Sawtooth Software

RESEARCH PAPER SERIES

Conjoint Analysis: How We Got Here and Where We Are (An Update)

Joel Huber,
Duke University

530 W. Fir St.
Sequim, WA 98382
(360) 681-2300
www.sawtoothsoftware.com

Conjoint Analysis: How We Got Here and Where We Are—An Update

Joel Huber
Duke University

This paper was originally published in the 1987 Sawtooth Software Proceedings. Because of the excellent quality of this work, we asked Joel to present this paper again at the 2004 Sawtooth Software Conference. We have published the original paper below, but with added footnotes, given a 2004 perspective. –Bryan Orme, Editor.

Conjoint analysis has had a profound effect on the conduct of research in many facets of business, particularly in the areas of product positioning and new product development. It is a field approaching the maturity stage of its life cycle. However, with the coming of inexpensive, user-friendly programs for conjoint, we can expect its use to increase substantially. Indeed, we will soon see the day when virtually all market research firms will offer conjoint studies as part of their standard repertoire. Managers will use conjoint not just for special projects, but as an indispensable tool enabling them to test the impact of proposed actions on the market. Conjoint is becoming less elite, its secrets no longer the property of a few, but available in its simpler versions to all.[1]

Today I would like to present my personal perspective on the history of conjoint analysis. The field is shaped by two fundamentally conflicting forces. First, there are the idealistic psychometric forces that started the field. Opposing these, while at the same time arising from them, are the pragmatic forces, practitioners who have determined the way conjoint is used. The tension between these forces has shaped the growth of the field and will continue to guide its future development.

The Psychometric Tradition

The term "conjoint[2]" has itself contributed to the mystery of the field. The term arose out of an attempt to apply extensive measurement to preference judgments. Extensive measurement refers to a method to build a scale by comparing relative lengths (extensions) of objects. For example, by comparing the lengths of different rods put end to end, one can form a scale on which it is appropriate to perform such operations as addition and subtraction. While such interval-level scales are relatively easy to generate from physical quanta such as weight, size and time, they have been notoriously difficult in the case of human preferences.

The difficulty arises because we know what it means to say that we like potatoes better than

---

[1] Indeed, the use of conjoint analysis has dramatically increased. Based on a 2004 Sawtooth Software customer survey, we project that between 5,000 to 8,000 conjoint analysis projects were conducted by Sawtooth Software users during the previous 12-month period. The relative proportion of projects by conjoint flavor among our users was CBC (61%), ACA (27%) and CVA (12%).
[2] Many researchers today believe, incorrectly, that the term conjoint comes from the idea that respondents are asked to "CONsider features JOINTly." Conjoint means "to join or become joined together," as in "conjoined twins."

rutabagas, but generally not what it means to say that our liking for potatoes over rutabagas is greater than our liking for artichokes over eggplant. This indeterminacy poses a problem to psychometricians who want the same solid base for measuring the psyche as physicists had for measuring weight. Without interval scales of preferences, it is difficult to specify what it means to have an additive model of preference.

The psychometricians reasoned that while our ordinary language pronouncements of preferences do not directly produce interval scales, certain kinds of preference judgments had to be based on utility values that do. One set of preference judgments that requires metric underpinnings refers to compound or conjoint objects.

Consider the statement that one prefers a $10,000 convertible to an $8,000 sedan. This statement implies that the benefit of a convertible over a sedan is greater than $2,000. Psychometricians were able to show that by putting together a number of such preference statements, it is possible to derive intervally scaled additive partworth utilities that could underlie these preferences. Further, they specified a number of tests to determine if such an interval scale is justified, given the preference orderings.

Conjoint measurement provided a theory for creating a measurement scale from judgments on compound or conjoint objects. It generated a great deal of excitement when first proposed. Conjoint was a "psychometric conjurer's stone"—a way to transform the dross of ordinal preferences into the gold of interval scales. At last the measurement of preferences might be put on par with measurement in the exact sciences.

The early contributions focused on finding sets of elegant axioms and/or conditions required to uncover the latent interval partworths. Some of these conditions, such as independence, are well known, while others, such as double cancellation, are less well known. The axiomatizations are best summarized in the classic Foundations of Measurement, Volume 1 (1971) by Krantz, Luce, Suppes, and Tversky. In the preface to that volume, reference is made to Volume 2 on applications. It is ironic and significant that that volume has not yet been published[3].

What happened? As soon as the psychometricians applied their models to human behavior, they found that the axioms were consistently violated. It was very similar to what is now occurring with respect to the Von Neumann-Morgenstern axiomatization of choice under uncertainty (Thaler, 1985). Virtually all the axioms were violated in relatively minor but systematic ways. Initially, it appeared that random error could account for the intransitivities and lack of additivity found. However, as more elegant and precise tests were devised, this escape was also blocked (e.g., see Falmange, 1976).

In hindsight, it is not surprising that if people cannot give consistent interval partworth values directly, then such metric rigor is unlikely to be hidden beneath more complex judgments on

[3] Since then, a follow-up volume did appear: Foundations of Measurement: Geometrical, Threshold, and Probabilistic Representations (Foundations of Measurement) by Patrick Suppes, David M. Krantz, R. Duncan Luce, Amos Tversky, Academic Press 1989.

conjunctive stimuli. There is no intervally scaled ruler hidden in the brain that can account for complex preference judgments.

Still, the psychometricians provided a clear and coherent tradition, aspects of which are still important today. That tradition includes the following components:

> First, the belief that individual preferences can be expressed in numerical terms that lead to behavior.
>
> Second, the focus on comparisons among conjunctive stimuli, defined on multiple attributes, so that the response requires trading off high levels on one attribute with low levels of others.
>
> Third, the tradition of using factorial designs in which the attributes to be tested are statistically independent of one another.
>
> Fourth, the emphasis on testing the assumptions, such as additivity, as a prior condition to estimating the partworth utilities.
>
> Finally, the orientation to ordinal responses from subjects as the primitive behavior being modeled, rather than direct magnitude or interval scales[4].

From Psychometric Swords to Market Researcher's Plowshares

The psychometric tradition is rigorous and idealistic, whereas its adoption by the market research community has been approximate and pragmatic. The market research community began with the same rigorous models, but soon found that the partworth utilities were managerially very useful despite the fact that the tests did not work. In effect, the operation failed, but the patient thrived. Useful aspects of the original conjoint measurement framework were adapted and less useful ones were dropped.

Rich Johnson's succession of conjoint models is perhaps most illustrative of the changes that occurred. Rich was trained as a psychometrician, and his original trade-off analysis used 3-by-3 trade-off matrices, in which respondents were to rank order alternatives defined on various levels of the two attributes (Johnson, 1974). Then, by computerizing the approach, he was able to avoid certain redundant questions and speed up the task. However, the price of this additional speed was a lessened ability to make consistency tests at the individual level. His next step expanded the task from one of categorical preferences to graded-pair comparisons. This permitted more information to be collected from respondents with very little additional cost in time or effort. Finally, he used the personal computer to merge direct attribute judgments with paired

[4] It is interesting to note that we have come full circle. The early researchers in conjoint measurement favored ordinal (ranked) data, rather than ratings judgments. In the late 70s and throughout the 1980s, researchers tended to favor the increased information (and ease of data collection and analysis) provided by ratings scales. Now, in 2004, interest has shifted back to non-metric scaling, this time embodied in choice (as in CBC).

comparisons and guide the selection of "optimal" pairs during the conjoint task[5]. All these changes helped to obtain information from respondents more efficiently and to formulate a better predictive model of their preferences. Still, these changes represent a substantial departure from the psychometric tradition.

Much of the ambivalence between idealism and practice in the marketing research community is found in Green and Srinivasan's (1978) classic review article on conjoint. In that article they differentiate conjoint analysis from the older conjoint measurement in order to make appropriate separation between the two fields. A dualism is evident in their discussion of the various ways to perform conjoint analysis, sometimes focusing on what is theoretically justified, while at other times succumbing to practical reality.

What did the marketing research community take from the psychometricians and what did they change? Generally, the trends are evident in Cattin and Wittink's (1987) review of practices in conjoint. First, the field continues to consider behavior as captured by partworth utilities and simple additive models[6].

> Second, in keeping with the psychometric tradition, they use compound stimuli which force individuals to trade off conflicting attribute levels.
>
> Third, they still rely on orthogonal arrays[7], although highly fractionated designs have replaced the original full factorials.

The first three components of the psychometric tradition have been passed down relatively unchanged. The last two, the structural tests and the nonmetric orientation, have rapidly eroded[8].

Consider, first, the tests of the structural composition rules (such as additivity) that were the major focus of the axiomatic systems. These tests are now virtually ignored, or worse, assumed away. Consider, for example, the common use of fractional main-effects design. While these offer far greater efficiency and permit main-effects estimates of many more attributes, they assume that interactions are zero. If there are interactions, the preference function will be biased or wrong. Further, because no test is possible, the analyst will never know that the results are

---

[5] Joel is referring to Rich Johnson’s Adaptive Conjoint Analysis (ACA) product, which went on to become the most widely-used conjoint method and software system for conjoint analysis is the late 1980s and throughout the 1990s.
[6] Interestingly enough, with respect to model specification, we have again come full circle. Early conjoint researchers favored parsimonious models that featured part-worth estimation at the individual level. Through the 1980s and 1990s, there was increased interest in linear estimation of terms, higher-order effects, and (largely due to necessity) pooled estimation from choice models. With the advent of HB analysis, evidence is in favor of simple main-effects models, with part-worth representations of attributes’ utilities.
[7] Largely due to efforts by Warren Kuhfeld (and co-authors), researchers today have recognized that strict orthogonal arrays are often not optimal (especially in cases involving asymmetric designs). Computerized searches that sacrifice some orthogonality in favor of increased level balance are able to achieve overall greater precision of estimates.
[8] Again, see the discussion in footnote 3 regarding the recent strong renewal of interest in non-metric data (choice).

biased[9].

The other major deviation from the psychometric tradition has been a move from a nonmetric to a metric orientation. This has occurred both in the kinds of data collected from respondents and routines used to analyze it. The original reason to use rank-order inputs over quasi-metric ones stemmed from a legitimate uncertainty about what respondents meant in responding to, for instance, a ten-point strength-of-preference scale. A number of nonmetric analysis packages, such as Kruskal's (1965) MONANOVA and Srinivasan's and Shocker's (1973) LINMAP permitted relatively easy analysis of input data about which only ordinal properties could be assumed. The shift from ordinal to metric inputs has been largely pragmatic. For example, putting 25 profiles onto a ten-category sort board is both easier for subjects and provides more reliable inputs than an exhaustive rank-order task. Using a rating scale permits one to generate predicted choices with equivalent reliability but fewer judgments.

Metric methods have also become more popular as methods of analysis. This shift is in part due to the ease of use of ordinary least squares over nonmetric routines. But it also stems from a realization of the value of the weak but errorful metric information in a rating or a sort-board task. Of course, such data can be analyzed by nonmetric procedures, such as monotone regression or LINMAP. The nonmetric procedures find a monotone transformation of the dependent variable that best fits the model. With such routines this transformation is very linear, indicating that the transformation provides little additional information. More significantly, the monotone transformation generally degrades the predictive ability of the model (e.g., see Huber, 1975). Quasi-metric data has some interval properties that nonmetric routines treat as noise, but metric routines are able to use. Nonmetric routines may be losing popularity simply because they do not appear to help predictions. However, this is simply a pragmatic criterion; we lack a good theoretical reason why metric routines should work better. Indeed, most theoretical considerations lead to the championing of a nonmetric orientation[10].

To summarize, the marketing research industry has taken some of the external trappings of conjoint measurement, but has generally deleted or modified its elegant inner workings—those dealing with hypothesis testing on strictly ordinal inputs. While these modifications may not, in a strictly predictive sense, matter, it is appropriate to inquire as to the kind of offspring that has emerged from this union of pragmatic and idealistic parents.

What Really Goes on During a Conjoint Exercise?

If we are not tapping a latent interval utility scale during conjoint, what are we doing? The answer depends on whether the stimuli are unitary or decomposable. Unitary stimuli are those

[9] Lately, conjoint analysts have come to appreciate that many interaction effects are actually due to heterogeneity, and individual-level part-worths used in choice simulators can reflect a variety of interaction effects (through sensitivity analysis), even though the effects were not directly specified in the model.

[10] Again, the orientation toward metric conjoint has waned, and researchers are using choice-based conjoint flavors more often now than metric responses. Even if using metric responses, attention has turned from OLS to HB analysis.

which respondents cannot easily break into component parts, such as foods, scents, or esthetic objects. With these kinds of stimuli, response to factorially designed stimuli is holistic and generally frustrates attempts to build simple models of that response. Main-effects designs are at a particular disadvantage in that the assumption of additivity is usually violated. Specialized designs are appropriate if the source of the interactions can be localized within a few variables, and if there is not too much heterogeneity across subjects, conditions that are sadly not often satisfied.

Because of the design problems with unitary stimuli, most conjoint analysis uses readily decomposable attributes and displays attributes in ways that make it easy for respondents to separate them. Respondents are given repeated questions with predictably arrayed attributes, and anyone who has watched a conjoint exercise knows the result. Respondents simplify the task by focusing on a few attributes. Each profile is evaluated by scanning these attributes and adjusting the valuation of the alternative accordingly. This process results in a very good fit of the additive model at the individual level. Typically, a small number of attributes are strongly significant and the rest are nonsignificant. Interactions are very rare—they require extra processing. Different tasks produce slightly different patterns of responses. For example, paired comparisons may produce more significant attributes. However, the general pattern of a strongly simplified evaluation strategy that has accurately captured an additive model emerges regardless of the particular task.

Evidence for such a simplification process appears in an anomalous result which I have found in my work, and I suspect many of you have as well. Following a conjoint exercise we often include a choice task, either using actual brands or profiles where the attributes have been scrambled to reduce the likelihood that the conjoint choice process will be trivially repeated. In these studies it is possible to note the correspondence between the internal fit of the conjoint and its accuracy in predicting the holdout choices. If the conjoint task is a correct representation of the holdout choices, then the better the fit of the conjoint and the more it satisfies the axioms, the better it should predict the holdouts. In other words, respondents who are well modeled by the conjoint should be ones we can best predict. However, my experience has not been consistent with this expectation. Correlations between conjoint fit and predictive accuracy are very low and often negative, particularly if one screens out the totally random subjects. How could this be? The simplest account focuses on differences in the conscientiousness of respondents. The conscientious ones try to consider as many attributes as possible in their conjoint task. Being mortal, they make mistakes, resulting in preference reversals and greater error levels. When they make the holdout choices, they conscientiously try to be consistent with their earlier judgments, resulting in greater predictive accuracy. For the less conscientious respondents the reverse holds. They simplify the conjoint task greatly in order to get through it, focusing on one or two attributes. This simplification permits remarkable fits to be the additive conjoint model. In the holdout choice task with distorted or scrambled attributes, using the same simplified decision rule is not easy. These less conscientious respondents shift strategies, basing their choices on different attributes than in the conjoint, resulting in a poor correspondence between the two.

The point is not to resolve this anomaly, but to raise an issue about conjoint. The professional success of conjoint practitioners attests to how well it works. It produces intuitively pleasing results that managers find very useful, although we do not have a clear account of why it works. Indeed, its problems could lead many to discard it. After all, it has shed its theoretical roots and appears only to capture a simplified and truncated version of choice behavior[11]. The next section considers why conjoint works, and this leads naturally into ways in which it might be improved.

Why Conjoint Works

1) Conjoint requires tradeoffs that are similar to those in the market.

A conjoint task is valuable because it forces the respondent to evaluate conflicting attributes, as between the type of car and its price. People typically try to avoid making such judgments by searching for unambiguous solutions involving dominance or following a relatively clear heuristic. However, the marketplace also requires such judgments and people make them when they must in the marketplace or the conjoint task.

The conjoint task, in which alternatives are compared on a number of dimensions, can be usefully contrasted with a direct elicitation approach, in which attribute utilities are directly assessed. There are two problems with direct elicitation that are ameliorated with conjoint. First, it is difficult with the direct elicitation approach to keep a respondent from seeing everything as important. Certainly, $2,000 is very important in selecting a car, but it may not be more important than the difference between a convertible and a sedan. Second, direct elicitation does not directly relate to a choice in the marketplace[12], but is a summary measure of those behavioral decisions. In contrast, the conjoint task is more directly analogous to market choice.

2) The simplification in conjoint may mirror that in the market.

---

[11] With the resurgence of interest and application of the non-metric tradition through choice, the conjoint analysis applied in 2004 is truer to its theoretical roots. Rather than use ratings to predict choices, we employ choices to predict choices. Still, true to Joel's assertions, our modern CBC questionnaires still reflect a simplified (truncated) representation of real-world choice behavior.

[12] Joel discounts the predictive validity of direct elicitation (self-explicated) methods. The debate continues nearly 20 years later, though the vast majority of the industry still favors tradeoff methods. Some well-known researchers have published papers defending the predictive validity of self-explicated methods, placing it on par or in some cases above that of conjoint methods. Improvements in the manner of collecting self-explicated judgments have led to its increased performance. For example, self-explicated judgments now require respondents to trade-off one attribute for another by techniques such as a constant-sum scale.

The simplification found in conjoint to a small number of attributes is only misleading if there is a very different kind of simplification in the marketplace. There is evidence that the decisions in the market are based on remarkably few dimensions (Olshavsky and Granbois, 1979). If so, then conjoint may indicate those few attributes on which the consumer bases his or her decisions.

Further, to the extent that conjoint is used to predict aggregate shares, it does not matter that an individual's selection of attributes is unstable over time. As long as conjoint captures an unbiased selection of attributes at the time, the aggregate market shares will also be unbiased. The criteria for conjoint to work at the aggregate level are considerably less stringent than for individuals.

To summarize, conjoint works by forcing respondents to make trade-offs among attributes. They then simplify the task by selecting a small number of attributes on which to base their judgments. To the extent that this pattern of simplification is mirrored in the marketplace, then conjoint market shares will predict quite well.

3) Conjoint profiles are orthogonal.

The use of orthogonal arrays is an aspect of the original psychometric formulation that has resisted modification by the marketing research community[13]. In particular, main-effects fractional factorials have been heavily used because they permit more attributes and levels. In the case of decomposable stimuli, the simplification by respondents typically assures that interactions will not be present.

The orthogonal nature of these designs is important in a way not generally appreciated. An orthogonal design is simply one in which the levels of different attributes across profiles are uncorrelated. Such designs assure that an estimate of one attribute is unaffected by the estimate of other attributes. It might appear that we could suffer moderate levels of multicollinearity without much harm. Most econometric models seem to thrive with much higher levels of multicollinearity. However, the fact that respondents regularly simplify the conjoint task leads to substantial difficulties if any attributes in the design are correlated. Let me illustrate with an example.

I was involved in a conjoint study dealing with snowmobiles. We were concerned with the impact of engine size on the acceptability of the snowmobiles. So that the profiles would be realistic, we increased the price by an appropriate amount for each of the engine sizes. Price was positively correlated with engine size in the design. This was not expected to be a problem, except that multicollinearity might render the estimates marginally less efficient. However, we found that for many respondents the coefficient of price had the wrong sign (high price is preferred) and for others the coefficient of engine size had the wrong sign (small size is preferred). We believe that respondents, in simplifying, had tended to focus on one of these two variables. For example, those who focused on engine size gave higher evaluations of profiles with larger sizes, but these, by our design, had higher prices. Thus, high price appeared to be

[13] See footnote 7 regarding the enlightened frame of reference with regard to orthogonal arrays.

desired by these subjects.

This account points to an advantage inherent in orthogonal arrays. For orthogonal arrays, the main-effect estimate for each attribute is independent of the others, whereas in the correlated case, this independence does not hold. If attributes are correlated, misspecification results in biased estimates. The particular misspecification that so often occurs in conjoint is simplification, where a number of attributes are effectively ignored. With orthogonal arrays, the estimated coefficients for the attributes remain unbiased. In the correlated case, misspecification results in distortions in the coefficients for both the attributes focused upon and those ignored. Thus, orthogonal arrays play an important role of increasing the robustness of conjoint by making it less likely that coefficients have counter-intuitive signs. This robustness contributes to much of the managerial satisfaction with conjoint.

4) Conjoint Simulators Account for Heterogeneous Tastes in a Market.

A final reason conjoint works relates to the way it is used. Typically, partworth functions are estimated at an individual level, then these are aggregated to produce estimates of market share under various conditions or scenarios. These simulations implicitly reject the notion that one homogeneous customer can account for choices in the marketplace. Instead, one is forced to deal with each customer having an idiosyncratic preference function, or at least with an explicit clustering in which strongly differing tastes are represented in different clusters. This practice of preserving the heterogeneity of individuals in simulators facilitates the representation of two important properties of markets that are difficult to achieve with other market research techniques. These are the properties of differential substitution and dominance.

Differential substitution refers to the notion that a new competitor in a market tends to take share differentially from those brands with whom it is most similar. For example, New Coke took most share from Classic Coke and Pepsi, and had relatively little impact on the lemon-lime soft drink category. Dominance refers to the idea that a brand that is equal on most attributes but slightly worse than its competitor on others gets very low share.

Managers understand these two phenomena and expect simulations to reflect them in positioning and new-product studies. Unfortunately, most models of market structure can account for differential substitution and dominance only in a very awkward fashion. By contrast, both phenomena arise naturally out of a conjoint simulator. For example, differential substitution occurs because individuals who like Pepsi tend to like Coke. Generally, changes in any brand will have a greater impact on similar brands than dissimilar ones. Dominance is represented in a conjoint simulator since a brand that is dominated consistently loses out to that competitor and achieves almost no share.[14]

[14] Joel's discussion of the success of simulators based on individual-level data well captures why HB analysis (which yields individual-level estimates) has been such a boon for choice researchers. Later developments such as Randomized First Choice have leveraged these principles of differential substitution and dominance, but achieving individual-level part-worth models has provided probably the most incremental benefit.

To summarize, conjoint works because it is derived from a task that forces respondents to trade off attributes in ways that may parallel actual buying behavior. The orthogonal design provides not only efficiency, but a strong degree of robustness against misspecification. Finally, the preservation of the utility function at the level of the individual or segment permits us to simulate a market that behaves in ways we expect.

The Future of Conjoint

There are three areas in which substantial changes are anticipated. The first area of change involves conjoint theory, the way we think about and organize the field. The second involves the task. The final area involves the ways conjoint is used.

1. Conjoint Theory: From Estimation of Utilities to Emulation of Behavior

In terms of theory, the psychometric framework has one fatal flaw—it assumes that utilities exist that account for preferences. Reality, unfortunately, is far more complex. Preferences between profiles are better described as being constructed, using various heuristics from the information at hand. The additive models in conjoint may reflect this process and at times correspond to it quite well, but conjoint certainly cannot reveal an interval scale in the brain. Instead of thinking that conjoint estimates latent utilities, it is more appropriate to consider that it emulates choice behavior.

There is a significant loss associated with giving up the idea that conjoint reveals a latent preference scale. The existence of a scale means that it is possible to formulate optimal experimental designs that result in the most efficient estimates of that preference scale. If, instead, conjoint is viewed as a paramorphic emulation of behavior, then it is no longer clear what makes a good design.[15]

There are some advantages with viewing conjoint as an emulation of choice behavior. First, we are no longer permitted to beg the question about the applicability of conjoint to the marketplace—something we can do if the same utility scale is presumed to underlie both choice and conjoint. Instead, we are forced to ask whether the conjoint task corresponds to the choice in the marketplace. For example, it is relevant to assess the number of dimensions that are actually used in the market, then choose a conjoint task that results in similar depth of processing. Second, in the behavioral perspective, one is freed from a rigid adherence to a particular question form to capture appropriately the choice process.[16]

[15] The search for ways to incorporate prior beliefs about latent utilities into the design of choice experiments has proven elusive over the years (See Johnson's paper on ACBC within this same volume). Greater utility balance usually leads to greater fatigue and respondent error. And, if prior beliefs aren't re-introduced when estimating final part-worth utilities, utility-balanced designs may also decrease precision.

[16] These points clearly argue in favor of choice-based experiments, with multiple product concepts available per choice task. Had there been a method for successfully estimating idiosyncratic models of preference from choice data in the late 1980s, Joel no doubt would have embraced choice-based conjoint methods at that time.

In contrast to utility, behavior can be captured in a number of paramorphic ways. Indeed, to get at some behavior it may be better to use different kinds of questions, such as combining direct elicitation and paired comparisons, rather than focusing on a particular question type. Switching to a different kind of question may discourage respondents from getting into a response pattern. Further, the differences in responses across question types will reveal the stability of the choice behavior.

2. Conjoint Task: From Monolithic and Rigid to Multifaceted and Adaptive

As our way of thinking about conjoint changes, so will the task we ask of subjects. Two important changes will occur, both adding new kinds of questions to the traditional conjoint task. The first involves the ability of routines to adapt to the idiosyncratic behavior of respondents, while the second adds a relatively realistic choice task at the end of the conjoint task to better assess the correspondence between conjoint and market choice.

We have already begun to see ways in which conjoint can profitably adapt to the needs of individual respondents. The Sawtooth Software ACA System uses "priors" to construct pairs so that paired comparisons are as closely balanced as possible. While this reduces the strict statistical efficiency of the design, it makes the questions more challenging and increases the correspondence between conjoint and subsequent choice (Huber and Hansen, 1986). The flexibility of the personal computer in administering conjoint will certainly lead to other adaptive mechanisms. Two such applications are particularly exciting—hierarchical conjoint and interaction testing.

Hierarchical conjoint permits the modeling of rich decision making despite simplification of the conjoint task by respondents. If respondents can only cope with two or three attributes at a time, the routine determines the partworth functions for these most important attributes, then fixes their levels. Subsequent test profiles only differ on the remaining attributes. For example, the values of location and price might be estimated first in an apartment study, followed by furniture style and room layout. Under standard conjoint, these less important attributes might not be revealed, whereas in hierarchical conjoint both their position in the hierarchy and relative importance could be assessed.

A second adaptive mechanism concerns the search for interactions. A promising technique might work as follows: First, a main-effects design would make rough estimates assuming no interactions. The residuals from these initial judgments would be tested for weak (and confounded) evidence of interactions. Then these potential interactions would be tested through specially designed questions. Such a method would avoid the current problem of having to assume that interactions are zero, and could be very helpful in studies where different interaction patterns are expected across respondents.

Perhaps the greatest prospect for improving conjoint involves the inclusion of a relatively realistic choice task at the end of the exercise. These choice tasks sometimes take the form of asking respondents to choose brands from a simulated store or having them evaluate actual products. A much less costly, if somewhat less realistic option, is to add choice questions at the end of the conjoint exercise. These might take the form of choices among alternatives defined on different attributes than those displayed in the original conjoint. This is easy to do within the Sawtooth Software ACA System by adding Ci2 System questions after the ACA System section. Further, new developments in personal computers will permit potential choice objects to be displayed in color video, thereby increasing the realism of the holdout choices even more.

The value of having a holdout choice task is twofold. First, in the field this holdout task is useful in identifying respondents whose conjoint responses are unlikely to correspond to their behavior. These respondents can then be given less weight in the simulation. Second, it permits an immediate assessment of the relevance of conjoint to choice. Where a conjoint model appears not to correspond to choice, it can be changed or improved[17]. This permits the testing of different forms of conjoint and leads to continuous improvement in its predictive validity.

3. Conjoint Simulators: From Complex and Opaque to User Friendly and Understandable

The third area in which we can expect conjoint to progress involves the way data are used in choice simulators. Elaborate choice simulators currently exist, permitting the analyst to ask virtually any question of the data. The positioning of a product can be optimized with respect to sales or profitability. Alternatively, one can assess the impact of changes on positioning or competition on the behavior of various segments. Unfortunately, these simulators are not particularly user-friendly. With time they will become easier to use and their use by managers should increase[18].

Even if made more user-friendly, there is still a problem. While simulators permit the manager to cope with heterogeneous tastes in the market, they remain a black box. The only way for a manager to understand a simulated market is by experimenting with a large number of runs. These simulation runs give managers a feel for the behavior of the market in the face of different positionings or competitive offerings. Developing this understanding is hard, relatively unstructured work. We need to develop ways to permit managers to understand more directly the behavior of the market being simulated[19]. Preference spaces may provide part of the answer, but

---

[17] Here Joel argues in favor of adjusting metric-based conjoint methods to better predict limited choice holdouts. Again, lacking a robust methodology to estimate individual-level models from experimentally-designed choice tasks alone, it naturally wasn't apparent at the time that we might forego the ratings-based exercise altogether and just focus on the more realistic choice scenarios as calibration tasks!

[18] Certainly, simulators have become easier to use and ubiquitous. Spreadsheets were available (Lotus 123) in the mid-1980s for developing conjoint simulators, and Microsoft's Excel was being released in the year Joel presented this paper at Sawtooth Software. Still, in 1987, it was hardly widespread practice to build conjoint simulators within spreadsheets. Today, researchers commonly build attractive spreadsheet applications for market simulations in Excel. Sawtooth Software's ASM (Advanced Simulation Module) has made product optimization within its commercial market simulator (SMRT platform) very straightforward.

[19] These problems by in large have not been solved. However, the application of cluster analysis or particularly Latent Class to the development of "needs-based" segments and the use of segment membership as banner variables for "cross-tab" style display in simulation output may increase the level of understanding of market structure and preference. Still, it takes a talented manager to absorb so much detail and come to sound conclusions.

it is very difficult to represent both respondent heterogeneity and central tendency on one map. Working with a small number of segments also helps. We would like to know how to specify a small number of segments, such that their aggregate behavior closely approximates the entire market. Defining such segments remains an unsolved question that evades simple solutions.

Just as the computer continues to make the conjoint task itself more appealing to respondents, it will also increase the ease by which the output of conjoint can be applied by managers. Once again, we may have been saved by the computer. We have come to realize that choice behavior cannot be captured by a simple scale of utilities. As we come to accept a behavioral base to conjoint, we can never return to the elegance and simple unity that characterizes the psychometric framework. However, with the computer we have a tool that may be powerful enough to mirror the complexity of behavior and display it in a manageable and understandable way.

REFERENCES

Cattin, Philippe and Dick R. Wittink (1987) "Commercial Uses of Conjoint Analysis: An Update" Working Paper, Graduate School of Business, Cornell University, Ithaca NY.

Falmange, Jean-Claude (1976) "Random Conjoint Measurement and Loudness Summation" Psychological Review, 83, 65-97.

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

Huber, Joel (1975) "Predicting Preferences on Experimental Bundles of Attributes: A Comparison of Models" Journal of Marketing Research, 12, 290-7.

Huber, Joel and David Hansen (1987) "Testing the Impact of Dimensional Complexity and Affective Differences ln Adaptive Conjoint Analysis" Advances in Consumer Research, 13, in press.

Johnson, Richard M. (1974) "Trade-Off Analysis of Consumer Values" Journal of Marketing Research, 11, 121-217.

Krantz, D.H., R. D. Luce, P. Suppes and A. Tversky (1971) Foundations of Measurements, Vol 1, New York: Academic Press.

Krusksal, Joseph B. (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data" Journal of the Royal Statistical Society, Series B, 27, 251-63.

Olshavsky, Richard W. and Donald H. Granbois (1979), "Consumer Decision Making—Fact or Fiction" Journal of Consumer Research, 6, 93-100.

Srinivasan, V. and Allan D. Shocker (1973) "Linear Programming Techniques for Multi-dimensional Analysis of Preferences" Psychometrica, 38, 337-69.

Thaler, Richard (1985), "Mental Accounting and Human Choice" Marketing Science, 4, 199-214.