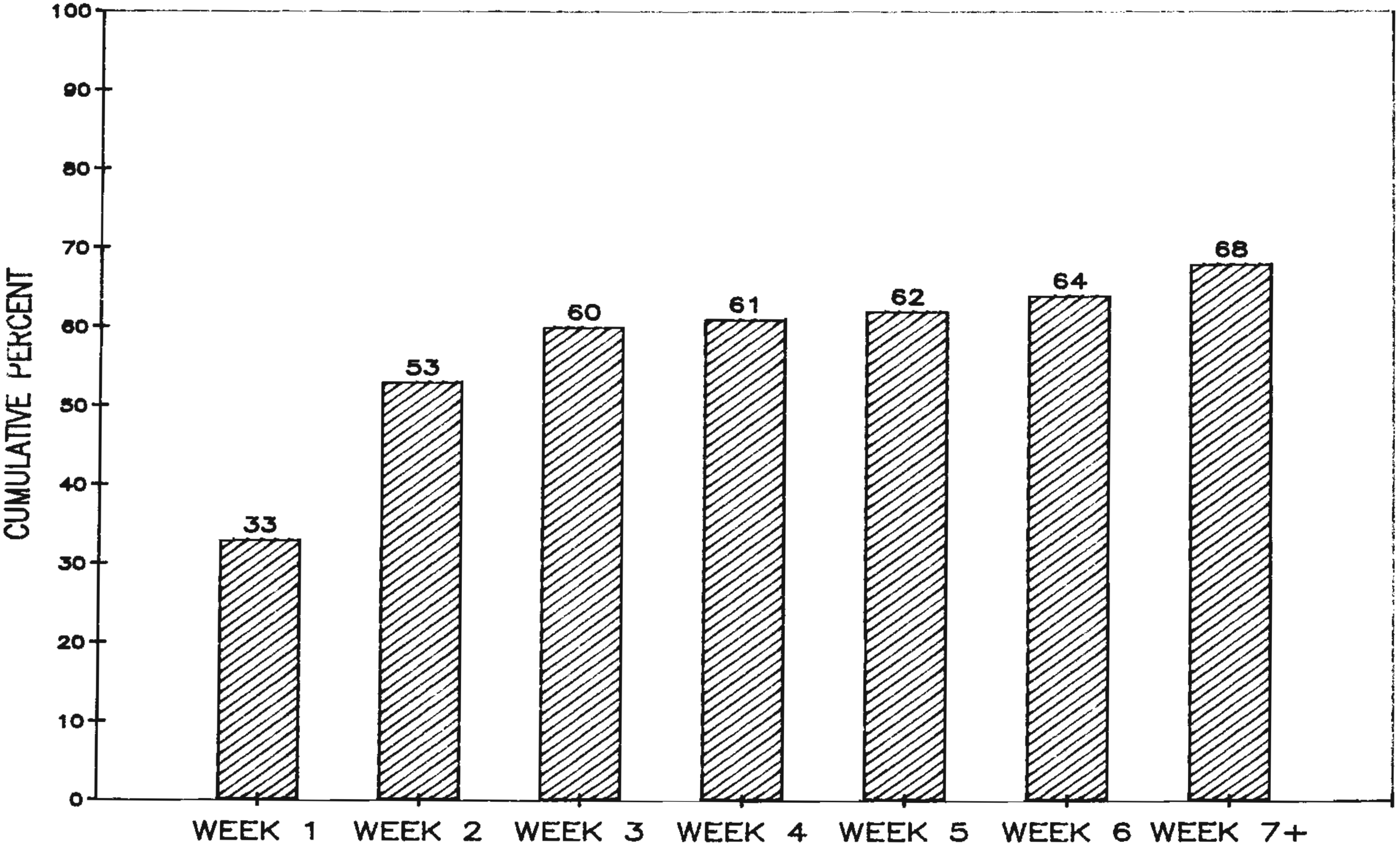
The effect of including incentives in a disk-by-mail survey was tested in this study. Fifty percent of the sample were given a $1.00 incentive, fifty percent received no incentive.

The inclusion of incentives did not result in a significant difference in the return rate.

Returns	
53%	Incentive Paid
47%	No Incentive

It appears that the inherent interest, high level of respondent curiosity, and relative ease of completing self-administered Ci2 interviews are sufficient to elicit participation and cooperation without monetary incentives.

FIGURE 1
A DISK—BY—MAIL STUDY
CUMULATIVE RESPONSE RATE



SOURCE: MORRISON & MORRISON. LTD.

FIELD EXPERIENCES

The nature of this disk-by-mail study required that only one respondent per disk be interviewed and tabulated. One way to accomplish this requirement would have been a laborious procedure of preparing the 500 field disks individually using Ci2's SETNUM procedure.

Fortunately a new procedure, QF2, had become available. The QF2 procedure is a small execution program which can be copied onto each field disk. It creates a unique respondent number by reading the user's machine clock and recording the time elapsed since the machine was booted as an identification code.

In theory, since interviews would not begin at the same moment in time, duplications were believed to be unlikely.

In practice, however, this survey with less than 400 responses yielded several duplicate respondent codes.

Frequency	Code
3	015
2	111
2	129

Researchers preparing studies which necessitate unique respondent numbers should be aware that the current QF2 program has a likely possibility of creating duplicate identification codes. Steps should be taken to check the data and, if necessary, renumber the respondent ID codes.

The requirement for one respondent per disk also created some special problems. Once the interview had been started, whether it was completed or not, the program would not permit a second entry.

Respondents who had interrupted the interview by removing the disk could not re-enter the program to complete the survey. In a self-administered interview, this condition can occur frequently. About six percent of the returned disks in this study were incomplete, and can be attributed to interrupted interviews.

Competing technologies in the computer industry have created a variety of incompatible and non-standardized machine types. These inconsistencies must be recognized and accounted for when fielding disk-by-mail studies.

Foremost, the researcher must establish that IBM or compatible computers are available to the respondent before fielding survey materials.

Yet even within the IBM/compatible PC market, important differences in machine types occur. High density disk drives, 3 1/2" disks, and machines with less than 256K of memory can make Ci2 interviews unusable by the respondent or data unreadable by the researcher.

Although these variations in machine compatibility occurred with less than 3% of the sample in this disk-by-mail study, future changes in the state of the PC industry may heighten these problems.

The variety of monitor displays available in the field also should be anticipated when preparing text screens for disk-by-mail surveys. The new Ci2 VID command allows respondents to indicate color or monochrome displays, and suppresses color commands accordingly.

The VID command cannot change the screen width of the text frames. Creating frames in the 40 column width format is often preferred because the text is large and easy to read when viewed on color monitors.

Viewed on monochrome monitors, however, the 40 column width frames appear only in the left hemisphere of the display. This can be disturbing for respondents, especially when the interview switches frequently from full screen display (80 column) to half screen (40 column).

As with other types of mail surveys, the handling processes of the postal service can damage disks, despite several layers of protective packaging. In one case, a disk returned to us was so severely warped and melted we speculated that somewhere a postal clerk had been issued a blow torch.

RESPONDENT REACTIONS TO DISKS-BY-MAIL

To gauge respondent reactions to the Ci2 interview, an open-ended question was included to allow participants to express their opinions about the disk-by-mail format.

Respondent reactions to the Ci2 disk-by-mail survey were overwhelmingly positive. The following are direct quotes from the respondents :

> "This was the first survey I enjoyed completing!"
>
> "It certainly is more interesting than the regular 'bingo' card method."
>
> "This was an interesting concept for a survey. I probably would have taken longer to answer a standard survey."
>
> "I am more likely to fill out and return a survey conducted like this versus one on paper."
>
> "This is a good idea. I normally throw away surveys sent to me, but this one caught my attention."

> "I rather enjoyed the experience and look forward to doing it again."
>
> "This was fun! I could hardly wait to get to the computer to see what was on the disk."

Participants in this study generally found the Ci2 interview to be fun, easy, and quick to complete. Even though the interview contained approximately 60 screens, the average length of time spent on the survey was less than ten minutes.

Several participants were so enthused about the Ci2 interview that they suggested similar surveys be accessible from a bulletin board or on-line service. This is an interesting concept which, for some market studies, may be worth pursuing.

ANALYSIS

Overall, the data base constructed from this disk-by-mail survey was easy to manipulate with standard statistical and tab package software.

However, a troublesome exception occurred when complicated user subgroups were requested.

This Ci2 interview utilized extensive brand lists, repeated over several screens to account for multiple product purchases. Therefore, in order to select some user subgroups, lengthy logic statements were required which proved to be beyond the management capabilities of our statistical software.

To avoid these problems, researchers are advised to limit brand lists to manageable rosters, and to incorporate "filtering" questions which can be used to select subgroups in a straightforward manner.

RECOMMENDATIONS

The incorporation of a RE-START procedure is highly recommended for any disk-by-mail study which is limited to one respondent per disk.

Interrupted and incomplete interviews are an inherent hazard to self-administered Ci2 surveys. The inclusion of a DOS batch job which can delete any respondent data, and permit re-entry into the interview should be considered. The file can be copied onto the field disks along with the Ci2 interview. Of course, respondents should be informed that this procedure is available.

Open-ended questions should allow for several lines of data. Regardless of the question being asked, respondents often use the open-ends for communicating additional ideas and comments which they feel are important to the study sponsor.

The X-BACK key should be explained thoroughly in text screens and reiterated periodically throughout the interview. Respondents in our studies have shown a tendency to forget the X-BACK procedure.

With sufficient preparation, foresight, and a fair dose of common sense, disk-by-mail studies can be easily and effectively constructed. The advantages of disk-by-mail studies are quite evident and undoubtedly will continue to produce higher respondent interest than traditional mail surveys.

INTERNATIONAL SELF-ADMINISTERED INTERVIEWS

Brent Dahle
Customer Satisfaction Research Institute

I. Introduction

An ESOMAR Symposium on International Marketing Research held a few years ago in Paris featured three papers on the viability of international telephone interviewing from a central location. In the recent decade, such an effort was also attempted from a location in the United States. With the advent of self-administered computer interviews, the question of international applicability surfaces. Our purpose, here, is to address a method whereby computer interviews can be effectively administered internationally.

In early 1987, we decided to develop an international telephone department. This department now consists of 30 interviewers and staff. Native-speaking, multi-linguals were hired and trained. The regular staff currently includes interviewers for the following languages: German, French, Italian, Spanish, Portuguese, Korean, Japanese, British English, Swedish, Norwegian, Russian, and Arabic. In addition, several dialects are represented, especially for Europe.

The highly educated team of native-speaking interviewers has conducted a great number of studies, from single-country to multi-continental surveys.

II. A Case in Point - Germany

A. Background on the Study: Client had a study in process using a self-administered questionnaire, on floppy disk. Market research (MR) firm was located in Germany. MR firm encountered difficulties in recruiting adequate numbers of respondents and phone numbers, resulting in extraordinary cost overruns (nearly double). Client and MR firm negotiated and renegotiated, trying to complete study. Impasse was reached and we were offered the job, with restraints as to the time allowed to complete. We were very busy, but to accommodate, we agreed to finish the work for a premium over our usual rates (given the very "tight" time constraint).

B. The Task: To complete 40 self-administered questionnaires (at businesses) in three (3) weeks, including mail time. Screen respondents, send the diskettes via courier, and receive the completed surveys back (also via courier).

C. How it was accomplished: Screened 135 respondents immediately (overkill due to time constraints), sent to all 135 respondents via courier. It was not easy to recruit the 135 since we had no phone numbers, no addresses, only company names and cities. However, we do have 335 phone directories for Germany in our city library. We sent an interviewer to the library daily to search out the numbers we needed to call each night. Even though we were assured as to computer compatibility in every case, some machines/softwares were not. We recruited heavily enough so that would not deter us from our goal. We also recruited heavily due to the need to get 40 back ASAP. We then sent all diskettes out via courier, prepaid, and began receiving them back.

D. Results: We received 39 returns in the time permitted. An additional 6 were received (good) after the deadline and in time to be included in the final results. Nine were received unusable, either incomplete or incompatible.

E. Costs: Our costs were less than quoted by 8%, and slightly less than the German firm's actual costs, including our courier costs each way. The original firm's quote had been grossly understated to begin with -- their actual costs had turned out to be in line with ours. Finally, had the courier costs not been incurred, we would have been 6% less in cost than the original quote from the German firm.

III. Some Do's and Don'ts in the International Arena

A. Why we were more successful than the local firm.

1. Higher response in screening due to uniqueness of method.
2. Fortunate access to phone directories for Germany in our library (German "directory assistance" is a misnomer) and international business club (Wer Macht Was).
3. Singularity of specially-couriered information packets arriving from overseas.
4. Detailed instruction sheets and cover letters.

B. What to consider in international self-administered interviews.

1. Do's:

    a. Look at each country individually; each country is unique.

    b. Assess the computer compatibility issue for each country; IBM does not dominate in every country, as it does in some.

    c. Consider mailing methods and costs.

        - Courier vs. remailer vs. regular mail

          Courier = control and speed;

          Remailer = cost savings and speed;

          Regular mail = cost savings.

        - Forwarding costs = courier (Italy & Sweden as examples)

        - Customs regulations ("not for resale" declarations) - courier

        - Return billing guarantees - courier (European cost examples)

    d. Ascertain experience of firm you plan to engage in your study.

    e. Be certain to include cover letters regardless of how much is explained orally in screening process (etiquette an important issue -- France as example).

    f. Make certain you choose the correct translation service: technical expertise, knowledge of (U.S.) English versus language considered, "currentness) of considered language, auditing/cross-check capability separate source on this).

    g. Be aware of regional differences and not-so-obvious language nuances (Italy, Germany & Belgium as examples).

2. Don'ts:

   a. Don't assume the sample is sufficient

      - incompatibility of computers

      - uncertain response rate

   b. Don't assume you can get the phone numbers readily (Belgium as example).

   c. Do not engage a firm (U.S. or local) without cost guarantees. Also consider your current exchange rate and whether that can work for you or against you. (Consider all variables and cost issues.) Watch out for Latin American inflation rates.

   d. Don't assume that the Far East cannot be done. It can.

   e. Don't forget to draw on advice from experts: consulates, embassies, US C of C in that country, MR firms, international research consultants, etc.

   f. Do not assume that international is the same as domestic with a few differences. It is actually <u>different</u> with many similarities (cost per interview shock to firms engaging in their first international study as example).

   g. Don't begin a study if there are loose ends and/or no contingency plans. Delays and cost overruns can kill you. Prepare in advance to ensure success.

   h. Don't be afraid of international. It is simple and profitable, once you know the rules.

DISKS-BY-MAIL: A NEW SURVEY MODALITY

Thomas L. Pilon and Norris C. Craig
IntelliQuest, Inc.

INTRODUCTION

The advantages of mail questionnaires have been well documented. The major advantages are low cost, respondent convenience, anonymity of respondent, accessibility to a widely diverse sample population, ability to administer lengthy interviews, and elimination of interviewer bias. The major disadvantages are the time to completion, low response rates, and nonresponse bias. (Alreck, Joselyn, Kress, Lehman, Peterson, Rossi)

While there have been numerous studies examining mail surveys from various perspectives (Armstrong, Childers, Furse, Harbicht, Huxley, Kanuk, Linsky, Weiss, Wolfe), only a few studies (Goldstein, Higgins) have been published on the use of diskette-based mail surveys (hereinafter referred to as DMSs).

This paper will first examine the feasibility of DMSs. Next, the advantages of conducting DMSs will be discussed. The third section will discuss the issues and options associated with DMSs. The advantages and issues and options sections will include a discussion of specific experiences with DMS projects conducted by IntelliQuest and other projects with which IntelliQuest is familiar. The concluding section will include a discussion of key factors for success in a DMS.

FEASIBILITY

A prerequisite for participation in a DMS is that the respondent have access to a personal computer (actually, at the present time, it must be an IBM PC compatible personal computer). Opportunities for conducting DMSs increase proportionately with the penetration of PCs in the population. Although only 25 percent of those that work at a desk have a PC compatible on it, approximately 75 percent of the people who work in offices have access to a PC compatible computer. (All figures in this section are from COMTEC, a subsidiary of the Gartner Group, Stamford, Conn.) However, these numbers vary considerably with the size of the firm. Only 20 to 25 percent of the firms with less than ten employees have one or more PC compatibles. Approximately 98 percent of those firms with 500 or more employees have one or more PC compatible(s).

The penetration of any one type of personal computer in the residential market is so low that it is not practical to consider using DMSs for consumer interviews at the present time.

ADVANTAGES

I. Response Rates

There is significant evidence that DMSs result in higher response rates than the traditional paper-based mail surveys (hereinafter referred to as PMSs). In the only published comparison available, Higgins et al. found a significant difference in response rates (78% DMS vs 63% PMS).

Although there were not any paper based control groups to enable direct comparisons, numerous DMS studies conducted by IntelliQuest have resulted in unexpectedly high response rates, when expectations were based on traditional PMS response rates. Response rates were usually in the 30 to 50 percent range.

II. Response Speed

Higgins et al also examined response speed. In their study, the average response time for a PMS was 8.85 days. Their average response time for a DMS was a significantly lower 6.68 days. Although the authors did not discuss possible reasons for the shorter response time, it should be recognized that one possible reason is the novelty of receiving a DMS. The novelty may wear off at some point and mitigate this effect.

The theoretical concerns with increasing response speed have to do with minimizing history and maturation effects. However, the practical concerns of delivering a study to the client earlier should not be overlooked. Since the data entry task is eliminated with DMS (except coding of open-ends), the time from the mail drop to when the data are in the computer, ready for analysis, is shorter still.

III. Response Accuracy

The wordiness of open-ended responses and the number of points made with those words were also examined by Higgins et al. In their study, the mean number of words was 31.02 for the PMS compared to 39.24 for the DMS. The mean number of points was 5.84 for the PMS and 6.89 for the DMS. Both differences were significant at the .05 level.

Other aspects of response accuracy include proper branching and response consistency checks. For example, if proper branching is forced, the researcher is never left to ponder what was meant when a respondent answered the "Why Not" question after responding affirmatively to the previous question.

Many researchers conduct response consistency checks as a routine part of their analysis. If a subject's responses are determined to be inconsistent, one popular solution is to discard that subject's data. With a DMS it is possible to program consistency checks into the questionnaire. If responses appear to be inconsistent, the respondent can be asked for an explanation or alternatively asked the question a third time.

IV. Questionnaire Complexity and Length

Researchers are familiar with the difficulty in obtaining responses to lengthy paper-based mail questionnaires. The respondent is able to estimate the time required to complete by paging through the questionnaire before to deciding whether or not to complete it. The longer the estimated time required to complete, the lower the likelihood of the respondent actually completing it. Of course, with a DMS it is not possible for the respondent to make this type of estimate in advance.

Higgins et al asked respondents to estimate the time to complete the DMS after completing it. The mean estimated time to complete was 23.12 minutes. The mean actual time was 30.18 minutes. (The actual time was determined by having the DMS software read the computer's internal clock at the beginning and end of the interview.)

Since respondents are not able to estimate the time required to complete a survey beforehand and afterwards think it took less time than it actually did, it may mean that researchers will be able to use slightly longer questionnaires using DMSs rather than PMSs.

Other aspects of questionnaire complexity and length include branching and rotations. In order to successfully administrate the traditional PMS, researchers are forced to avoid using complex branching and rotations. Researchers are not faced with this limitation in DMS. Also, by controlling the flow of the questionnaire, researchers are able to prevent respondents from reading ahead and biasing order sensitive questions.

Since proper branching is forced and respondents are not able to respond to questions not intended for them, the questionnaire length may be shorter.

V. Project Costs

Higgins et al. found that DMSs were no more expensive than PMSs once the initial program development costs were absorbed. IntelliQuest has had similar experience. Cost savings result from eliminating the typesetting and printing, and the elimination of the chores of data entry and typing of verbatims.

Additional costs result from having to program the questionnaire, duplicate the disks, and buy diskette mailers (it is possible to reuse most of the returned mailers). For shorter surveys, a DMS may require slightly more postage than a PMS.

ISSUES AND OPTIONS

I. Non-response Bias

Perhaps the biggest concern to researchers about DMS is the potential for non-response bias since, for many studies, not all potential respondents in the sample have access to a PC compatible. Non-response bias occurs if those that do not respond are different from those that do respond and if the difference affects what is being measured by the study. No one is in a better position to estimate the potential for this bias than the researcher conducting a particular study.

The researcher must trade off the various types of non-sampling error (non-response bias, interviewer error, data entry errors, etc.) against one another and against sampling error (caused by sampling instead of conducting a census) and trade both of these off against budgets and deadlines (i.e., in a blind world, a 1-eyed person would be King). The ultimate objective is to minimize total error to the extent that is possible given budgets and deadlines.

II. Response Rate

There have been many studies about various ways of increasing response rates to paper mail surveys (Childers, Furse, Harbicht, Kanuk, Linsky, Weiss, Wolfe). Most of these results should generalize to diskette based mail surveys.

IntelliQuest has experimented with several methods and combinations of methods:

- Pre-qualifying respondents by telephone is a very useful way to clean a list. It helps to eliminate what could be termed as pseudo non-response bias. It is not non-response bias if an unqualified respondent does not respond. However, if a respondent is not pre-qualified, a researcher would erroneously count unqualified non-respondents in his calculation of response rate.

  The pre-qualification telephone call can also be used to enlist a qualified respondent's agreement to complete the forthcoming questionnaire.

- In one study, first class postage was used for one-half of the diskette mailers. Bulk rate postage was used for the other half. The first class postage group had a 32 percent response rate, while the bulk rate postage group had only a 26.7 percent response rate.

- In another study, four groups were selected to receive the four combinations of $1 incentive/no incentive and reminder card sent 5 days later/no reminder card. The group that received both the $1 incentive and the reminder card had a 46 percent response rate. The group that received neither had a 33 percent response rate. The $1 incentive/no reminder card had a 39 percent response rate and finally the no incentive/reminder card group had a 34 percent response rate. For this study, it seems that $1 incentive worked well by itself and better in conjunction with the reminder card. The reminder card did not seem to work well unless it was used in conjunction with the $1 incentive.

- In yet another study which included a $1 incentive, followup telephone calls were found to significantly increase the response rate.

III. System Compatibilities

Many of those that have tried the DMS methodology have technical and support horror stories to tell. For the most part, these horror stories have to do with system incompatibilities (within the framework of supposed PC compatibles). These incompatibilities center around three components of the PC system: Monitors, Disk Drives, and Clocks.

Monitors

The most frequent complaint about monitors is "the screen is fuzzy, I can't read it." This problem is usually caused when a composite color monitor (two-color: black and either green or amber) displays a color program. Since it is not possible to test (from within the software) what type of monitor is attached to a color graphics adapter, it is advisable to use only color combinations that provide a high degree of contrast. Another alternative is provide the respondent with a simple means of solving this problem if it occurs. Remember that although your respondents may have access to PCs, they may know virtually nothing about them.

Disk Drives

The are two potential problems relating to disk drives. The first is diskette size. Presently, the majority of PC compatibles use 5 1/4 inch diskettes. However a rapidly increasing number use 3 1/2 inch diskettes. It is still a good bet to assume a respondent will have access to a PC compatible with a 5 1/4 inch drive, but if respondents are pre-screened, the opportunity to ask should be taken.

The other potential problem relating to disk drives is that diskettes written on by high density drives on many AT or 286 class machines cannot be read by most PC or 8086/8088 class machines. High density drives should not be used to write on diskettes for distribution. Also, it is important to have access to a machine with a high density disk drive because respondents may use one to take the survey.

System Clock

The problems with the system clock center around some versions of MS-DOS which cannot read the system clocks on certain machines (e.g., PC DOS, Compaq DOS, and Leading Edge DOS can't read the clock on Corona brand machines). If a questionnaire includes commands that depend on the system clock (i.e. the NOA command in Sawtooth's Ci2 (Computer Interactive Interviewing), the respondent's system will hang if the software cannot read the clock.

The only guaranteed solution to this problem is to not use commands that depend on the system clock. Another good solution is to not put any version of DOS on the diskettes and to provide respondents with instructions that will enable them to start the interviewing program from an already booted system. (It should be noted that it is technically not legal to put DOS on the DMS diskettes under current copyright and licensing laws.)

KEY FACTORS FOR SUCCESS

I. Use only when Appropriate

Like any other methodological tool, attempts to use DMSs when they are not appropriate can be disastrous. The researcher must either check or have strong reason to believe that an overwhelming proportion of the sample frame has access to a PC compatible. The researcher must also be sure that certain types of respondents are not systematically excluded. For example, since the incidence of PC compatibles in small businesses is lower than the incidence of PC compatibles in larger businesses, a random sample of all businesses may result in a systematically higher non-response bias for the smaller businesses.

II. Organization and Logistics

If the keys to real estate are location, location, and location, the keys to a successful DMS are organization, organization, and organization. Check lists such as the ones shown in Figures 1 and 2 greatly facilitate planning and organization. For larger studies, it is helpful to have access to a disk duplicating machine.

III. Instructions/Support

As with any new technology, ease of use is a critical issue. Each respondent should be provided with clear instructions. While many subjects will be PC experts, others may need to be told how to turn the computer on and that it is necessary to hit the big key (enter, return, etc.) after entering DOS commands. To avoid boring the expert but, at the same time, providing the necessary level of detail to novices, consider enclosing two sets of instructions (perhaps labeled "Experts" and "For the Rest of Us").

It is critical to include instructions for the unexpected. What are respondents to do if they get bumped off the computer they were using when they were only half way through? What do they do if they hit the wrong key or change their minds? Perhaps the best fail-safe for instructions is to provide an 800 number that respondents can call for help.

IV. Paper Mail Surveys Rules

Most of the rules that have emerged in the literature with respect to paper mail surveys will also apply to diskette based mail surveys. Writing a good cover letter (the hook), keeping the questionnaire length reasonable, making the questions interesting, and paying attention to format and appearances are all still very important.

REFERENCES

Alreck, Pamela and Robert B. Settle, The Survey Research Handbook, (Homewood, Illinois: Richard D. Irwin, Inc., 1985).

Armstrong, J.S. and T.A. Overton, "Estimating Nonresponse Bias in Mail Surveys," Journal of Marketing Research, August 1977, 396-402.

Childers, Terry, William M. Pride, and O.C. Ferrell, "A Reassessment of the Effects of Appeals on Responses to Mail Surveys," Journal of Marketing Research, August 1980, 365-70.

Furse, David H., David W. Stewart, and David L. Rados, "Effects of Foot-in-the-Door, Cash Incentives, and Followups on Survey Response," Journal of Marketing Research, November 1981, 473-8.

Goldstein, Harris, "Computer Surveys by Mail," paper presented at the 1987 Sawtooth Software Conference.

Harbicht, Robert and Terry Atwill, "Nontraditional Incentives Spark a Response from an Upscale Audience," Marketing News, September 13, 1985, 10.

Higgins, C. A., T. P. Dimnik, and H. P. Greenwood, "The DISKQ survey method," Journal of the Market Research Society, Volume 29, number 4.

Huxley, Stephen J., "Predicting Response Speed in Mail Survey," Journal of Marketing Research, February 1980, 63-76.

Joselyn, Robert W., Designing the Marketing Research Project, (New York, New York: Petrocelli/Charter, 1977).

Kanuk, Leslie and Conrad Berenson, "Mail Surveys and Response Rates: A Literature Review," Journal of Marketing Research, November 1975, 440-53.

Kress, George, Marketing Research, 3rd Edition, (Englewood Cliffs, New Jersey: Prentice Hall, 1988).

Lehman, Donald R., Marketing Research and Analysis, (Homewood, Illinois: Richard D. Irwin, Inc., 1979).

Linsky, Arnold, "Stimulating Responses to Mailed Questionnaires: A Review", Public Opinion Quarterly, Spring 1975, 82-101.

Peterson, Robert A., Marketing Research, (Plano, Texas: Business Publications, Inc., 1982).

Rossi, Peter H., James D. Wright, and Andy B. Anderson, Handbook of Survey Research, (New York, New York: Academic Press, 1983).

Weiss, Louis, "Prepaid Incentives Yield Higher Response Rates to Mail Surveys," Marketing News, January 4, 1985, 30.

Wolfe, Arthur and Beatrice Treiman, "Postage Types and Response Rates on Mail Surveys," Journal of Advertising Research, February 19, 1979, 43-8.

FIGURE 1

Check List of Materials for Mail-Outs

1. Incentive (new $1 bills)

2. Printed materials
    A. Stationery (for cover letters)
    B. Envelopes
    C. Labels (address labels and diskette labels)
    D. Diskette Sleeves
    E. Diskette Mailers
    F. Postcards
3. Diskettes
4. Postage
    A. First Class or Bulk
    B. Business Reply Mail (put money in the account)
    C. Postcards

FIGURE 2

Project: ____________________

Mail-Out Survey Procedures & Steps

DATE **RESPONSIBILITY**

I. PREPARE DISKS

______☐ 1. Provide approved survey ______
______☐ 2. Copy approved survey onto disks ______
______☐ 3. Design survey diskette label ______
______☐ 4. Determine respondent numbering scheme ______
______☐ 5. Print disk labels with appropriate respondent numbers ______
______☐ 6. Apply labels to disks ______
______☐ 7. Run "SETNUMX2 B M" program on each disk to assign respondent numbers ______

II. PREPARE OTHER INSERT MATERIALS

______☐ 1. Write, edit, proof, & spell-check insert materials ______
______☐ 2. Print cover letters and other materials to be inserted. Mail-merge names as required. ______
______☐ 3. Sign all cover letters as necessary ______
______☐ 4. Obtain necessary dollars or other incentives to be included. Provide adequate security & double check methods as appropriate. ______ ______

III. PREPARE ENVELOPES, POSTAGE & MAILING LABELS

______☐ 1. Provide mailing list ______
______☐ 2. Prepare return diskette mailers ______
______☐ A. Design, print & apply return address labels ______
______☐ B. Determine, obtain & apply appropriate return postage ______
______☐ 3. Prepare out-going envelopes ______
______☐ A. Print (mail-merge) labels matching the cover letters ______
______☐ B. Determine, obtain & apply appropriate out-going postage ______

IV. PREPARE, FILL & SEAL ENVELOPES

______☐ 1. Determine that all disk numbers, cover letters & address labels are in the same correct order. ______
______☐ 2. Insert diskette with a dollar (or other incentive) into return mailer ______
______☐ 3. nsert a cover letter & the return mailer containing the diskette with the number assigned to the respondent on the cover letter into the out-going envelope addressed to the same person. ______
______☐ 4. Insert any other materials as required ______
______☐ 5. Seal envelope ______

V. SORT FOR POST OFFICE & DELIVER TO POST OFFICE

______☐ 1. Deposit anticipated postage fee @ Post Office ______
______☐ 2. Sort by ZIP codes as directed by Post Office for the appropriate postage rates that have been used ______
______☐ 3. Deliver to Post office ______

IV. RETRIEVE RETURNED DISKETTES FROM POST OFFICE

______☐ 1. Open & sort returned diskettes by survey project ______
______☐ 2. Aggregate data files from disks ______

DISKS-BY-MAIL: A NEW SURVEY MODALITY

Harris K. Goldstein
Trade-Off Marketing Services, Inc.

As marketing researchers our mission is to help our companies and clients get to the right place at the right time, namely the bottom line. Our information gathering and analytical responsibilities are analogous to the role of "scouts ahead of the wagon train." Our survey methodologies serve to access the options available.

Today's scouts must do more than explore the physical aspects of the territory. In this era of "market-driven" economies, it is imperative that we get feedback from the "natives"...the customers who will buy or reject our products and services. There are several options available for these studies: secondary data reviews, behavioral observations, focus groups, high traffic intercepts, telephone interviews, mail panels, etc.

Computer interactive surveys via mail represent yet another approach. And, in the opinion of someone who has conducted hundreds of mail surveys, a superior option whenever appropriate. "Appropriateness" is critical. Simply stated, this means your target respondents must have easy access to computers, more specifically, IBM or compatible computers. Common sense and/or some preliminary screening will provided the necessary "go" or "no go" answer. Business-to-business studies represent the most logical applications.

SELLING THE CLIENT ON THE METHOD:

There are several benefits in conducting disks-by-mail surveys. These benefits will differ in value depending upon the circumstances of the project objectives, timing, budgets, etc.

The benefits are:

1. Costs are significantly less than conventional personal or telephone interviewing. Disks via mail will cost from 1/2 to 2/3 less than telephone surveys. Disks via mail will cost from 1/2 to 3/4 less than personal interviews.

2. Response rates tend to be higher than conventional mail surveys. The evidence from our studies, plus what we have learned from other firms conducting disks-by-mail surveys, suggests response rates normally in the 40%-50% range. This compares to the typical 10%-20% response rates from conventional, non-mail panel studies.

While we can find no specific reason for the favorable response rates, it appears to be a function of the nature of the respondents...people who have access to computers are interested in things that involve using their machines. Plus, many people have said they think this is a "novel" way to collect information. Another theory involves anticipated questionnaire length. Long paper-and-pencil surveys can scare off a portion of potential respondents. With diskette-based surveys the respondents do not know what's coming next so their involvement level can be sustained.

3. The quality of information is deemed by many practitioners to be superior to conventional surveys involving interviewers. Published information from Xerox, John Morton, and The Wall Street Journal suggest respondents give more "truthful" answers when interfacing with computers. Disks-by-mail eliminate interviewer bias, allow people to answer when they choose and at their own pace, and can include reference props as needed.

4. Time schedules are not significantly increased. Mail surveys normally take 3-4 weeks for data collection. It is possible to complete this phase in two weeks by mailing to a larger sample, sending reminder cards 3 days after mailing the disks, and if necessary, with pre-study and follow-up phone calls to the targeted sample. Conventional studies, especially with "professional/business"-type respondents, usually require 2-3 weeks for data collection.

5. Mailed diskette surveys allow for comparative testing of independent variables. Test and control cells can be easily created to isolate variables such as product design features, price elasticity, brand image, promotional inducements, etc.

Thus, when we take all of these advantages into consideration it is usually easy to "sell" clients on disks-by-mail. The cost savings are usually the most important variable.

APPROPRIATE TESTS TO CONSIDER:

The "tests" to consider include:

1. Pilot testing the questionnaire.

This first phase should involve personal computer input. The questionnaire author(s) should be in attendance. Probe for confusing instructions, inappropriate answer categories, adequate open-end opportunities. We often use focus groups with computers to "shake-down" the questionnaire. Besides probing the document as written, be sure to ask about things left out, not covered adequately, or even over-emphasized.

2. Test the master disk in several different computers.

Remember, you don't want to "recall" your survey. So take time to check it out on the most common machines (IBM, Compaq, AT&T, etc.).

3. Test your cover letter and incentive offers.

Maximizing the response rate is an important objective. Therefore, take the time to do a test mailing with different cover letter appeals and incentive inducements. What you say and what you offer can affect responses.

4. Test reminder techniques.

Many people want to respond but put the diskette aside. Assume that's the rule not the exception. A friendly reminder is a good idea. Sending a second disk is the best, but can be costly. A postcard, or first class letter, planned to arrive about 5-7 days after the initial diskette is a cost-efficient method. Another possibility is a phone call. The importance of the respondents' participation is underscored by the call. If you cannot speak with the respondent try to get the secretary to act as your ally.

SAMPLE CONSIDERATIONS:

The quality of the sample is critical. Make sure you start with a representative list. Be sure the client is comfortable. Have them sign off. Always try for a list with specific names and titles. These lists cost more but heavy mail survey users say it is worth the extra price.

If possible, call the targeted respondents before mailing. This is an excellent way to confirm the list, get alternative names where appropriate, and initiate a commitment to participate. Pre-calling costs about $8-$10 extra per respondent. I think it is worth the incremental expense. Try to budget the costs in the basic proposal.

INCENTIVES:

We have had excellent luck with crisp $1.00 bills. The addition of a "sweepstakes" for those who respond works for many firms. Common prizes have included vacation trips, home electronics, restaurant meals, and cash. There are no right or wrong incentive programs. A sincere cover letter with $1-$2 will usually get the response you need.

RESPONDENT QUESTIONS:

Answering respondent questions is difficult, so do your homework up front before you mail the diskettes. I suggest you establish an 800 number hot line. Put the 800 number on the cover letter and diskette label. Be sure to clearly indicate when the line is being staffed.

THE FUTURE FOR DISKS-BY-MAIL:

I am very optimistic about this research process. Mailing diskettes represents good value. And, as stated earlier, I feel it is a superior option. I would like to close with a quote from Theodore Levitt's recent book, The Marketing Imagination, "Customers don't buy things they buy betterness. To create betterness requires knowing what customers think betterness to be. This precedes all else in business."

I think surveys-via-mail are an excellent example of betterness.

STATISTICAL ANALYSIS AND THE MARKET RESEARCHER

Anthony Babinec, SPSS Inc.

1 Introduction

The title of this session is "Statistical Analysis and the Market Researcher." In the brief time allotted to me, I want to sketch where we've been and where I think we are headed.

To do anything more than the simplest tallying and calculating, you use a computer. One of the great trends of the last fifteen years has been that computing power has gone up as the cost of computing has gone down. With the advent of personal computers, you can acquire computing power at a modest cost. What is more, you can avail yourself of "organized intelligence"--a sometimes forgotten factor of production--in the form of software produced by experts.

2 Review of Standard Techniques

Let's consider some of the statistical techniques available through software. Figure 2 shows a schematic organization of statistical techniques (which is geared to SPSS procedures but need not be).

Figure 2 Schematic presentation of techniques

- Data Description and Presentation: the "workhorse"/bean counters
  - frequencies, summary statistics, plots(FREQUENCIES)
  - joint frequency tables(CROSSTABS)
  - summary statistics for groups(MEANS)
  - presentation-quality statistics(REPORT)
  - presentation-quality and/or complex tables(TABLES)
- Group Differences
  - two groups(T-TEST)
  - several groups(ONEWAY)
  - two or more factors(ANOVA)
  - two or more factors,covariates,dependent variables(MANOVA)
- Multivariate Techniques
  - regression analysis(REGRESSION)
  - nonlinear regression(NLR, CNLR)
  - logistic regression
  - discriminant analysis(DISCRIMINANT)
  - canonical correlation analysis(MANOVA)
  - principal components analysis(FACTOR)
  - factor analysis(FACTOR)
  - causal models with latent variables(LISREL)
- Categorical Data
  - hierarchical models(HILOGLINEAR)
  - logit models(LOGLINEAR)
  - latent structure analysis(Cliff Clogg's MLLSA)
- Cluster Analysis
  - hierarchical clustering(CLUSTER)
  - K-means approach(QUICK CLUSTER)

Let us consider each area in turn.

First, what most people do most of the time is produce frequency tables and contingency tables. Sometimes this becomes an end in itself. However, you might wonder how

```
CROSSTABS ALL BY ALL
```

can reveal what is going on in your data, since it is easy to generate masses of tables! Making inferences from tabular data is unwieldy with more than a handful of variables. Plus, you risk "mining" the data for evidence of association, thereby spoiling any attempt to keep the Type I error rate down. Extensions of tabular analysis include stub-and-banner tables and tables presenting not only counts but also means and standard deviations of quantitative variables. Research houses and supplier houses make a living preparing these sorts of tables for clients.

Second, there exists a set of statistical techniques for detecting group differences where one or more dependent variables are quantitative. The simplest situation is the two-group t-test, which exists in two forms: independent samples and paired comparisons. From the independent samples t-test flows factorial analysis of variance, while from the paired comparisons t-test flows repeated measures MANOVA.

Third, there exists a set of multivariate techniques for analyzing quantitative data. We use "multivariate" in the loose sense of "many variables." In the broadest sense, some of these techniques assess either the dependence of one or more variables on one or more other variables, or the relationship between two sets of variables when neither set depends on the other. Yet other techniques assess interdependence among a single set of variables.

Various forms of regression analysis, along with discriminant analysis, are standard techniques for examining dependence. Regression techniques estimate a model in which a dependent variable is predicted by a set of independent variables. The standard model is linear in the parameters. Nonlinear regression estimates models which are not linear in the parameters. These models occur much more frequently in the physical/chemical sciences than in market research. A special nonlinear regression technique increasingly seen in market research is logistic regression, where the dependent variable is dichotomous. A related technique is discriminant analysis, where the dependent variable is a categorical variable signifying group membership, and your aim is to use independent variables which help you distinguish cases in the different dependent variable categories as much as possible.

Canonical correlation analysis finds a weighted combination of variables in one set and a weighted combination of variables in a second set which are maximally correlated. Intuitively, you use canonical correlation analysis to ascertain the extent to which two sets of variables are alike.

Principal components analysis and factor analysis are standard techniques for analyzing interdependencies in a set of variables. Principal components analysis is used on a single set of variables to find major "directions" of variation in the data. Principal components analysis is often used as a dimension-reduction technique, for it is often the case that the first few components capture much of what is going on in the set of variables under analysis. Factor analysis attempts to account for the intercorrelations between variables in terms of a relatively small number of "factors." Factor analysis finds subsets of variables which are more highly correlated within the set than across sets.

When you analyze nonexperimental data, you can perform causal modeling via LISREL, which lets you model LInear Structural RELations. These models combine features of regression analysis and factor analysis.

Fourth, when your data are predominantly categorical, an established approach is to use loglinear models or logit models. When you apply loglinear models, the variables in the analysis are treated as jointlydependent. When you apply logit analysis, a categorical dependent variable is predicted by categorical predictors. While these models are appropriate for the data, they can be difficult to apply and interpret. Not seen much yet in market research is latent class analysis, which attempts to account for the observed relations between categorical variables in terms of an underlying unobserved categorical variable. In this respect, latent class analysis is akin to factor analysis on categorical variables. Virtues of the latent class model are that you need not assume multivariate normality nor need you assume continuity of measurement.

Fifth, cluster analysis techniques are used to empirically determine groupings of variables, or more often, of subjects or objects under study. Two common clustering techniques are hierarchical cluster analysis and k-means clustering.

In hierarchical cluster analysis, you first compute a distance matrix (or proximity matrix) between the objects. In doing so, you must decide what measure of dissimilarity to use. Having done so, you must decide on a method for combining cases into clusters or clusters into clusters. Hierarchical cluster analysis has the property--in some circumstances a drawback--that once cases are combined into a cluster they cannot be separated. Another drawback is that the distance matrix gets very large for large sample sizes. A convenient feature of hierarchical clustering is that you can easily examine different orders of solution and make a tentative conclusion regarding the number of clusters.

In k-means approaches, you must first decide on the order of the cluster solution. Having done so, you must specify initial cluster centers. It is important that the initial cluster centers be well chosen, that is, well separated. Some researchers tend to use the first k cases in the data file as initial cluster centers, but there is no reason to think that the first k cases in your file are necessarily well separated. SPSS's QUICK CLUSTER program can select good initial clusters for you, so you should let it do so unless you have somehow established your own values. The k-means algorithm then passes the data and assigns cases to one of the clusters. Each time a case is assigned to a cluster, the cluster center is updated. Thus, after passing the data, the cluster centers are most likely different from the initial cluster centers. In the k-means approach, the data are then passed again, evaluated against the existing cluster centers, and assigned to a cluster. Unlike hierarchical clustering, a case can be re-assigned to another cluster in one of the data passes. Moreover, k-means clustering does not need to keep the distance matrix around, so you can analyze larger files with it. In principle, the k-means algorithm should be applied to the data until no cases are re-assigned. In SPSS, QUICK CLUSTER passes the data a total of three times, but there is no reason to think that this is sufficient. You should consider saving the final cluster centers from QUICK CLUSTER and using them as initial cluster centers in a subsequent invocation of QUICK CLUSTER. You can also save the cluster memberships of each of the cases, and it is easy to employ CROSSTABS on saved cluster memberships to determine whether no cases were re-assigned in the last invocation of QUICK CLUSTER in a session or run. Because you must specify the number of clusters in an invocation of QUICK CLUSTER, you might run QUICK CLUSTER several times while varying the number of clusters.

This completes our rapid survey of common statistical techniques which have been used in market research. We will now briefly consider some newer techniques which have seen some application in market research.

3 New Developments

Figure 3 shows two families of techniques which have been around for a long time and are being applied with increasing frequency to market research data.

Figure 3 Quantification of qualitative data

- Multidimensional Scaling
  - one-mode two-way symmetric scaling(ALSCAL)
  - unfolding(ALSCAL)
  - two-mode three-way scaling,INDSCAL(ALSCAL)
- Nonlinear (nonmetric) Multivariate Analysis
  - homogeneity analysis(HOMALS)
  - correspondence analysis(ANACOR)
  - "nonlinear" principal components(PRINCALS)
  - ...

By "qualitative data" we mean categorical data. If the variable in question is nominal, its numeric scores are arbitrary and its categories have no inherent ordering. If the variable in question is ordinal, its numeric scores are also arbitrary but the scoring now reflects the order of the categories.

In multidimensional scaling, you begin with a matrix of proximities between a set of objects, much as you would in cluster analysis. The proximities might not be actual quantitative distances, but instead ratings along a similarity-dissimilarity scale which you wish to treat as no better than ordinal. Multidimensional scaling produces a plot of the objects in a low-dimensional space, such as two dimensions. If the fit is good, the relative positions of the objects in the two-dimensional space provide a good representation of the "real" distances between the objects in multidimensional space. Examination of the configuration of points in the plot might suggest groupings or market segments. Unfolding models enable you to analyze matrices of preferences, while INDSCAL models allow for differences between individual subjects or groupings.

The expression "Nonlinear multivariate analysis" is used in a special sense, namely, the usual multivariate techniques indicated above are extended to incorporate situations where your data are a mix of nominal, ordinal, and quantitative variables. The nonlinear multivariate routines produce optimal scores for the nominal and ordinal variables in the analysis, while quantitative variables in the analysis are already scored. The scores, or categories, of the variables in the analysis are represented in a low-dimensional space such as two dimensions. Examination of this plot reveals the nature of the relationship between the variables in the analysis.

4 An Example

We now present an example illustrating homogeneity analysis, one of the simpler applications of this approach.

In this example, we look at two variables from the 1984 General Social Survey, which is a multi-stage national probability sample of roughly 1500 noninstitutionalized adult Americans. The variables are PARTYID, that is, party identification, and POLVIEWS, that is, political views along a liberal-conservative spectrum. While this is not explicitly a marketing example, the example is intuitive and illustrates the technique quite nicely.

Figure 4a shows the frequency distributions of the two variables for those cases in which both variables have valid responses.

Figure 4a Frequency distributions of PARTYID and POLVIEWS

```
PARTYID   POLITICAL PARTY AFFILIATION


                                                             VALID     CUM
 VALUE LABEL                  VALUE  FREQUENCY  PERCENT   PERCENT  PERCENT

STRONG DEMOCRAT                   1        249     18.0      18.0     18.0
NOT STRONG DEMOCRAT               2        271     19.6      19.6     37.5
IND,NEAR DEM                      3        204     14.7      14.7     52.2
INDEPENDENT                       4        142     10.2      10.2     62.5
IND,NEAR REP                      5        155     11.2      11.2     73.7
NOT STRONG REPUBLICA              6        242     17.5      17.5     91.1
STRONG REPUBLICAN                 7        123      8.9       8.9    100.0
                                         -------  -------   -------
                              TOTAL       1386    100.0     100.0

VALID CASES     1386      MISSING CASES      0

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

POLVIEWS  THINK OF SELF AS LIBERAL OR CONSERVATIVE

                                                             VALID     CUM
 VALUE LABEL                  VALUE  FREQUENCY  PERCENT   PERCENT  PERCENT

EXTREMELY LIBERAL                 1         29      2.1       2.1      2.1
LIBERAL                           2        130      9.4       9.4     11.5
SLIGHTLY LIBERAL                  3        175     12.6      12.6     24.1
MODERATE                          4        557     40.2      40.2     64.3
SLGHTLY CONSERVATIVE              5        272     19.6      19.6     83.9
CONSERVATIVE                      6        183     13.2      13.2     97.1
EXTRMLY CONSERVATIVE              7         40      2.9       2.9    100.0
                                         -------  -------   -------
                              TOTAL       1386    100.0     100.0

VALID CASES     1386      MISSING CASES      0
```

Both PARTYID and POLVIEWS have numeric scores ranging from 1 to 7.

Figures 4b and 4c show the crosstabulation of PARTYID and POLVIEWS.

Figure 4b Crosstabulation of PARTYID and POLVIEWS--start

```
-------------------------------------------------------------

PARTYID  POLITICAL PARTY AFFILIATION
by  POLVIEWS  THINK OF SELF AS LIBERAL OR CONSERVATIVE

                 POLVIEWS                            Page 1 of 2
         Count  I
        Col Pct IEXTREMEL LIBERAL  SLIGHTLY MODERATE SLIGHTLY
                IY LIBERA          LIBERAL           CONSERVA   Row
                I       1I       2I       3I       4I       5I Total
PARTYID     --------+--------+--------+--------+--------+--------+
              1 I     17 I     43 I     29 I     94 I     28 I   249
 STRONG DEMOCRAT I  58.6 I  33.1 I  16.6 I  16.9 I  10.3 I  18.0
                +--------+--------+--------+--------+--------+
              2 I      2 I     29 I     44 I    116 I     52 I   271
 NOT STRONG DEMOC I   6.9 I  22.3 I  25.1 I  20.8 I  19.1 I  19.6
                +--------+--------+--------+--------+--------+
              3 I      2 I     19 I     50 I     84 I     34 I   204
 IND,NEAR DEM    I   6.9 I  14.6 I  28.6 I  15.1 I  12.5 I  14.7
                +--------+--------+--------+--------+--------+
              4 I      4 I     19 I     14 I     68 I     23 I   142
 INDEPENDENT     I  13.8 I  14.6 I   8.0 I  12.2 I   8.5 I  10.2
                +--------+--------+--------+--------+--------+
              5 I      2 I      4 I     20 I     62 I     33 I   155
 IND,NEAR REP    I   6.9 I   3.1 I  11.4 I  11.1 I  12.1 I  11.2
                +--------+--------+--------+--------+--------+
              6 I      1 I      9 I     14 I    103 I     77 I   242
 NOT STRONG REPUB I   3.4 I   6.9 I   8.0 I  18.5 I  28.3 I  17.5
                +--------+--------+--------+--------+--------+
              7 I      1 I      7 I      4 I     30 I     25 I   123
 STRONG REPUBLICA I   3.4 I   5.4 I   2.3 I   5.4 I   9.2 I   8.9
                +--------+--------+--------+--------+--------+
         Column       29      130      175      557      272    1386
(Continued)  Total      2.1      9.4     12.6     40.2     19.6   100.0

-------------------------------------------------------------
```

Figure 4c Crosstabulation of PARTYID and POLVIEWS--finish

```
-------------------------------------------------------------

PARTYID  POLITICAL PARTY AFFILIATION
by  POLVIEWS  THINK OF SELF AS LIBERAL OR CONSERVATIVE

                    POLVIEWS     Page 2 of 2
            Count  I
          Col Pct  ICONSERVA EXTRMLY
                   ITIVE     CONSERVA   Row
                   I       6I       7I Total
PARTYID    --------+--------+--------+
                1  I    29  I     9  I   249
  STRONG DEMOCRAT  I  15.8  I  22.5  I  18.0
                   +--------+--------+
                2  I    19  I     9  I   271
  NOT STRONG DEMOC I  10.4  I  22.5  I  19.6
                   +--------+--------+
                3  I    11  I     4  I   204
  IND,NEAR DEM     I   6.0  I  10.0  I  14.7
                   +--------+--------+
                4  I     9  I     5  I   142
  INDEPENDENT      I   4.9  I  12.5  I  10.2
                   +--------+--------+
                5  I    28  I     6  I   155
  IND,NEAR REP     I  15.3  I  15.0  I  11.2
                   +--------+--------+
                6  I    36  I     2  I   242
  NOT STRONG REPUB I  19.7  I   5.0  I  17.5
                   +--------+--------+
                7  I    51  I     5  I   123
  STRONG REPUBLICA I  27.9  I  12.5  I   8.9
                   +--------+--------+
           Column     183        40     1386
            Total     13.2       2.9    100.0

-------------------------------------------------------------
```

It is natural to ask: are the two variables related? Figure 4d shows several chi-square measures as well as a number of measures of association for the two-way table.

Figure 4d Measures of association for PARTYID/POLVIEWS table

```
--------------------------------------------------------------

     Chi-Square                Value          DF               Significance
--------------------        -----------      ----               --------
Pearson                      263.51536        36                  .00000
Likelihood Ratio             238.93986        36                  .00000
Mantel-Haenszel               93.11606         1                  .00000

Minimum Expected Frequency -    2.574
Cells with Expected Frequency < 5 -     7 OF    49 ( 14.3%)

                                                                Approximate
     Statistic                 Value        ASE1       T-value  Significance
--------------------         ---------    --------    -------   --------Phi
.43603                          .00000 *1
Cramer's V                     .17801                             .00000 *1
Contingency Coefficient        .39969                             .00000 *1

Lambda :
   symmetric                   .05813       .01205     4.70920
   with PARTYID  dependent     .08251       .01722     4.62380
   with POLVIEWS dependent     .02533       .01072     2.33793
Goodman & Kruskal Tau :
   with PARTYID  dependent     .03066       .00428                .00000 *2
   with POLVIEWS dependent     .03185       .00496                .00000 *2
Uncertainty Coefficient :
   symmetric                   .04887       .00622     7.79608    .00000 *3
   with PARTYID  dependent     .04519       .00578     7.79608    .00000 *3
   with POLVIEWS dependent     .05322       .00674     7.79608    .00000 *3
Kappa                          .04859       .01431     4.97666

Kendall's Tau-b                .21148       .02163     9.71544
Kendall's Tau-c                .19744       .02032     9.71544
Gamma                          .26285       .02666     9.71544
Somers' D :
   symmetric                   .21114       .02160     9.71544
   with PARTYID  dependent     .22369       .02285     9.71544
   with POLVIEWS dependent     .19993       .02053     9.71544

Pearson's R                    .25929       .02645     9.98777    .00000
Spearman Correlation           .25949       .02655     9.99580    .00000
Eta :
   with PARTYID  dependent     .29150
   with POLVIEWS dependent     .28722


*1 Pearson chi-square probability
*2 Based on chi-square approximation
*3 Likelihood ratio chi-square probability

--------------------------------------------------------------
```

The chi-square statistics lead us to reject the null hypothesis of independence between PARTYID and POLVIEWS. The measures of association enable us to characterize the strength of association between PARTYID and POLVIEWS, but the question arises: are PARTYID and POLVIEWS nominal or are they ordinal variables? If both variables are ordinal, then you can interpret more of the measures shown. Both measures have probably been treated as ordinal by many researchers, but there is strong evidence that they should not be so treated! To show this, we employ homogeneity analysis.

Homogeneity analysis has been around for a long time under various names: homogeneity analysis, method of reciprocal averages, dual scaling, multiple correspondence analysis, or principal components analysis on an optimally quantified data matrix. The University of Leiden (Netherlands) Department of Data Theory has developed a program named HOMALS, which stands for HOMogeneity analysis by means of Alternating Least Squares. HOMALS makes the weak and easily met assumption that the variables in the analysis are measured at the nominal level--that is, the values of the variables represent unordered categories. Thus, HOMALS is suitable for data conventionally represented in multi-way contingency tables. HOMALS produces numeric results and plots. HOMALS attempts to separate the categories of a variable as much as possible. Like categories will be close and unlike categories will be far. What is more, HOMALS does this for all variables in the analysis, so that categories of different variables will be juxtaposed or separated, as the case may be. Examination of the plots from HOMALS reveals aspects of the relationship between the variables in the analysis.

Figure 4e shows the optimal scores for the variables given a two-dimensional solution.

Figure 4e Optimal scores for PARTYID and POLVIEWS

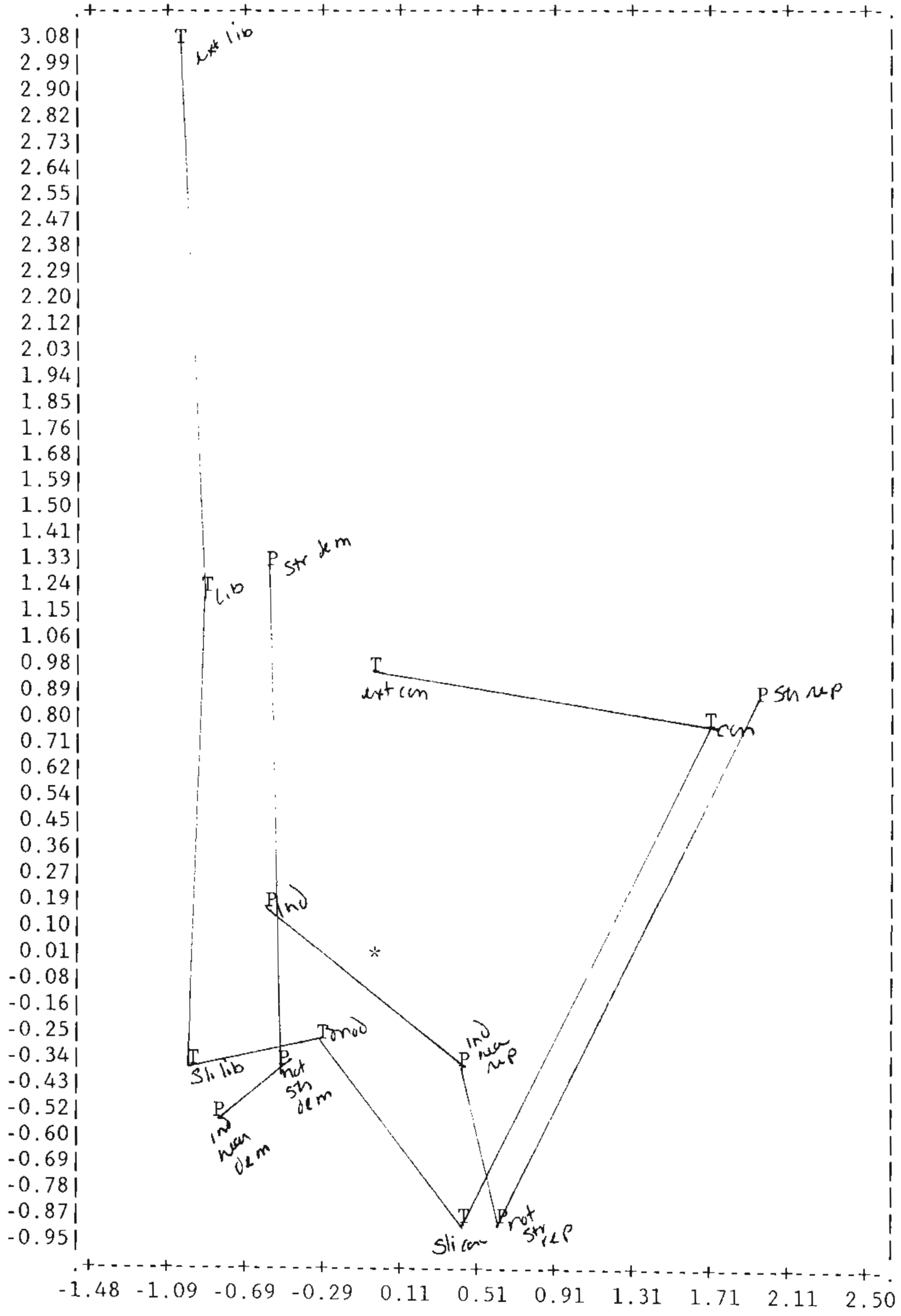
```
-------------------------------------------------------------

VARIABLE: PARTYID     POLITICAL PARTY AFFI
---------
                           MARGINAL    DIM.: 1       2
CATEGORY                  FREQUENCY    CATEGORY QUANTIFICATIONS
--------                  ---------    -----------------------

     1    STRONG DEMOCRAT       249          -0.51    1.34
     2    NOT STRONG DEMOC      271          -0.45   -0.35
     3    IND,NEAR DEM          204          -0.77   -0.52
     4    INDEPENDENT           142          -0.52    0.18
     5    IND,NEAR REP          155           0.48   -0.31
     6    NOT STRONG REPUB      242           0.65   -0.91
     7    STRONG REPUBLICA      123           2.02    0.90

          MISSING:                0

--------------------------------------------------------------------

VARIABLE: POLVIEWS    THINK OF SELF AS LIB
---------
                           MARGINAL    DIM.: 1       2
CATEGORY                  FREQUENCY    CATEGORY QUANTIFICATIONS
--------                  ---------    -----------------------

     1    EXTREMELY LIBERA       29          -1.01    3.08
     2    LIBERAL               130          -0.88    1.23
     3    SLIGHTLY LIBERAL      175          -0.94   -0.36
     4    MODERATE              557          -0.24   -0.25
     5    SLIGHTLY CONSERVA      272           0.46   -0.82
     6    CONSERVATIVE          183           1.73    0.76
     7    EXTRMLY CONSERVA       40           0.00    0.94

          MISSING:                0

-------------------------------------------------------------
```

Let us consider the first dimension category quantifications for both variables. The rank ordering of the dimension-one scores for PARTYID is 3-4-1-2-5-6-7. This suggests that PARTYID is not ordinal as scored in the General Social Survey, for if PARTYID had been ordinal, the dimension-one scores for PARTYID would have been monotonic with the GSS scores 1 to 7. In other words, while we often think in terms of the ordering Democratic-Independent-Republican, the dimension-one scores for PARTYID instead suggest the ordering Independent-Democratic-Republican. For POLVIEWS, the category quantifications are not in the same order as the GSS scores; in particular, the extreme conservatives are put between the moderates and the slight conservatives. We do not try to account for the ordering here, but you might speculate that we are seeing different types of conservatives: for example, libertarian conservatives, strong defense conservatives, and moral conservatives. If we have on hand variables indicating attitudes toward a strong state, strong defense, or toward various social items, we could include them in the analysis.

Let's turn to the dimension-two scores. Why are there a second set of scores? If you think by analogy with regression, a variable with many categories can be represented by a set of dummy variables. In similar fashion, a variable treated as nominal has multiple possible quantifications. For PARTYID the category quantifications contrast "strong" identifiers with "moderate" identifiers. The category quantifications for POLVIEWS similarly contrast the political extremes with the political middle.

Figure 4f dramatically suggests some groupings of interest between PARTYID groups and POLVIEWS groups.

Figure 4f Category quantifications in two-dimensional solution



The asterisk (*) shows the (0,0) origin. Homogeneity analysis has the property that larger groups are relatively closer to the origin while smaller groups are farther from the origin. It is useful to connect the category points for a variable with lines, as is done in the figure.

The plot suggests the following affinities:

1 strong republicans and conservatives;

2 not strong republicans, independent near republicans and slightly conservative;

3 independent, independent near democrat, not strong democrat and moderate,slightly liberal;

4 strong democrat and liberal;

5 extreme liberals off by themselves;

6 extreme conservatives off by themselves and somewhat strangely positioned.

It is possible, although we don't pursue it here, that a three-dimensional solution might fit the data better. For, in the two-dimensional solution, the category points are "flattened," that is, projected into the two-dimensional plane. A three-dimensional solution would let the category points "stretch into space." This might show, for example, that the extreme conservatives are really far away from the other categories.

5 Conclusion

Homogeneity analysis is one of a number of promising new techniques for analyzing sets of categorical variables. A major plus of the technique is that it is visually compelling. By comparison, what might the usual contingency table analysis or a loglinear approach have shown us about the relationship between PARTYID and POLVIEWS? In the usual contingency table analysis, you produce a many-way table in which it is difficult to see what is going on. Or, you produce a series of two-way tables which do not fully reveal what is going on in the multi-way table. If you instead apply loglinear models, you must first fit a model. For our two-way table, the model you fit must lie somewhere between the independence model and the saturated model. Upon fitting the model, you may have a many-parameters model in which it is difficult to interpret the signs and magnitudes of the parameters and relate them to the cells of the fitted table.

In conclusion, since so much of the data we typically analyze is categorical, and graphic approaches aid our understanding, the future of homogeneity analysis, correspondence analysis, and other techniques seems promising. If I have whetted your appetite for "nonlinear" multivariate analysis and you would like to read something on these techniques, you might read the following articles as well as references they cite.

Goodman, Leo A., "New Methods for Analyzing the Intrinsic Character of Qualitative Variables Using Cross-Classified Data," in American Journal of Sociology, Vol. 93, Number 3 (November 1987), 529-85

Hoffman, Donna L. and George R. Franke, "Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research," in Journal of Marketing Research, Vol. 23 (August 1986), 213-27

Hoffman, Donna L. and William D. Perreault, Jr. "Market Research: Consumer Preference and Perception," in Multidimensional Scaling: History, Theory, and Applications by Forrest W. Young and edited by Robert M. Hamer, 1987, Lawrence Erlbaum Associates

Perreault, William D. and Forrest W. Young, "Alternating Least Squares Optimal Scaling: Analysis of Nonmetric Data in Marketing Research," in Journal of Marketing Research, Vol. 17 (February 1980), 1-13

van der Heijden, Peter G. and Jan de Leeuw, "Correspondence Analysis used Complementary to Loglinear Analysis," in Psychometrika, Vol. 50, no. 4, 429-447

SURVEY RESEARCH SOFTWARE: FROM EXPERT SYSTEM SAMPLING THROUGH COMPUTER INTERVIEWING, DATA ANALYSIS AND PRESENTATION, TO PUBLICATION

Edwin H. Carpenter
University of Arizona

For presentation purposes let's assume that the survey research process can be represented in six stages (Figure 1).

These six stages do not include the first and critical step -- research problem formulation. However, they should serve the purpose of providing a framework for this presentation. The lower half of Figure 1 presents a brief outline of what I plan to discuss. First are the important considerations regarding the appropriate specification and execution of research endeavors. Then I will discuss current and future hardware and software capabilities. And lastly, I will provide brief glimpses at exemplary software packages that could be employed at each stage of the research process.

Survey researchers are generally familiar with the tasks that they must perform to accomplish a given research endeavor. I suspect that it is a rather rare individual that can specify all the procedures and details required to implement all the tasks that fall within these six general stages. For example, while the researcher may be quite adept at the procedures and details of data analysis, he or she may lack the training necessary for specifying such items as the sampling frame or the experimental design. Past solutions to the problem of incomplete or inappropriate research process specification were to:

1. Request that knowledgeable colleagues or experts provide the needed information free of charge.

2. Hire colleagues or consultants to provide the needed information.

3. Gloss over that stage of the process with an educated guess as to what is appropriate.

The latter is unacceptable for obvious reasons, while the former has attendant problems such as cost and getting colleagues to cooperate. Now there is another approach to solving the problem of providing expertise -- microcomputer software. The entire research process can now be advised by/or accomplished with available microcomputer software.

Hardware and Operating System Developments

Microcomputer hardware recently took a large leap toward being as powerful as mainframes. The IBM PS/2 Model 80 class machines, such as this Zenith 386 and the Apple Macintosh II, are both 32 bit architecture.

New operating systems like OS/2 for the IBM World and Multifinder for the Macintosh will allow the ease of use and the capabilities of research-related software to increase dramatically (Figure 2).

There is another factor that should guarantee that what I am suggesting will come to pass. It is called the free enterprise system. Developments in the computer world have accelerated dramatically since the development of the microcomputer. The cost of entry for development of both software and hardware is a fraction of what it was when mainframes were the only computers. As a consequence, the number of entrants has increased and their striving for a share of the market has provided the consumer with an incredible number of products that are increasingly easy to use.

Important Software Capabilities

The current crop of research-oriented software varies in capability and ease of use (Figure 3).

There are three fundamental requirements for good useful software:

1. A complete set of the intended capabilities must be provided.

2. The capabilities must be technically accurate.

3. It must be easy to use.

Let's take statistical software as an example of the progress being made in software capabilities. Technically accurate statistical software has been available since the mid to late 1960's when it might take all night to get a simple statistical job accomplished on a mainframe. In addition to being slow, it took a professional statistician to tell it how to accomplish the job. Nowadays, these same statistical problems can be solved on a microcomputer in a few minutes by people with a moderate amount of statistical training. In the not too distant future, expert-system-driven statistical software will allow virtually anyone with a minimum of training to do routine statistical analysis. And down the road a ways, artificially intelligent statistical software will allow the lay person to accomplish statistical analysis. I used statistics software as an example of where we are today, but it should be pointed out that other software that I will discuss may be at a different point on the continuum.

Sample or Experimental Designs

Let's assume that you have a research project that will require a random sample or experimental design. Some of the software products that are available are found on the upper-left quadrant of Figure 4.

I am going to show you the beginning frames from the expert system product called, "EXPERtIMENTAL DESIGN" -- A package that tells you what experimental design to use for the project that you are working on.

What I really want to convey is how the software can serve as an expert, and at the same time provide an interface that an nonexpert can understand and use to his or her advantage. You can get terms defined, or you can find out why the question was asked. You can even tell the program how certain you are of your response using the certainty scale. Depending on your answers either singly or in combination with the certainty score, the software decides the next appropriate question. The process goes on and the answer is provided regarding which experimental design is appropriate given the needs of the research project. You cannot see it, but there is an inference engine in the software that helps define the appropriate next question and decides on the appropriate outcome based on the responses you provide. Next, we will move quickly to another expert system package, but this time it is for calculating the size sample you need for your project based on a variety of inputs, including the types of analysis you plan to use and the number of variables that will be in that particular analysis.

This package, like the previous package, allows you to start from a previous stopping point, a welcome feature since there can be a lot of questions to answer before the results are provided. Also, this package, like the former, provides a printed output of the input information and the results and the flexibility to change input information and receive a new result. One of the most appealing features of both of these packages is the ability to get clarification of words or questions when needed by pressing the appropriate keys.

Data Collection

Among the various software packages that are available (see top right quadrant of Figure 4) for computer-aided interviewing, Ci2 (Computer Interactive Interviewing) provides one of the largest and most varied sets of capabilities and attributes. Among the capabilities is the ability to take results of the survey into several statistical software packages, now including SPSS/PC+. This recently announced addition fits nicely into this presentation since SPSS/PC+ is the software that I chose to demonstrate how microcomputer statistical software can aid the survey research endeavor. More on that in a moment.

Provision of the ability to link, to other software is an indication that the software vendors are concerned about users' ability to accomplish the task at hand; a task that spans the domain of both software companies. It is also an indication that both companies judge the other to be sufficiently important to the well being of the user -- and likely to stay in the business -- that it is worth the effort to implement the linking step.

Data Entry, Cleanup and Transfer

If the data to be analyzed arrive on paper, like a mail survey, or by phone link to another computer, then data entry, cleanup, and transfer software is needed. If you find yourself in this situation, then one of the software packages shown in the bottom left quadrant of Figure 4 may be what you need. The most capable package for entry and/or cleanup is

SPSS/PC+DATA ENTRY II. The key is its ability to apply simple tests to verify variable values and to apply complex cleaning rules to relationships among variables. For example, a check can be made for sex, pregnancy status, and age that raises a flag when certain conditions are met, like pregnancy at age 82 or male pregnancy. The rules can be applied in an interactive session, or a report can be generated so that flagged records can be identified and handled appropriately.

Several packages can transfer data and variable names and labels. Stat/Transfer provides this one basic and useful function, if you want to move files among SPSS/PC+, SPSSx, SYSTAT, STATA, GAUSS, LOTUS, dBASE and other programs that use the same data file structures.

Data Transformation and Management

The worst task in survey research is data transformation and management. It is a tedious and thankless task that must be done. Most statistical software packages have routine capabilities available. However, if you need data aggregation or similar advanced capabilities, then one of the advanced packages listed in the bottom right quadrant of Figure 4 will help. These packages have data transformation and management capabilities to accomplish the task set forth in the following example: Three years of data on more or less the same 1400 clients are available in four data files: two for the first year, when people signed up for the program at two different research installations; one for the first follow-up year; and one for the third year, when there was a possibility that some clients had been seen up to 40 times, with a separate record for each visit. Fortunately, each client has a unique identification number; but unfortunately, each client also has somewhere between 1 and 43 records full of information. The data need to be in one file for analysis to proceed. To accomplish the task, case-sorting, file-appending, nonparallel file-merging, record-aggregation, and record-mapping capabilities are required. The two data files from the first year are joined together into one file, which becomes the base data file for the 1400 cases. The second-year file, with 1340 cases, is merged into the base file, and missing data codes are given to variables for the 60 cases that do not exist in the second-year file. Data from the third year are subjected to the aggregate procedure, which produces one record containing summary data for each of the variables collected in the third year for each client. A mapping of each record is produced so that all the actions above can be checked. With capable data transformation and management software in hand, what could have been a frustrating task becomes a tractable one.

Data Analysis

Now let's go back to the data file created by Ci2 and take it across to the SPSS/PC+ analysis package. I will start the SPSS/PC+VERSION 2 software. I am going to use the menu mode, which is one of the three modes available, to provide the commands necessary to run the software. The other two modes are batch and interactive. One of the hallmarks of a user-friendly software package is the ability to choose how you would like to interact with the package. For example, if you use the package on a regular basis, then batch or interactive mode is probably the bast since you can remember the commands. If you use the package infrequently and do not remember all the commands, then menu mode is quite handy. However, to be forced to use a menu each time you use the package becomes quite tedious and burdensome.

As you can see, the screen is divided into three parts: a list of topics is in the top left box, to the right is information about the topic highlighted in the left box, and below is the area where job control commands are to be written. We move through the menu to the current topic and press return which causes a new menu of only the appropriate choices to appear. As this happens the information in the right box changes to provide up-to-date information about the topic highlighted in the left box. This process continues till the job control commands are written in the lower part of the screen. The commands can be saved to a file for later use or they can be executed, or both. The beauty of this type of interface is that the user is prompted with only that set of commands, and the information about those commands, that are needed to accomplish a given objective. My experience suggests that using the manual will become a thing of the past, except for examples of how to interpret findings from a specific analysis procedure. One other point -- if you ask for help, the help you receive will be keyed to the place in the command list that you asked for help. This is known as context-sensitive help.

Now we will instruct the SPSS/PC+ software to get the data file prepared within the Ci2 package (the file was prepared by Ci2 for immediate use by SPSS/PC+) and then we will do a frequency analysis of responses to some of the questions. I should add, as an aside, the statistical software listed at the top left of Figure 5 is software that provides a reasonably full complement of statistical routines. As you can see, the bulk of what is available is for the IBM world as compared to the Apple Macintosh world.

Presentation of Results

The results must be presented. There is no reason to expect anything less than high-quality text and graphical presentation to suffice for findings so diligently sought. The top right portion of Figure 5 shows capabilities available in the statistical packages. What I will show you comes from the SPSS/PC+ package and NUMBER CRUNCHER.

When you purchase SPSS/PC+ Version 2, you also receive a handy presentation package called GRAPH-IN-THE-BOX. Use of this package is integrated into SPSS/PC+ as FASTGRAPH under the menu heading "graphics." Use of this capability allows you to select or capture numbers from your screen and graph them instantly. For this example, we will capture the frequencies of one of the questions that came from Ci2 and provide an instant graph. This bar graph can be changed into a line graph with a couple of key strokes. Also, GRAPH-IN-THE-BOX can be used independently of SPSS as a "terminate-and-stay" resident package, which means it resides out in memory until needed and is available instantly when summoned from within any application software.

Now let's move to what is perhaps the most spectacular graphical display of data that is currently available. It is provided in NUMBER CRUNCHER. We just finished looking at a two dimensional display, so now let's take a look at a three dimensional display and then let's rotate it around both the "X" and "Y" axis. Let's not stop there, how about a five dimensional display of data? What I will show you are five test scores for sixteen different people. The five scores are presented by Chernoffs' faces. Each test score is assigned to a facial characteristic, such as the eyes or nose or mouth. As you can see, it is pretty easy to tell who made good scores on all the tests and who made bad scores. I really do not know to what extent Chernoffs' faces have gained acceptance as a legitimate way of presenting data, but I do know that we all are capable of reading faces. In fact, it was probably one of the first things we did in life and continue to get better at throughout life. Perhaps one day you will evaluate what city to live in based on a facial expression representing the various quality of life attributes of city life such as crime, pollution, unemployment, etc.

Desktop Publishing

Quickly, I would like to share my experience trying to cause the graphics produced from the statistical packages to go into and out of PAGEMAKER. At the bottom of Figure 5 are the packages that will, or soon will, interface with PAGEMAKER. Overheads that are reproduced in Figure 6 show some examples of graphics that were produced through PAGEMAKER.

The Survey Researcher Reigns

I have not spent much time discussing the impact that the Apple Macintosh and related software is going to have on our future well being (Figure 7). I suspect that we will see - in fact we are already seeing - evidence that we will have benefit from the competition. This is evident in the graphical user interface, high-powered hardware, open architecture, and vertical links to mainframes. Competition amongst software developers will force advances that will employ expert systems and artificial intelligence to come increasingly into use, especially when that will make the software easier to use. The upshot of it all is that the survey researcher will be able to do a better job at a lower cost, faster, and do it "in-house."

Bibliography

Archer, J., & King, A. (1987, June 30). IBM and MICROSOFT Set the Record Straight. PC Week, pp. 43-50.

Barcikowski, R.S. (1986). SPSS/PC+. The American Statistician, 40, 225-227.
Carpenter, E.H. (1987). The Evolving Statistics and Research Process Using Microcomputer Statistical Software. Social Science Microcomputer Review, 5(4), 529-545.

Carpenter, E.H. (1985a). The Future For "User Friendly" Microcomputer Statistical Software: SPSS/PC and SPSS/PC+ as Examples. Social Science Microcomputer Review, 3, 352-362.

Carpenter, E.H. (1985b). Statistical Packages for Microcomputers: New Vistas for Social Science Researchers in 1985. Social Science Microcomputer Review, 3, 183,204.

Faust, M. (1987, June 30). The Forecast Looks Bright for Business Planning Packages. PC Week, pp. 53-64.

Fridlund, A.J. (1986, September 1). Statistics Software, Infoworld, pp.31-39.

Garson, G.D. (1987). Academic Microcomputing: A Resource Guide. Beverly Hills, CA: Sage.

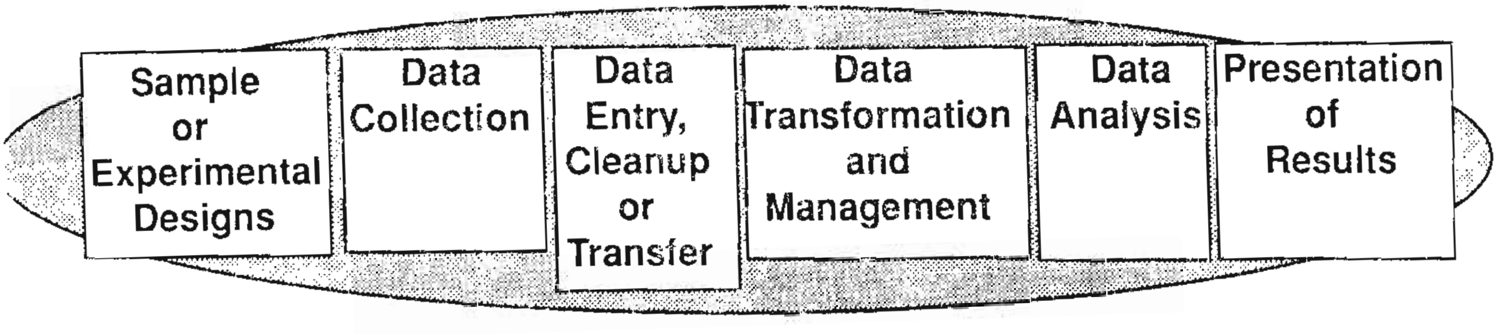
LePage, R. (1987). STATGRAPHICS 2.0 Statistical Graphics System. The American Statistician, 41, 64-67.

McKenzie, J.D. Jr. (1985). NUMBER CRUNCHER STATISTICAL SYSTEM (NCSS). The American Statistician, 39, 315-318.

Turner, D. L. (1987). STAT and STATA GRAPHICS. The American Statistician, 41, 68-70.

Figure 1

Stages of the Research Process



Sample or Experimental Designs

Data Collection

Data Entry, Cleanup or Transfer

Data Transformation and Management

Data Analysis

Presentation of Results

PRESENTATION OUTLINE

1. Considerations regarding appropriate specification and execution of research endeavors.
2. Hardware and software capabilities - now and in the future.
3. A brief glimpse at exemplary software packages.

Figure 2

HARDWARE AND OPERATING SYSTEM DEVELOPMENTS

IBM WORLD

HARDWARE -- 80386 class machines, IBM PS/2 Model 80, COMPAC 386, Zenith 386.

OPERATING SYSTEMS -- OS/2, Presentation Manager, Windows 386, DesqView 386, DOS 3.3 and EMS 4.0.

APPLE WORLD

HARDWARE -- 68020 class machine, Macintosh II.

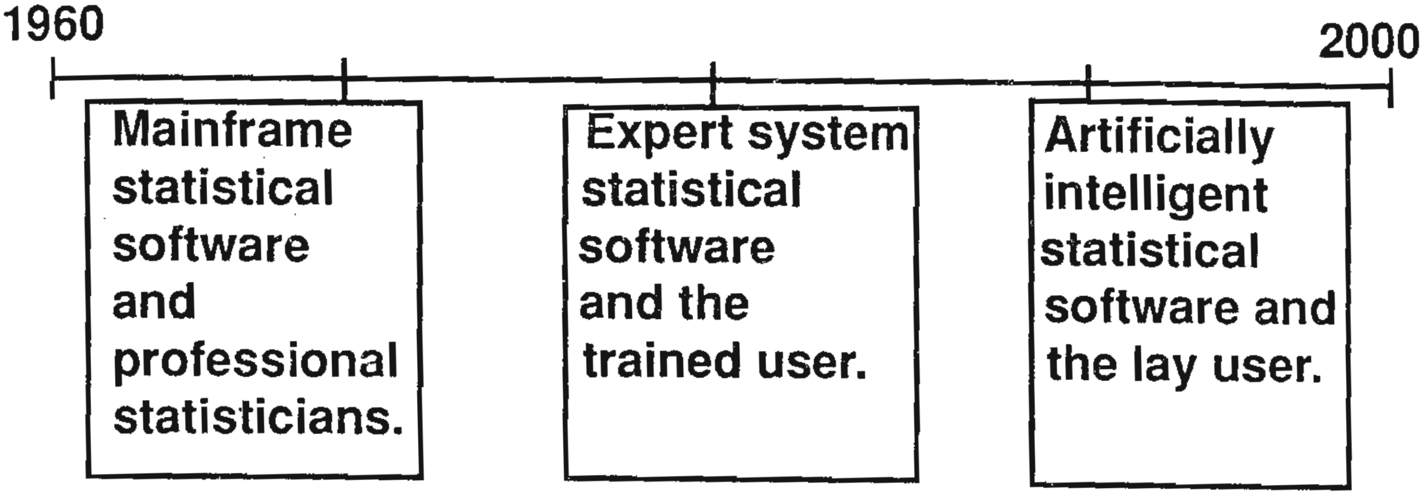
OPERATING SYSTEM -- Multifinder.

Figure 3

Important Software Capabilities

Three fundamental requirements for useful software.

1. Complete set of capabilities.
2. Technically accurate.
3. Easy to use.



1960

2000

Mainframe statistical software and professional statisticians.

Expert system statistical software and the trained user.

Artificially intelligent statistical software and the lay user.

Figure 4

SAMPLE OR EXPERIMENTAL DESIGN SOFTWARE PACKAGES

SAMPLE DESIGN

EX-SAMPLE
STATPAC GOLD
SYSTAT'S DESIGN
STATGRAPHICS

EXPERIMENTAL DESIGN

EXPERIMENTAL DESIGN
STATGRAPHICS
SYSTAT DESIGN

DATA COLLECTION

SOFTWARE PACKAGES

CAPPA
Ci2
Ci2 CATI
Q-FAST
QUIZ WHIZ
MATI
MaCATI

DATA ENTRY, CLEANUP AND TRANSFER

ENTRY

COSTAT
With Rules
SPSS/PC+ V.2
DATA ENTRY II
with verification
STATPAC GOLD

CLEANUP
With Rules

COSTAT
CRUNCH
SPSS/PC+ V.2
DATA ENTRY II
STATPAC GOLD

TRANSFER
Beyond Just ASCII

ABSTAT
SAS
SPSS/PC+ V.2
TRANLATE
STATGRAPHICS
STAT TRANSFER

DATA TRANSFORMATION and MANAGEMENT

MOST PACKAGES HAVE ROUTINE CAPABILITIES

Advanced Capabilities Like Aggregate

BMD/PC
CRUNCH
SAS
SPSS/PC+V.2
STATPAC GOLD

Figure 5

DATA ANALYSIS

IBM WORLD (PC & MS DOS)

- ABSTAT
- BMD PC
- COSTAT
- CRUNCH
- CSS
- MINITAB
- NUMBER CRUNCHER
- NWA STATPAK
- SAS
- SIGSTAT
- SPSS/PC+
- STATA
- STAT80
- STATGRAPHICS
- STATPAC GOLD
- SYSTAT

APPLE WORLD (MACINTOSH)

- SYSTAT
- STATVIEW 512+
- STAT80
- NWA STATPAK

PRESENTATION OF RESULTS

MAPS

SPSS/PC+ V 2.0
and MAP MASTER

STATGRAPHICS
and ATLAS GRAPHICS

3D GRAPHICS

- COSTAT
- NUMBER CRUNCHER
- SPSS/PC+V.2
  and Microsoft CHART
- STATA
- STATGRAPHICS
- STATVIEW 512 (MACITOSH)
- SYSTAT (MACINTOSH)

DESKTOP PUBLISHING

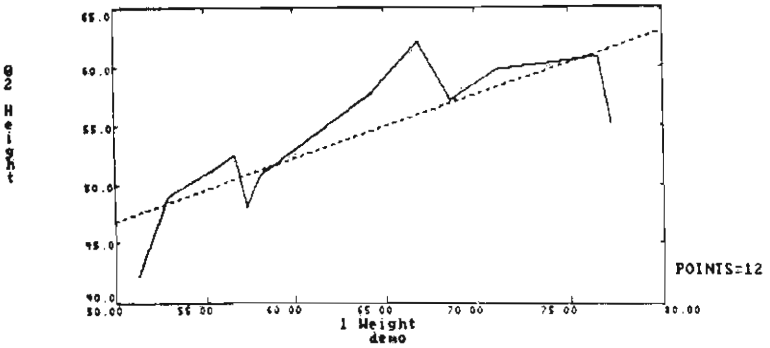
INTERFACES WITH PAGEMAKER

(Beyond ASCII)

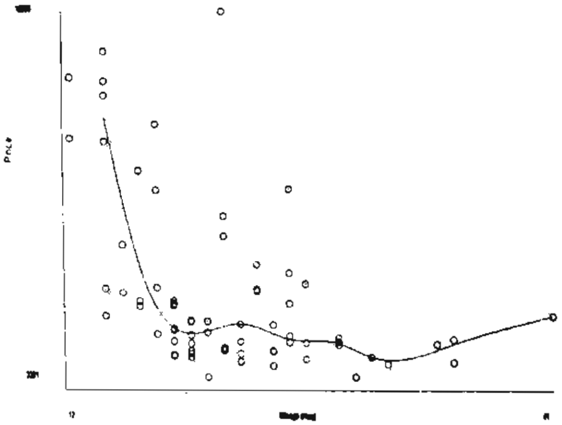
- ABSTAT
- CRUNCH (Near Future)
- NWA STATPAK
- SPSS/PC+V.2
- STATA
- STATGRAPHICS (Near Future)
- STATVIEW 512+ (MACINTOSH)
- SYSTAT (IBM) (Near Future)
- SYSTAT (MACINTOSH)

Figure 6

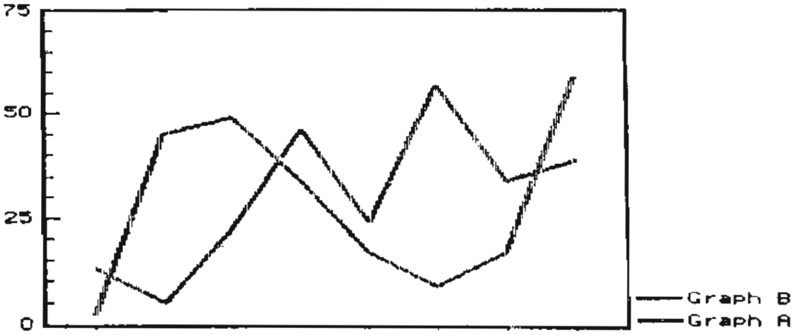
Example of ABSTAT (screen captured) file



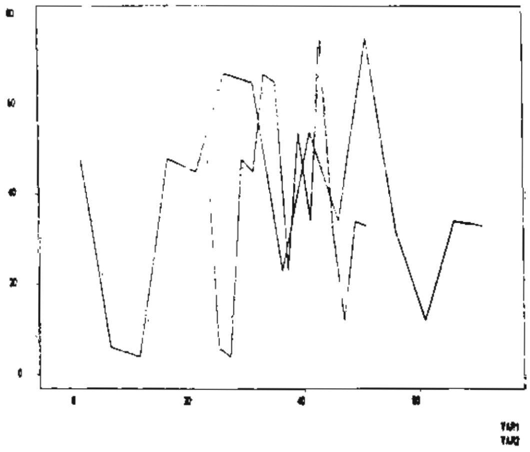
Example of STATA file



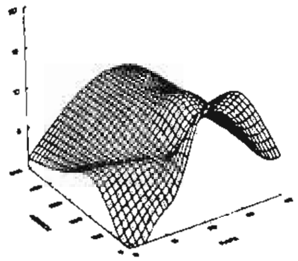
Example of SPSS/PC+ and GRAPH-IN-THE-BOX (screen captured) file



Example of NWA STATPAK file



Example of SYSTAT (MACINTOSH) file



Example of STATVIEW 512+ (MACINTOSH) file

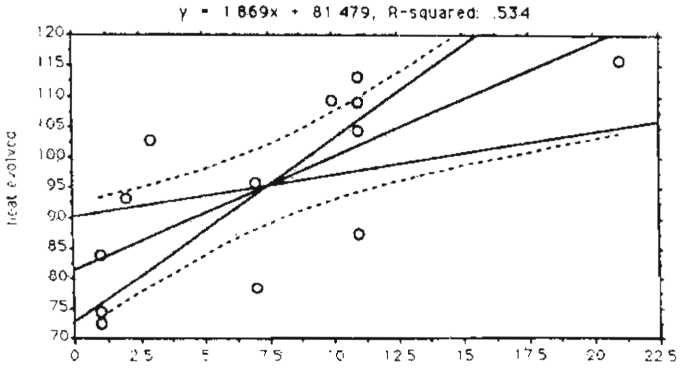
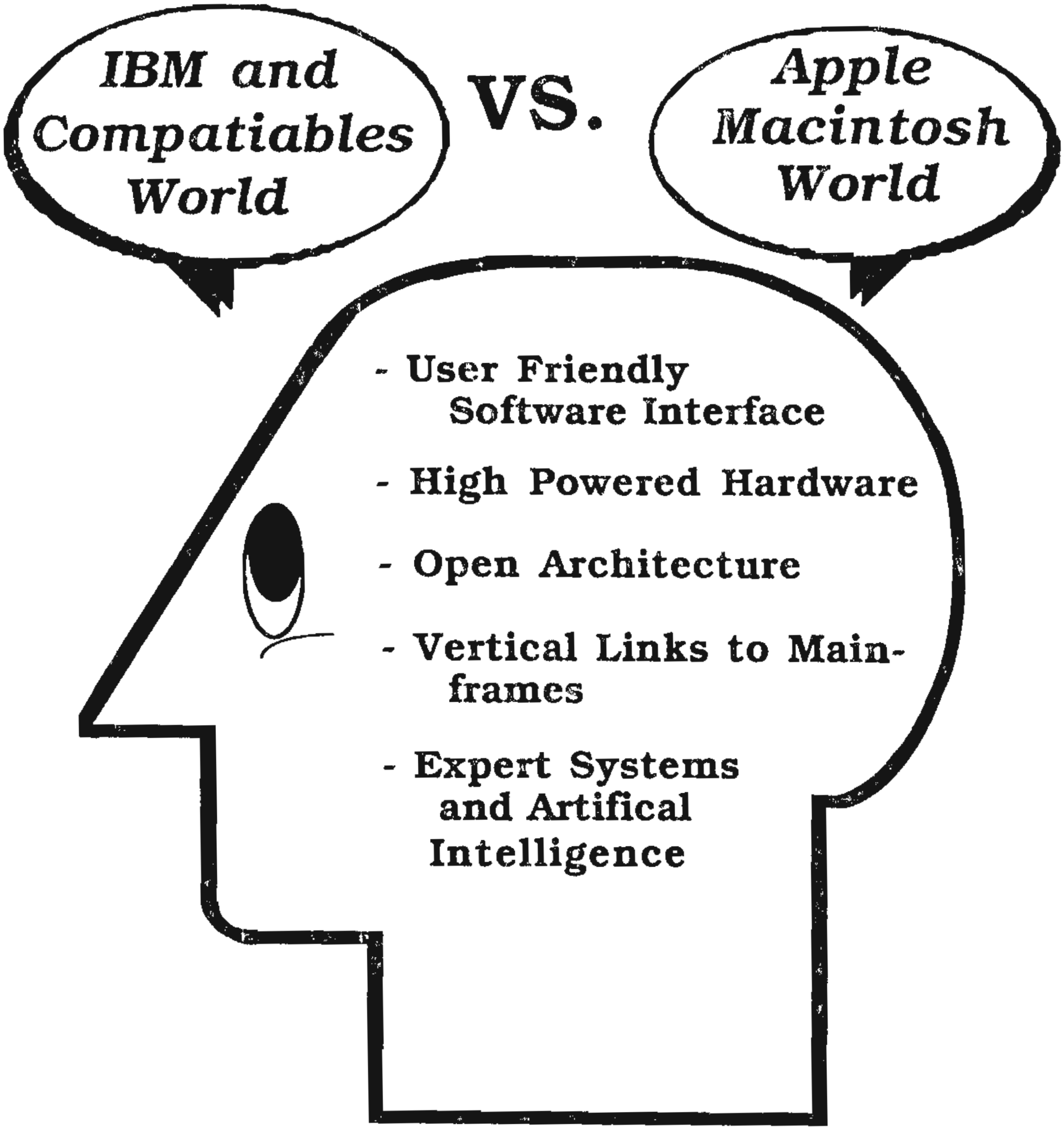


Figure 7

THE RESEARCHER REIGNS

CONCLUDING THOUGHTS



- User Friendly Software Interface
- High Powered Hardware
- Open Architecture
- Vertical Links to Mainframes
- Expert Systems and Artifical Intelligence