# Sawtooth Software

***RESEARCH PAPER SERIES***

# What We Have Learned from 20 Years of Conjoint Research:

## When to Use Self-Explicated, Graded Pairs, Full Profiles or Choice Experiments

Joel Huber,
Duke University, Fuqua School of Business
1997


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

# What We Have Learned from 20 Years of Conjoint Research: When to use Self-Explicated, Graded Pairs, Full Profiles or Choice Experiments

**Joel Huber**
**Duke University**

Conjoint analysis as a commercial method has been available for more than 20 years (Green, Wind and Jain 1972). Ten years ago, 1987, at a Sawtooth conference, I raised the question of how conjoint could work so well, given obvious differences between the task and market choice. I proposed that conjoint works primarily because the simplification in a conjoint task mirrors the simplification in the marketplace. That is, the complexity of the marketplace encourages people to make choices based on relatively few attributes, effectively selecting the attributes that they will attend to and value in a given choice. In the same way, the complexity of a full-profile conjoint task also encourages respondents to attend to a subset of the attributes. Thus conjoint works by simulating the *attribute selections* process that occurs in actual choices.

Much has changed in the last decade. The biggest change is the expanded arsenal of methods that enable us to measure people's values (Green and Srinivasan 1990). There are sophisticated self-explicated methods that break down choice into values for levels of attributes, weights for different attributes, and perceptions of alternatives on those attributes. ACA in its new version permits different kinds of self-explicated assessments and allows various scales to be used. Full-profile can be built with more general choice designs (Kuhfeld, Tobias and Garratt 1994). Perhaps the biggest change has been the availability of easy-to-use choice-based conjoint systems, which measure tradeoffs with responses to hypothetical choice sets. The question for the next millennium is not whether conjoint works, but defining the contexts in which different methods are appropriate.

One fact is irrefutable: different methods, and even the same method in different contexts, give different results. Consider the following three examples:

1. On high technology products such as computers, ACA's price partworth for knowledgeable respondents needs to be doubled to accurately predict subsequent choice-based conjoint (Pinnell 1994). That is, choices among brands are best predicted if the raw ACA utilities are doubly weighted before being entered into a choice simulator.

2. The relative value of price over brand doubles from the early choice tasks compared to later ones (Johnson and Orme 1996).

3. In contrast to ACA and full-profile conjoint, choices reflect 20%-30% greater emphasis on the most important attributes and put 30%-40% more weight on the least preferred levels of each attribute. (Zwerina and Huber 1997, Orme, Alpert and Christensen 1997, Huber, Ariely and Fischer 1997).

These results have been profoundly disquieting to me, and should be to you. They make clear that the myth of measuring a person's true utility structure is just that, a myth. However, in a positive sense they also suggest that our goal of formulating one way to best measure values needs to be replaced with a goal of matching the characteristics of decisions in the marketplace with those of the task. The three results above are not anomalies. I propose that we can understand the differences among techniques exhibited by examining three characteristics of methods to measure values: the attention that they place on various attributes, the way they alter competitive expectations, and, most importantly on the kinds of values they promote. Depending on the market being simulated, different methods may be appropriate. The purpose of this paper is to provide some guidelines to help you make such a matching.

It is important for me to acknowledge a strong, and somewhat controversial belief that I bring to this discussion. I believe, and hope to convince you, that the purpose of the typical trade-off study ought *not* to be the prediction of short-run market behavior. Short-run behavior is both quite predictable and of minimal strategic value. If we want to know what people will choose, the best predictor is what they chose last time. Part of the reason market behavior exhibits so much inertia is that most market choices take very little time. Even when time is taken, decisions are made using heuristics that permit reasonable choices despite poor comparative information and relatively little effort on the part of consumers. Instead of asking what a heuristic laden and opportunistic customer would buy, I believe that companies need to know:

1. What customers will choose if they *attend* to the attributes.
2. What they will do if customers' competitive *expectations* change.
3. What customers will do if they think about how much they *value* the attribute.

Put differently, if a company needs to know what customers do now, that is best approximated through econometric analysis of current sales data. For the majority of problems, short-term prediction is less important than knowing what a customer would do if and when the competitive environment changes. What happens if a price is lowered and customers gradually notice? What happens if the competitors promote a new feature? What happens if a consumer magazine makes side-by-side comparisons of different competitors? The sections below detail three ways in which various tradeoff tasks alter these three characteristics of the simulated purchase experience: they increase attention to displayed attributes, they alter expectations about the competitive offerings, differentially encouraging the evaluation of various attributes.

**Three Properties of Evaluation Tasks**

*Increase Attention.* Evaluation tasks intentionally force respondents to attend to attributes that they might otherwise not notice. In doing so, attention can elevate the importance of particular attributes to a level that is greater than would occur in the marketplace. For example, featuring the attribute "surge protector" may make this attribute salient even though it may not be salient in actual choices.

Drawing on the ideas given above, this lack of correspondence to the market may be seen as an advantage rather than a disadvantage. It certainly suggests that measured value of an attribute should be viewed as *conditional* on the respondent noticing the attribute. Being conditional on

attention is advantageous since in most markets important but unnoticed attributes tend to become noticed over time. Consumers in markets may be slow, but they are not stupid. If a product has better features or is lower priced, it will be noticed eventually, initially by vigilant consumers, and thereafter through word-of-mouth, rating services and by retailers. Finally, if a product feature needs to be noticed to affect choice, then advertising or promotion of that superior feature is a relatively simple task. The important point is that by forcing attention on specific features or prices, various evaluation tasks provide a prediction of marketplace behavior as customers become aware of those attributes.

*Alter Competitive Expectations.* An elaborate set of beliefs assist us in our choices. Two kinds are relevant here. The first involve reference levels within attributes. For example, we have price and performance expectations among competitive offerings, enabling us to spot an important new feature or a price that is out-of-bounds. The second kind of expectation involves associations between attributes, as when one uses one attribute to draw inferences about other attributes. These beliefs also help us to simplify choice by using one attribute as a proxy for others.

Research has shown that these expectations are important but fragile. Once successfully challenged in the market, then the new expectations alter the way we process information and make choices. To the extent that measurement tasks also break down beliefs, they simulate what is likely to happen if competitive conditions change. In doing so they assess in a short period of time changes in behavior that the market may take longer to accomplish.

*Encourage Evaluation.* All trade-off methods, to a greater or lesser extent, encourage respondents to think about the meaning of an attribute in terms of their lives. The primary mechanism generating these thoughts is the trade-off task itself, which encourages respondents to think about the value of one attribute compared with another. Notice that the type of task can encourage or discourage this evaluation. For example, by loading the alternatives with features, a number of attributes may never be noticed; alternatively, by loading an attribute with a large number of levels, it can draw attention to itself. (Wittink, McLauchlan and Seetharaman 1997) .

Some studies only gauge reaction to the idea of a feature, while others ask respondents to use the different versions. These latter studies simulate the effect of trial use and evaluation of the feature. The point here is that the degree and depth of the evaluation is part of the study design. In the next section we will explore how particular tasks differentially elicit values.

**Differences Among Four Tasks: Self-Explicated, Pair Comparisons, Full Profile and Choices**

We examine four tasks commonly used to measure values: self-explicated methods, graded pair comparisons, full profile ratings and choices. We initially explore simple and somewhat stylized versions of each method, examining the ways they focus attention, alter competitive expectations and ultimately transform the values generated. After having considered these standard forms, we will then discuss how varying implementations of the tasks further alter the values generated by each method.

*Self-Explicated Methods.* Self-explicated methods typically involve two data collection

stages. First, the respondent provides the relative value of levels within each attribute. For example, ACA's default asks for a ranking of the levels of the attributes, while a different method might assign 100 points to the most preferred level and corresponding values to other levels. Then the self-explicated model needs to determine a weight for each attribute. ACA's default accomplishes this task with a 4-point scale, while other options permit a continuous scale or point allocation. The value of an alternative is then the weighted combination of the values of each of its levels.

However they are operationalized, self-explicated models are best applied to gauge reaction to a particular offering, in the absence of an explicit competitive offering. They are also useful where there are a great many attributes or outcomes associated with the choice. For these reasons, self-explicated models are commonly used for services, such as how much more positively you would feel about a hotel if its service level improved. They also work well for actions lacking comparable alternatives, such as how likely you would be to trade in a car. This distinction between within- and between-alternative orientation is important. The self-explicated models (e.g. Fishbein, weighted-additive models) were primarily developed to measure attitudes for an alternative, to assess how the characteristics of a product lead you to like it. Unfortunately, positive attitudes often do not translate into purchase decisions. Witness the strongly vocal Macintosh users who nonetheless buy DOS or Windows-based machines. As this example illustrates, attitude toward a brand may be a poor predictor of purchase in a competitive setting.

Because direct evaluation draws attention to attribute levels, it tends to overweight attributes that might otherwise be unimportant in a competitive context. It is easy to think of reasons why, for example, an audiovisual feature (such as Bose speakers) might be important in evaluation of a computer but much less important in choice. In the market, this attribute may become overshadowed by more important attributes, or not perceived to differ sufficiently.

Self-explicated models also become problematic in the face of correlations among attributes, where there are strong expected associations between attribute levels. For example, computers that handle numerical calculations quickly are often faster at handling strings. The customer may assume one attribute is a surrogate for the other, both being measures of 'speed'. Suppose, however, the manufacturer needs to know the relative importance of each. Should each be included as a separate attribute? Does the value respondents provide in a direct rating of numerical speed include an implicit component for string handling? There are no simple answers to these problems within the context of the self-explicated methods.

While self-explicated models can become unstable or ill-defined in the face of correlated attributes, standard forms of conjoint get around this problem by breaking up expected associations among attributes. When exposed to computers with both high reliability and low weight, the heuristic of using one to 'stand for' the other is both less credible and less useful. By contrast, self-explicated tasks do little to break down pre-existing assumptions about the world. Indeed, they tend to reinforce both beliefs about the available levels of attributes and their associations. Accordingly, self-explicated tasks are most useful in simulating those markets whose offerings are stable.

There is some evidence that the self-explicated process works quite well when the

alternatives reflect stable and veridical beliefs about the world. For example, in two studies of MBA job choice from offers received, the self-explicated model provides a better prediction than a conjoint model (Srinivasan and Chan Su Park 1997). In my view that makes sense because the job offers are likely to reflect expected levels and correlations. Thus, the expectation-based heuristics inherent in the self-explicated model work well. By contrast, where the offerings are substantially different from expectations, then a task that breaks from those beliefs should work better.

Summarizing, self-explicated models are best for:

1. Contexts in which many attributes are important for the decision.
2. Markets where expectations about levels and associations among attributes are stable.
3. Decisions where the action depends on the attitude towards an individual alternative or action, rather than in the context of competitive offerings.

*Graded pair comparisons.* The graded pair comparison task puts two alternatives next to each other and asks how much more one is liked than another, and in doing so immediately draws attention to the *differences* in attribute levels for the pair. For example, in the comparison between a 500MB laptop at $2400 against an 800MB model at $2900, the focus is on whether 300MB additional memory is worth the $500 price difference.

Three evaluative effects follow from the pairwise task. First, the task tends to flatten attribute importances by making otherwise unimportant attributes salient. Since attribute differences are easy to assess in a pairwise task, the importances are spread out across different attributes. This flattening of attribute importances tends to be particularly strong when there are only a few attributes differentiating the pair, as occurs in ACA's default. In tasks where the dominant attributes are missing, respondents come to value attributes that might otherwise be overshadowed.

Second, the pairwise orientation tends to reduce the importance of external reference levels. In particular, in valuing the differences between levels, there is less attention placed on their absolute levels. For example, suppose there is a real resistance to paying more than $3000 for a computer. However, in a pair task, there will typically be relatively little differences in respondents' reaction to computers with $2400-$2900 prices, compared to those with $2600-$3100 prices. The problem arises because thinking about the $3000 resistance level takes an extra processing step after the difference has been evaluated. This focus on differences becomes stronger and stronger as respondents repeat the pairwise task.

Finally, the focus on differences increases the relative value of attributes about which such differences are easy to value. It is easy to assess the implications of a $500 difference in price, but how should one value the difference between, say, Compaq and Dell? More generally, numerical attributes about which it is easier to characterize the difference, such as price, size, rating, will have greater weight in pair tasks than categorical attributes such as brand name, product family or country of manufacture. Nowlis and Simonson (1997) have shown that when one's focus is on individual alternatives then the categorical attributes have more weight, while a pairwise task emphasizes continuous attributes.

In summary, pair comparisons are most appropriate when:

1. Modeling a market in which alternatives are explicitly compared with one another.
2. Approximating a deeper search where the consumers draw information from a broad range of attributes.
3. Contexts in which within-attribute value steps are smooth and approximately linear.

*Full Profile Ratings.* A full profile rating is a very common form of conjoint in which respondents evaluate a group of, say, 16 alternatives, each defined as a bundle of characteristics. A typical task requires that each alternative be evaluated on a simple scale, say, a 1-7 attractiveness rating, or a 0-10 purchase likelihood. Regardless of the scale used, the critical defining aspect of this task is it encourages respondents to evaluate each profile *individually*. Put differently, attention is focused on the acceptability of an alternative's attributes, rather than differences between alternatives as we found for pairs. This seemingly innocuous attentional difference produces strong effects on shifts in expectations and on values that emerge from the task.

To understand the impact of a within-alternative focus on expectations, think about rating the attractiveness of a laptop. You might have feelings about it, for example, that the price is too high or the processor untrustworthy. However, to give it a rating it is important to know how it compares to others in the set. You need to learn to identify the kind of laptop that gets a "2" rating compared to one that gets a "5". Respondents achieve this mapping quite easily by getting a sense of the range and the average value of the profiles. Evidence for the speed and strength of this adaptation process can be seen by noting how the average rating hovers for most respondents within one point of the center of the scale, *regardless of the respondent's general attitude.* Indeed, in most analyses of ratings the mean is treated as a nuisance variable to be discarded, rather than a measure of attitude toward the category.

This adaptation of respondents to the average profile has important implications for the implementation of ratings-based conjoint. Since the adaptation does not occur instantly, respondents generate more reliable ratings if they understand the range of the profiles. In our work, we have found that full profile ratings predict better when preceded by an ACA task than when followed by it (Huber, Wittink, Fiedler and Miller 1988). The self-explicated and pair sections in ACA permit people taking the subsequent full profile task to have a good sense of the attribute ranges and how they combine to make more or less attractive alternatives. More generally, warm-up tasks are very important in full profile conjoint. Louviere (1985) recommends two warm-up tasks, first rating an alternative that is poor on most attributes, followed by one that is good. These two tasks familiarize the respondent to the scales and stabilize subsequent ratings.

In addition to efficiently shifting respondents from their external reference levels to the average within the set, the full profile task also is quite effective in breaking up associations between attributes. Respondent's associations are weakened by profiles that go against their expectations; for example, when they find quality processors with low power, or light-weight

laptops with long battery life. Indeed, the orthogonal designs common in most full profile exercises require that any pair of attribute levels have an equal likelihood of being paired within a profile, thus assuring that respondents will experience profiles that violate their prior expectations. The net effect of the full profile conjoint task is to generate decisions that are relatively free from simplifications that come from reference and associational effects. Even more so than graded-pair comparisons, full profile ratings *decontextualize* respondent values.

In addition to producing values that are relatively context free, full profile's focus on the individual alternative changes the resulting values compared with graded pairs in three ways: fewer attributes are featured, those featured tend to be qualitative, and greater weight is attached to the lowest levels of each attribute. Each of these value shifts is considered below.

The first value shift is a focus on a few attributes. There is no logical reason why ratings-based conjoint should limit attention to a small number of attributes, but that is what happens, study after study. At the individual level the pattern is clear; out of say seven attributes, two or three attributes will be significant, while the rest are virtually zero.

In addition to a small number of attributes becoming prominent in the full-profile task, those featured are more likely to be qualitative. (Simonson and Nowlis 1997). This qualitative focus follows from the full profile's orientation to the individual alternative, and contrasts with the focus on numerical attributes in pair comparisons discussed earlier. Qualitative attributes tend to have attitudes attached to them regardless of context. Consider your immediate evaluation of the brand name, Packard Bell, or the feature "multi-media." By contrast, how you feel about a 100MhZ processor depends critically on whether it is compared with an 80MhZ or a 130MhZ. In a conjoint rating task the standard of comparison is implicit, while in pair comparisons the contrast is with the other pair. The implication is that if the market action you wish to simulate by the task involves such explicit comparisons, then a pairwise method is preferred. To the extent that each alternative is evaluated alone (like homes, cars, recordings) then a full profile task is more appropriate.

Finally, there is evidence that full profile puts greater weight on the negative levels of attributes than pair comparisons. For positively-coded attributes, this orientation is reflected in diminishing returns, so that the gain from a one- to a two-year warranty is greater than the gain from a two- to a three-year warranty. For negative attributes, it is expressed as increasing aversion to the negative attribute, so that the loss of moving from 4 to 6 pounds for a laptop is less than the move from 6 to 8 pounds (Orme, Alpert and Christensen 1997, Huber, Ariely and Fischer 1997). The process driving this curvature is a combination of task simplification and loss aversion. The task is simplified by downgrading alternatives containing the least-preferred attribute levels, while avoiding these low levels protects against losses associated with making a bad choice.

To summarize, full profile conjoint is most appropriate when:

1. It is desirable to abstract from short run level and associational beliefs.
2. The market choices demonstrate substantial simplification both in a limited number of attributes being processed and greater weight on the most negative levels.

3. The focus of the decision is within alternative so that the explicit comparisons between pairs of option are rare.

*Choices.* A choice task can be viewed as a group of full profile concepts, where, instead of requiring that each be individually rated, the respondent is asked to indicate which is best. However, this formal similarity to the conjoint rating task belies strong processing effects that derive from the act of choosing. Choosing shifts attention away from assessing how much better one alternative is compared to another and towards processes that lead one to be reasonably confident that the one chosen is best. This goal encourages even greater simplification than a rating task. Some of this simplification is evident in the time taken. A 9-attribute, 3-alternative choice task took about 30 seconds per choice, while a rating task took about 30 seconds for each alternative (Orme, Alpert and Christensen 1997). Clearly, respondents are not evaluating each of the alternatives and choosing one with the highest score.

What are they doing? First, respondents are looking for dominating, easy choices. If they find none, they look to see if they can exclude any of the alternatives, typically those with low scores on important attributes. Once the choice is down to two, then a quick scan of important attribute differences completes a satisfactory selection process. In part because of such strong simplification, the process is not very reliable. In choice experiments where one choice set has been repeated, the same alternative (5 attributes, 4 alternative) is chosen only 70%-80% of the time (Huber, Wittink, Fiedler and Miller 1993).

What is the impact of choices on expectations? As with pair comparisons and full profile tasks, respondents quickly leave their old reference levels as they adapt to the alternatives provided. Of course, if the alternatives are too bad then people react negatively to the entire process, and if they are too good, they may make very little effort to choose, since all are satisfactory. However, within bounds, respondents in choice tasks quickly learn to evaluate each alternative compared with the local competition within the choice set.

Associations between attributes are more difficult for respondents to see in choice sets, but learning does eventually take place. The Johnson and Orme (1996) study shows that the relative importance of brand name relative to price drops by 30% in the course of 3-4 initial choice sets and to 50% after 10 tasks. Initially, brand name is important. Soon, however, respondents realize that brand name is not predictive of price or features, and evaluate its contribution, *holding other aspects constant.* In the same way, they also learn to evaluate each attribute independently of other attributes commonly associated with it.

Choice tasks shape values in three ways. First, greater simplification leads to even fewer attributes being featured (Orme, Alpert and Christensen 1997, Zwerina and Huber 1997). Second, the attributes featured are different. Since choices combine both within-alternative processes (like full profile) and between alternative judgments (like pair comparisons), the focus is not with respect to quantitative or qualitative, but follows from the fact that choices are more *immediate* and *real.* Choices lack the abstract and hypothetical quality of ratings--respondents are being asked if they would actually choose the alternative. Attributes whose impacts are immediate and concrete come to the fore compared to those that are distant or abstract. Consider the following two examples. First, IntelliQuest (Pinnell 1994) has found that the utility values

for price have to be doubled to make their ACA values match the subsequent choice task. Second, in our study of refrigerators, we found that long-term cost of annual energy use was more important in ratings than in choice, whereas the immediately due sales price was more important in choice over ratings. (Huber et al. 1993).

The third way in which choices shape values is by putting even greater weight on the poorest attribute levels. This tendency is manifest in large utility differences between poor-middle levels, and relatively small differences between middle-best levels of the attributes. (Orme, Alpert and Christensen 1997, Zwerina and Huber 1997). The mechanism here is the same combination of loss aversion and simplification found in the full profile task, but the effect is stronger. Rather than getting lower ratings, alternatives with low levels on important attributes are more likely to be simply dropped from consideration.

To summarize, choice is most appropriate when:

1. Simulating immediate response to competitive offerings, especially brand and price studies.
2. Decisions are made on the basis of relatively few, well-known attributes with substantial aversion to the worst levels of each attribute.
3. Consumers make these decisions on the basis of competitive differences among attributes given.

**What About Different Versions of the Methods?**

To simplify the exposition, the preceding sections have deliberately focused on relatively pure types of the self-explicated, graded pair, full profile and choice tasks. Of course, most implementations of these tasks differ importantly from these pure forms in terms of the ways they affect attention, competitive expectations and values. However, the logic used to understand the task effects of the pure forms can be used to predict the impact of these modifications. A few examples are given below.

*ACA:* ACA combines a self-explicated and a pair comparison task. The self-explicated task permits a good introduction to attribute levels and tends to bring more attributes into consideration. The pairwise task further increases attention to less important attributes, since attribute differences are so easy to process. Finally, the focus on differences and the linear priors tends to result in quite linear steps in utility between adjacent levels.

If desired, ACA can be made more like choice by encouraging simplification and loss-aversion. Curvature in the ranking of levels can be encouraged by changing the task to having respondents assign the best level of each attribute 100 points and proportional values to lower levels. Furthermore, simplification emulating choice can also be encouraged by having pairs differ on more attributes, say 4-5 attributes differing rather than the default two or three. One may not want to make this modification, however, as there is evidence that ACA works best with 2-3 attributes differing (Huber and Hansen 1988).

*Sort Board for Full Profiles:* A common way to make full profile ratings more like choice is to give respondents a deck of cards and ask them to sort them on, say, a board with 10 categories

from groups from worst to best. This task brings in attentional properties that mirror some aspects of pair comparisons and choice. Respondents typically group the cards into rough categories followed by a more detailed evaluation of alternatives in the same pile. This latter pairwise focus tends to bring attention to less important attributes, since the alternatives sorted together often have the same values on the most important ones. Further the two-stage process of an initial screen followed by more detailed pairwise assessment of final alternatives mirrors what happens in a number of choice contexts (Payne 1976).

*Simplified choices:* Not only are there ways ratings can be made more like choice, but choices also can be made more like ratings. Since the major property of the choice task is that it encourages simplification, a common way to limit this tendency is to reduce the processing required. Two ways are possible, reducing the number of alternatives per choice (Pinnell and Englert 1997), or reducing the number of attributes differing (Chrzan, Fellerman, 1997). Both these methods lessen the statistical power of the design, but increase the ability of respondents to respond consistently to the task. Generally, simplifying choice can be expected to increase the number of attributes that are processed and decrease the weight put on the least preferred attribute levels.

**Summary: A Framework for Evaluating Methods**

Table 1 provides a summary of the different methods and their impact on attention, competitive beliefs, and resultant values, permitting a link between the market decisions and the appropriate task. Below, I reiterate the important ways the measurement task shifts attention, competitive beliefs and resulting values, and suggest ways that this knowledge can be used to guide the development of useful commercial studies.

*Attentional shifts* are an integral part of any value measure. Simply mentioning an attribute increases its importance, raising the specter of attributes appearing important that otherwise would be ignored in the market choices. One way to limit this problem is to load the task with enough attributes so the unimportant ones are ignored in the task process. This task simplification screening is particularly strong in full-profile ratings and choices. The risk here is that the task may encourage respondents to ignore too many attributes, in which case pair comparisons or self-explicated tasks may be more appropriate. In any event, it is important to think about ways attributes can become important in the market context, for example through a promotional campaign, shelf talkers, or simply gradual understanding of the market over time. It is those attributes that should be featured in the task.

*Competitive beliefs* are changed by value measurement tasks. Except for self-explicated methods, all the tasks discussed decontextualize judgment by shifting reference levels and changing associations. Reference levels refer to expected levels and ranges of each attribute. These levels assist our market decisions by gauging whether a particular offering is appropriate or not, and enable us to make reasonable decisions in very little time. However, these reference levels are also quite sensitive to the competitive context. Consider the following two examples. What seems like an outrageous price can quickly become acceptable in the face of higher-priced competitive offerings. What seems like an appropriate modem becomes outmoded when compared with the faster models.

To my mind, decontextualizing values from particular or reference levels is an advantage of the various trade-off methods. Just as conjoint easily shakes people from their reference points, so market forces also shift these reference levels. Sticker shock may make people put off buying a car, but eventually they adapt to the new competitive level. Thus, a well-designed tradeoff study can anticipate the effect of adapting to new market offerings. The caveat is that the conjoint context needs to match the future market.

The second change in competitive beliefs is with respect to associations. Like reference levels, associations allow people in markets to make reasonable choices quickly, by selecting a trusted name, store or price tier. Breaking down these associations requires that people really assess the value of, for example, the brand name in itself. Thus, the process of breaking down associations can be seen as a way of approximating what a person would do if expectations are not used to simplify the decision. While it may result in somewhat worse predictions of short-term decisions, it can better approximate the effect of extended thought or discussion.

If measurement tasks focus attention and shift competitive expectations, they also tap different values. We have reviewed three ways in which the task can alter derived values: through an orientation to individual items versus a comparison with others, through the immediacy of the situation it evokes, and through a need to simplify the task. Below is a summary of each of these distortions along with suggestions as to how they might be better handled.

Tasks can evoke either a *comparative or individual-alternative* orientation. As argued earlier, pair comparisons result in greater weight to those attributes whose differences are easy to calculate, whereas full profile ratings put greater weight on categorical attributes such as brand name. The choice of which to use depends on the degree to which the market decision is a comparative one. Thus, the choice of laptop is typically a comparative process, whereas the choice of a job or a house is generally more focused on the fit to one's own values. Further, decisions where the alternatives are not comparable, such as when to sell or whether to buy in a category, focus attention within alternatives and thus are best modeled by ratings-based conjoint or even a self-explicated model.

Tasks can be *immediate or reflective*. Immediate tasks, such as choice experiments, ask respondents which they would choose today. As discussed earlier, the more immediate tasks increase the importance of attributes with short-term implications, such as price, and attributes with visible performance characteristics. The more reflective tasks (say, asking for a tradeoff between two pounds of weight and an hour of battery life) are both hypothetical and relatively timeless. One does not consider one's next business trip, but instead the general pattern of such trips. The implication should be clear. To the extent that the market decision being simulated is based on immediate and short-term considerations, then choice experiments are appropriate. Long-term and repeat purchasing contexts, by contrast, are better modeled by procedures that encourage respondent to abstract from current considerations.

Finally, tasks can evoke varying degrees of *simplification*. Respondents simplify tasks both across and within attributes. Across attributes, respondents simplify by attending to the most important attributes at the expense of less important ones. Within attributes, they discard alternatives that have low levels on important attributes, typically producing the appearance of

strong diminishing returns in the partworths. Choices result in the most simplification, followed by full-profile conjoint, pair comparisons and the self-explicated task. Thus the various tasks provide a way of simulating more or less simplification.

As we examine the ways these tasks focus attention, alter competitive beliefs and change revealed values, the dilemma posed at the beginning of this paper emerges. If we wish to predict short-term, heuristic-bound behavior, then none of the methods reviewed are very good, although some, like choice, may be better than others. However, I believe marketing research and marketing firms will be better off if they err on the side of encouraging more elaborate over less elaborate processing: drawing attention to more attributes rather than less, encouraging a long term rather than an immediate focus on the problem, and by breaking apart the problem for the respondents so that they do not too grossly oversimplify it for themselves. This leads to a recommendation to use pair comparisons and self-explicated methods, and to be particularly cautious with choice-based methods.

There are two reasons for taking this posture. First, individuals certainly use all kinds of shortcuts in making market decisions. However, while individual decisions may make them vulnerable to opportunistic marketers, customers do learn, both individually and collectively. Thus we want a technique that enables people to be most happy with outcomes of the decisions that they make, not to make the decision that they will see later as foolish or short-sighted. The more aspects customers consider, the more long term their orientation, the more they are able to cope with the complexity of the problem, the more satisfied they will be with the offerings they choose.

Konosuki Matsushita said it best:

> "Don't sell customers goods that they are attracted to. Sell them goods that will benefit them." (Fortune, 1997)

I would like to close with a thought experiment. Suppose as part of this conference you win a laptop; but you do not get to choose the laptop. Instead, you get to choose the method that will select the laptop for you. You choose from among the four methods discussed here: self-explicated, pair comparisons, full profile or an individual choice experiment. You will then get the laptop that optimizes your values as expressed by your chosen technique.

Most knowledgeable people do not like this question, seeking control over the choice rather than the method to choose. However, when pushed, most knowledgeable people prefer the biases and distortions from the more thoughtful exercises such as self explicated or pair comparisons compared with full profile ratings or the choice-based task. Choices are seen as evoking in too much simplification, with too much focus on near-term consequences, and too much emphasis on avoiding the 'worst' levels of an attribute. Pair comparisons and self-explicated tasks, by contrast, may err by putting too much weight on unimportant attributes, and perhaps not enough weight on the worst levels, but they are robust and reliable.

The question then is, knowing what you know about the strengths and weakness of the different tasks, which method would you choose to select your own laptop? Perhaps more relevant, which method would you select for your own customers?

**SUMMARY OF DIFFERENCES AMONG VALUE MEASUREMENT TASKS**

| | Self-Explicated | Graded pair comparison | Full profile ratings | Choice |
|---|---|---|---|---|
| Attentional Focus | Individual alternatives | Pair differences | Alternatives in a general context | Alternatives in a competitive context |
| Impact on Competitive Expectations | Reinforces prior expectations | Shifts expected trade-offs (e.g. price-quality) and levels. | Orthogonal arrays break down associations. | Initial reference levels are dominated by attribute differences. |
| Valuation Focus | Feelings towards attributes | Tradeoffs between levels | Selection of attributes and levels | Short term and concrete attributes, loss avoidance |
| Emphasis | Less important attributes | Quantitative attributes | Qualitative attributes | Near-term, concrete attributes |
| Ideal use | Non-competitive contexts, many attributes | Stable trade-offs | Gauge simplification strategies | Immediate competitive effects, simple choices. |

**REFERENCES**


Chrzan, Keith and Ritha Fellerman (1997), "A Comparison of Full- and Partial-Profile with Best-Worst Conjoint Analysis," *Sawtooth Software Conference Proceedings.*

Green, Paul, E, Yoram Wind and Arun K. Jain (1972), "Preference Measurement of Item Collections," *Journal of Marketing Research, 9,* (November) 371-377.

Green, Paul E. and V. Srinivasan (1990), "Conjoint Analysis in Marketing, New Developments with Implications for Research and Practice," *Journal of Marketing, 54,* (October), 3-19.

Huber, Joel and Klaus Zwerina (1996), "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research, 23,* (August) 307-317.

Huber, Joel (1987) "Conjoint Analysis: How We Got Here and Where We Are," *Sawtooth Software Conference Proceedings,* Ketchum ID: Sawtooth Software, 237-253.

Huber, Joel, Dan Ariely and Gregory Fischer (1997), "The Ability of People to Express Values with Choices, Matching and Ratings," Working Paper, Fuqua School of Business, Duke University.

Huber, Joel, Dick R. Wittink, John A. Fiedler and Richard Miller (1993) "The Effectiveness of Alternative Elicitation Processes in Predicting Choice," *Journal of Marketing Research* (February), 105-114.

Johnson, Richard M. and Bryan K. Orme (1996), "How Many Questions Should You Ask in Choice Based Conjoint?" Presented at the ART Forum, Beaver Creek CO.

Kuhfeld, Warren, Randall D. Tobias, and Mark Garratt (1994) "Efficient Experimental Design with Marketing Research Applications," *Journal of Marketing Research,* (November), 545-57.

Louviere, Jordan (1988), "*Analyzing Decision Making: Metric Conjoint Analysis,"* Newbery Park, CA: Sage Publications.

Orme, Bryan K., Mark I Alpert and Ethan Christensen (1997) "Assessing the Validity of Conjoint Analysis—Continued," *Sawtooth Software Conference Proceedings* .

Pinnell, Jonathan (1994) "Multistage Conjoint Methods to Measure Price Sensitivity," Presented at the ART Forum, Beaver Creek, CO.

Pinnell, Jonathan and Sherry Englert (1997) "Design Considerations in Choice and Ratings-Based Conjoint" *Sawtooth Software Conference Proceedings.*

Payne, John W. (1976) "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance, 16,* 366-387.

Srinivasan, V. and Chan Su Park (1997) "Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement," *Journal of Marketing Research, 34,* (May) 286-291.

Wright, Peter and Mary Ann Kriewall (1980) "State of Mind Effects on the Accuracy with Which Utility Function Predict Marketplace Choice," *Journal of Marketing Research, 17,* (August) 277-293.

Zwerina, Klaus and Joel Huber, "Deriving Individual Preference Structures for Practical Choice Experiments," Working Paper, Fuqua School of Business, Duke University.