The assumptions of choice modelling: Conjoint analysis and SUMM

Eric Marder
Eric Marder Associates, Inc.

Conjoint analysis is based on the assumption that the values of product characteristics to customers cannot be measured directly and must be inferred from overall ratings of integrated offers; SUMM (Single Unit Marketing Model) is based on the opposite assumption. These assumptions are examined in the light of logical consistency and available empirical evidence.

The impact of potential marketing strategies on market share can be evaluated by three generic methods — marketing experiments, choice experiments, and choice models. These methods make progressively more assumptions and accept progressively less rigour in exchange for progressively greater capacity (ability to assess a large number of potential strategies) as Table 1 shows:

Table 1

Number of assumptions required		Capacity: Low	Capacity: Intermediate	Capacity: High
	Almost none	marketing experiment		
	Mainly one		choice experiment	
	Several			choice model

The demand by marketers for quick evaluations of many alternative strategies (changes in existing brands and/or definition of new ones) has resulted in a willingness to trade off rigour for capacity, which in turn has led to the widespread popularity of choice modelling. This article examines the core assumptions of choice modelling and of two particular methods of choice modelling, conjoint analysis and SUMM (Single Unit Marketing Model). It presents no new empirical data, but draws on the empirical evidence that is available in the literature. It explores whether, and to what extent, the assumptions of conjoint analysis and SUMM are sustained by logical coherence and by the empirical evidence. To facilitate this, it begins by putting choice modelling in its proper context, spelling out the relationships among the three generic methods, before going on to discuss the main subject.

Marketing experiments, choice experiments, and choice models

Conceptually, the most straightforward way of measuring the potential impact of marketing strategies is by a marketing experiment. Such an experiment requires randomly equivalent geographies in which randomly equivalent customers shop in randomly equivalent stores that carry the same array of competitive products at the same prices. This is easy enough to imagine on paper, but extremely difficult, if not altogether impossible, to implement in practice. The closest approximation is a direct-marketing experiment; but since the number of people who respond to direct marketing offers tends to be low (around 1%), such an experiment is appropriate only for products that are ordinarily marketed that way.

If one is willing to relax the requirement that the experiment be conducted under actual market conditions, one can use a choice experiment instead. This is any experiment that assesses the impact of different characteristics of a test brand on respondents' choices in a competitive frame. Provided that the competitive frame, the accessibility, and the respondents' information are the same in a test situation as in the market, the choices observed in the test situation will be the same as the choices (market shares) in the market. Assuming that this proposition, which has been formalized as the First Law of Choice, is true, one can treat a choice experiment as a surrogate for a marketing experiment.

Choice experiments can be conducted in many ways. One such experiment, for which a large amount of empirical evidence based on thousands of studies, is available, is STEP (Marder 1997). Each page of a STEP booklet is devoted to a different brand, and shows a picture, a price, and a brief statement summarizing the brand's principal benefits. The respondent divides ten adhesive labels among the brands, in effect making ten separate choices among them. The resulting shares have been shown to be highly predictive of market shares and of individual buying behaviour.

In the choice experiment, two or more randomly equivalent groups of respondents receive STEP booklets that are identical in all respects, except that one page, devoted to one brand (the "test" brand) is systematically varied from group to group. The picture may be different, or the name, or the statement, or the price, measuring the effect of the relevant stimulus on choice. If it is desired to evaluate physical products, the respondents in the different groups may be exposed to different products before receiving the STEP booklet. The particulars are unimportant. The only requirements are that:

- the respondent is exposed to all brands in the competitive frame;
- the test brand is not singled out;
- each respondent is exposed to a single test stimulus; and
- the respondent makes some choice among the brands of the competitive frame.

Suppose we want to evaluate the potential value of two brand names, A and B. We set up a choice experiment in which the test brand is presented in its full competitive frame, identified by name A in group 1 and B in group 2. When we observe shares of 12% and 15% in groups A and B, respectively, we know with certainty that name B's incremental three share points are causally attributable to the name because, aside from sampling error, the name was the only thing different between the two groups. Suppose we also want to know what difference, if any, might be deserved by offering the brand at ten calories per glass versus 150 calories per glass. All we have to do is split the samples. We now have four groups representing all combinations of names A and B, and low and high calories. Conceptually, this process can continue indefinitely, and will provide definitive information about the shares deserved by each of the variables studied, and by all interactions among these variables. The problem is that, as the number of variables increases, the size of the experiment grows exponentially and soon becomes totally impractical. Twenty dichotomous variables, for example, would require $2^{20} = 1{,}048{,}576$ experimental groups. It is at this point that the need for choice modelling arises.

Choice modelling came into marketing research in the early 1970s on two parallel tracks, conjoint analysis (Green & Rao 1971; Green & Wind 1973; Johnson 1974; Green & Wind 1975) and SUMM (Marder 1968, 1973, 1974) addressing essentially the same problem, but proceeding from different, in some respects incompatible, assumptions. Their history has also been quite different. Conjoint analysis, supported by a huge academic literature, has become the most widely known and used method of choice modelling. SUMM has been developed privately in the interaction between a commercial research organization and its clients, with no publications between 1974 and 1997, and has until recently been virtually unknown in the academic community.

A choice model allows the user to ask "What if" questions of the form "How much business (share) would I gain (lose) if I changed my brand, or offered a new brand, with characteristics A, B, C ...?" Accordingly, a choice model estimates the results that would have been obtained in a choice experiment, if it had been practically feasible to conduct that experiment. The choice model is not an end in itself, but rather a screening device. Conservatively, its results should be confirmed by a choice experiment.

How then does one evaluate a choice model? Since the choice model is used only because the number of variables to be studied is too large for a choice experiment, the appropriate criteria for assessing a choice model are:

- its capacity, or how flexible it is in allowing us to study a large number of variables, and
- its predictive power, or how well it predicts the results that would have been obtained from a choice experiment if it had been possible to do such an experiment.

Assumptions shared by conjoint analysis and SUMM

All choice models, including conjoint analysis and SUMM, make one fundamental assumption, the "partitioning" assumption, which asserts that the value of a product is an aggregation of the values of its characteristics. This assumption is inherent in the decision to construct a choice model to begin with, because choices that are not amenable to partition cannot be modelled. They can only be studied the hard way, by choice experiments.

Both conjoint analysis and SUMM begin by defining a "map" of characteristics in terms of which the products of a competitive frame are to be described. Ideally, this map includes all characteristics likely to influence a respondent's choice. Conjoint analysis and SUMM use different nomenclatures to refer to the elements of this map. In conjoint analysis, colour is called an "attribute"; and red, blue, and green are called "attribute levels". In SUMM, colour is called a "topic"; and red, blue, and green are called "attributes". To facilitate referring to both models simultaneously and to avoid confusion due to the different ways the term "attribute" is used by the two methods, I will call colour a "dimension", and red, blue, and green "characteristics" of that dimension.

Both conjoint analysis and SUMM use individual-by-individual analysis. In principle, choice modelling can be implemented either on the individual or on the aggregate level. From the outset, SUMM used individual analysis, which is embedded in its name, "Single Unit Marketing Model"; some early conjoint publications used aggregate analysis. Nowadays, conjoint analysis is "usually carried out at the individual level" (Green & Srinivasan 1990), so there is no conceptual difference between conjoint analysis and SUMM in this respect.

Both conjoint analysis and SUMM employ the same general approach to modelling. After the map has been created, the value to each respondent of each of the characteristics that comprise the map is measured or computed. The characteristics that describe the brands are entered into the model. For each respondent, each brand is credited with an aggregation of the values of its characteristics. This is usually a linear function (a sum) but could be some other aggregating function, for example a product. In the end, the brand that is most valuable to a respondent is designated as that respondent's chosen brand, yielding shares for the various brands in the competitive frame.

Notwithstanding these similarities there are fundamental differences between conjoint analysis and SUMM, both in the measurement methods they employ and in the assumptions on which they are based.

The measurement methods

The common denominator underlying the many variants of conjoint analysis has been defined as follows:

> Conjoint analysis is any decompositional method that estimates the structure of a consumer's preferences (i.e., estimates preference parameters such as part-worths, importance weights, ideal points), given his or her overall evaluations of a set of alternatives that are specified in terms of levels of different attributes (Green & Srinivasan 1990).

Typically, the major focus of conjoint analysis is on estimating the value to each respondent of the various characteristics that comprise the map. Towards that end, a number of "profiles" are constructed, each of which represents one possible combination of characteristics (one characteristic from each dimension), and the respondents are asked to evaluate these profiles in their entirety, either by ranking the profiles, rating them, or successively selecting the preferred profile from pairs of profiles, or some analogous process. In the full-profile method, the respondents rate every possible profile. Given a map of five dimensions, they rate $2^5 = 32$ profiles. Assuming that the overall ratings are a linear function of the implicit unknown values of the characteristics, these values are then inferred computationally from the overall ratings. The number of possible profiles increases rapidly as the number of dimensions increases, which has been a practical problem for conjoint analysis from the beginning. Defining conjoint analysis as a decompositional method is not merely a conceptual statement; it has practical consequences. It leads to data collection systems that become totally unimplementable as the number of dimensions grows large. This has been recognized by proponents of conjoint analysis:

> The full-profile method of conjoint analysis works very well when there are only a few (say six or fewer) attributes. As indicated by Green (1984) industrial users of conjoint analysis have strained the method-

> ology by requiring larger numbers of attributes and levels within attributes, thus placing a severe information overload on the respondents (Green & Srinivasan 1990).

Ironically, as long as the number of dimensions is relatively low neither conjoint analysis nor SUMM are really appropriate, because it is possible to conduct a choice experiment, which is preferable to any model. Once the number of dimensions becomes so large that a choice experiment is no longer feasible, practical considerations place an increasing burden on conjoint analysis as well.

The proponents of conjoint analysis have attempted to deal with this limitation in two ways. They have introduced "hybrid" models (Green, *et al.* 1981; Green 1984), which are only partially conjoint, in effect abandoning the insistence on decomposition. And they have explored mathematical techniques to reduce the total number of profiles to which the respondent must be exposed, "exploring the feasibility of estimating the heterogeneity of the part-worths when fewer profiles per subject are used as compared to more traditional methods" (Lenk, *et al.* 1996) These models can become complicated, but proponents of conjoint analysis forecast that they represent the wave of the future.

> We believe that the newer hybrid models along with ACA, BRIDGER, and the models listed above will play an increasing role, at least in large-scale industry applications (Green & Krieger 1996).

Thus the newer forms of conjoint analysis rely on mathematical complexity to overcome their fundamental capacity limitations of conjoint analysis. To the extent that they are intended to remain conjoint, however, they must infer the values of the various characteristics from overall ratings of stimuli defined by combinations of these characteristics. Once these values have been determined, they are entered into the model, together with objective inputs that define which characteristic of each dimension describes each of the brands. The model is then interrogated with "what if" questions.

Whereas conjoint analysis infers the values of characteristics from overall ratings of products, SUMM measures these values directly. SUMM grew out of an old idea, that of evaluating choices by weighing pros and cons. Benjamin Franklin formulated it explicitly in 1772, referring to it as a "moral or prudential algebra" (Dawes & Corrigan 1974). The gist of the idea was to make a list of factors, assign weights to them, and use the sum to evaluate options, a routine practice of people seeking to clarify their feelings in everyday life. This idea was the basis of brand choice studies at Kenyon and Eckhardt in the 1950s — articulated at that time as: "Brand choice is determined by desires for product characteristics and beliefs about the extent to which different brands possess these characteristics."

In the meantime, investigators from a range of disciplines were independently promulgating the same idea under such labels as multi-attribute utility model, subjective evaluation model, and multi-criteria model (Huber 1974), encompassing contributions from economics, management science, social psychology, marketing research, and related fields (Bass, *et al.* 1972; Dawes 1971; Fishbein 1963; Lancaster 1966; Ratchford 1975; Rosenberg 1956; Wilkie & Pessemier 1973). This convergence is not surprising. It is an everyday experience to describe the overall value of an object as the sum of (or more generally as a function of) the values of the characteristics attributed to it — leaving room for innovations in how the variables are measured and used.

The original version of SUMM measured the values of characteristics by a two-stage process. First, respondents rated all the characteristics of a dimension, assigning a +10 to their most preferred characteristic, and rating the remaining characteristics in relation to this top characteristic on a scale from +10 to –10. Then they divided 50 labels among the top characteristics of the various dimensions. This yielded an importance weight. The value of each characteristic was obtained by multiplying its rating by the importance weight of the dimension. This approach has been called "self-explicated" in the literature.

Quite a few self-explicated models, including the self-explicated stages of hybrid models, have appeared (Edwards & Newman 1982; Green 1984; Green *et al.* 1981; Green, *et al.* 1991; Huber 1974; Johnson 1991;

Srinivasan 1988; Srinivasan & Park 1997; Srinivasan & Wyner 1989). These models have been similar to each other (Green & Srinivasan 1990). Apart from the fact that SUMM measures both beliefs and desirabilities, they have also been quite similar to the way the values of characteristics were measured in the original version of SUMM (Marder 1974, 1997). SUMM has, however, evolved since then; each new version has become progressively simpler. The important turning point came with the creation of the unbounded write-in scale (Marder 1984).

One reason the early generations of SUMM, as well as other self-explicated models, typically called for multiplying the ratings of the characteristics by the importances of the dimensions was that ratings generated by numeric scales were not expected to be comparable from dimension to dimension. After the introduction of the unbounded write-in scale, which requires respondents to write an unrestricted number of Ls or Ds to report the strength of their likes or dislikes (Marder 1997), the hypothesis was proposed that this scale might be measuring not only the relative value of characteristics within dimensions but also across dimensions. If this proved to be the case, it would allow a final simplification of SUMM. It would become possible to eliminate the importance weights altogether and to measure the values of all characteristics, both within and across dimensions, in a single step. This proved to be the case and, on the basis of conclusive evidence (Marder 1997), the single-stage method of measuring the values of characteristics, called the absolute method, became the standard of SUMM.

The unbounded write-in scale has implications for a refinement that has been proposed for self-explicated models. Srinivasan (1988) used a conjunctive stage to remove "totally unacceptable" levels. There have been conflicting reports of the merits of such a stage (Green & Srinivasan 1990). The first SUMM study done in 1970 contained a similar provision, which was called the "veto". The respondents were strongly urged to use the veto sparingly; which they did, assigning the veto to only 5% of all characteristics. But the study had 34 dimensions and 170 characteristics. Given this scope, more than half of the 903 respondents registered at least one veto for each of the three principal brands, which rendered the veto useless for that study, and probably for any study using a large number of characteristics. Conceptually, Srinivasan was right in calling attention to the need for a conjunctive stage; in practice, however, even a veto is a matter of degree. A respondent may, for example, dislike the idea of a leather sofa, and may cross out "leather" in a preliminary conjunctive stage. But, given a sufficient number of compensating characteristics, she may, on balance, accept the leather after all. The crux of the matter is whether her negative feelings towards leather have been quantified properly. To the extent the unbounded write-in scale does this, it eliminates unacceptable characteristics automatically, and the issue is rendered moot.

In addition to the values of the characteristics, SUMM collects each respondent's beliefs about each brand. The value of a brand is then defined as the sum of the values of those characteristics which the respondent believes the brand possesses. The model is used by changing respondent beliefs in the computer, and determining the share gains or losses resulting from these changes.

Conflicting assumptions

Conjoint analysis and SUMM are based on two sets of conflicting assumptions, decomposition versus direct measurement and objective reality versus beliefs.

Decomposition versus direct measurement

Conjoint analysis assumes that people cannot directly report the relative value to them of different product characteristics and that it is, therefore, necessary to estimate these values by analytic decomposition of overall preferences of profiles —the decomposition assumption.

SUMM assumes that people can directly report the relative value to them of product characteristics, provided the measurements are made properly — the direct measurement assumption.

Hybrid models retain the label "conjoint analysis"while accepting inputs from self-explicated ratings on the ground that "these tasks are relatively easy to implement and proceed rapidly in comparison with conjoint tasks" (Green 1984). But if conjoint analysis cannot survive without assistance from the self-explicated measures it was designed to supplant, there can be only one reason for holding onto it, a strong belief that it is fundamentally more valid than self-explicated measures, and that the decomposition assumption is so superior conceptually that it is better to hold onto it, even if only in part, than to give it up altogether. This is, in fact, how the use of hybrid models is justified. "Hybrid models combine the ease of administration of self-explicated data with the greater realism afforded by decompositional models" (Green & Krieger 1996). This idea that the decomposition assumption somehow affords greater "realism" crops up repeatedly in the conjoint literature. In effect, it says that the preferences people have and the choices they make are not made up of separate pieces, but are an overall "gestalt" response to the product as a whole. Though this proposition is usually asserted as a self-evident truth, it is really an assumption about how people make choices, an assumption that may turn out to be true, but one which, far from being a proven fact, warrants closer scrutiny.

Before going on to examine this assumption, it will be helpful to digress briefly to clarify a point on which there is occasional confusion. Analysts sometimes think that, because conjoint analysis is based on overall preferences, it automatically takes interactions between dimensions into account, while self-explicated methods and SUMM do not. This is not so. Both conjoint analysis and SUMM treat the overall preference for a product as an aggregation of pre specified parts, or characteristics. To be sure, it is possible to include interaction effects in conjoint models, but this is not done ordinarily.

> It has been typical in conjoint studies to estimate only the main effects and assume away interaction effects. In certain cases, interaction effects, particularly two-way interaction effects, may be important ...
>
> Empirical evidence (Green 1984, Table 1) indicates that the model with interaction terms often leads to lower predictive validity-that is, the increased model realism obtained by incorporating interactions is small in comparison with the deterioration in predictive accuracy caused by including additional parameters (Green & Srinivasan 1990).

Conversely, two-way and three-way interactions are often incorporated into SUMM by means of compound dimensions. For example, expecting an interaction between luxury appointments and size of car, we might define a single "luxury/size" compound dimension with characteristics like:

- a big car with ordinary appointments
- a big car with luxury appointments
- a small car with ordinary appointments
- a small car with luxury appointments

Thus, both conjoint analysis and SUMM can tackle interaction effects when that is judged appropriate, each in its own way. This brings us back to the question at issue. Is the decomposition more realistic than direct measurement? This depends on how people make choices — on the basis of a single overall judgment or by piecing them together out of parts.

Studies of the detailed steps respondents take in selecting their preferred "brand" (profile) from a list of up to 12 profiles, each consisting of up to 15 characteristics, have shown that respondents do not evaluate profiles in their entirety. They eliminate some profiles on the basis of one or two important characteristics that have negative value to them, judgments that certainly qualify as "self-explicated". More importantly, they go on to make their final decisions by doing the very thing conjoint analysis has been designed to avoid; they assess the various characteristics approximately one at a time (Lussier & Olshavsky 1979; Payne 1976). Reflection indicates that it could hardly be otherwise. As the number of characteristics used to describe a product increase, respondents become unable to keep all of the enumerated characteristics in mind simultaneously. They are then forced to "analyze" the description, to break it apart into pieces, and to evaluate the pieces separately in order to arrive at a meaningful evaluation of the whole. Since the direct measurement of characteristics helps them do just that, this method may correspond more closely to how people actually make decisions than the ostensibly more "realistic" data gathering system of conjoint analysis.

We can confirm this by considering everyday experience. Suppose we are shopping for a home and are shown two houses. Do we instantly make one integrated judgment, saying, "I like this one better"? If one of the houses has obvious overriding advantages on all counts, we may indeed do that. But if the utilities of the two houses are more nearly matched, we are more likely to consider the dimensions one at a time. This house has the extra bedroom that will come in handy for guests. The other one is closer to the train station. This one is in a better neighbourhood. The other one has a larger back yard. As we consider our options, we sometimes draw up a list, explicitly assign a value to each characteristic, and add these values to determine which house gets the higher score, thus filling out a SUMM questionnaire in the ordinary course of life. If we had been dealing with a respondent in a survey, would it really have been more "realistic" to demand that she first decide which house she preferred overall? Presumably this judgment, together with other judgments of the same type, would have then become an input for estimating indirectly the relative value to her of the extra bedroom, the closeness to the railroad station, the size of the back yard — something she would have been willing and able to tell us to begin with if we had asked her to do so directly. This does not detract from the fact that some scaling instruments will yield better data for this purpose than others.

But aren't there cases in which overall preferences provide more appropriate information than self-explicated measurements of parts? Certainly. Overall preferences for the beauty of paintings, for example, are doubtless more meaningful than self-explicated ratings of different aspects of the paintings. But what then? The mathematics are neutral — it might be possible to compute values, or part-worths, for such dimensions as: the painting has reds, greens, blues, yellows; or it has large brush strokes, small brush strokes. This information, however, won't help us create more beautiful paintings, because the beauty of paintings is not an aggregation of parts in the first place. Freud put it this way:

> I can only tell you of my personal experience. When making a decision of minor importance, I have always found it advantageous to consider all the pros and cons. In vital matters, however, such as the choice of a mate or a profession, the decision should come from the unconscious, from somewhere within ourselves (Reik 1948).

We may suppose that Freud would have regarded the kind of decisions we deal with in choice modelling as decisions of "minor importance", to be handled by "considering all the pros and cons".

Following Freud, the situation can be summed up as follows: There are psychological judgments that are not amenable to modelling of any kind, either by conjoint analysis or by SUMM, because the choices cannot be partitioned. These cases can be studied only in their entirety, by choice experiments in which each respondent is exposed to a single test stimulus embedded in a full competitive frame. There are other cases in which the psychological total is an aggregation of parts — only these cases are amenable to modelling. The direct measurement assumption holds that, in these cases, our nervous system arrives at overall judgments by performing precisely the kind of summing that is simulated in SUMM, and that direct measurement rather than decomposition is more realistic. It is impossible to develop more than circumstantial evidence concerning how our nervous system actually processes information, and is not necessary or even appropriate to try to settle this issue on physiological grounds. Instead we must accept that neither theory can be proven and must ask ourselves which is more likely to prove useful in practice.

Objective reality versus beliefs

Conjoint analysis assumes that products (brands) can be described adequately in terms of their objective characteristics without recourse to beliefs that vary in the eyes of the beholder — the objective reality assumption.

SUMM assumes that a brand's characteristics are not objective "facts", that they depend on what people believe about the brand, and that this is not only true for subjective characteristics like styling, reliability, or user friendliness, but also for so-called objective characteristics like the weight of a computer notebook or the fat content of a food — the beliefs assumption.

There is a fundamental difference between the way conjoint analysis and SUMM have framed the problem. For conjoint analysis, the principal challenge has been to estimate the value, or part-worth, of product characteristics. In pursuit of this objective, conjoint analysis ignores respondent beliefs and treats product characteristics as objective facts, uniformly applicable to all respondents. For SUMM, measuring the value of product characteristics is only a means to an end, an essential element but not the sole one, because brand choice depends not only on what people want but also on what they believe. By definition, brand characteristics in SUMM differ among respondents and are whatever the respondent believes them to be. In theory, beliefs could be incorporated into conjoint analysis; in practice, capacity limitations make this difficult.

If one is primarily interested in product development and poses the question, "What characteristics should I build into my product to increase the probability that people will buy it?", it may appear that the question deals exclusively with objective characteristics of the product — with how to make it, big or small, fast or slow. Beliefs can be left for later, for the advertising people. The R&D people don't manufacture beliefs. Before one can speak of beliefs, there must be a product to believe things about. This product must be real, objective, with specific characteristics. And the purpose of the choice model is to tell us what these characteristics should be.

Coming from an end user, this may seem to be a sensible request, but a model builder who takes the request literally falls into a trap, because the proposed new brand will not be launched not into a vacuum, but into a market made up of competitive brands. Each of these brands has some characteristics, and the potential share the new brand will capture will depend both on its own characteristics and on those of the other brands in the competitive frame. It could happen, for example, that there is much greater customer interest in a big luxury car than in a small luxury car, but if the market already has ten big luxury cars, a new small luxury car might capture a larger share than a new big luxury car. Thus all characteristics are important, not only those of the proposed new brand, but also those of the other brands. But what are these characteristics — can they be defined by laboratory data, by engineering specifications? When it comes to estimating which brand a respondent will choose, those specifications are irrelevant. The respondents' choice will not depend on what the specifications actually are, but on what they believe them to be.

There are times when beliefs can be expected to correspond so closely to objective facts that it is not necessary to measure them explicitly; for example, it may be safe to assume that 100% of all respondents would report that a four-door model automobile has four doors. If it is desired to limit an entire study to such "objective" characteristics, as is customary in conjoint analysis, SUMM can accommodate such a requirement by simply entering "answers" to the beliefs questions into the model as though these questions had actually been asked. The moment we are dealing with such characteristics as safety, reliability, pick-up, and ease of handling, however, this is no longer appropriate. In the general case, we are not entitled to assume either that the respondents' beliefs correspond to anything that can be specified objectively or that these beliefs will be held uniformly by all respondents. Instead, SUMM usually provides for measuring each respondent's beliefs, and then defines the value of a brand for a particular respondent as the sum of the values of those characteristics which that respondent, correctly or incorrectly, believes the brand to possess. Only those characteristics drive the choice.

Empirical evidence

Theory can be debated indefinitely; in the end, the only arbiter is empirical evidence. The more strongly conflicting hypotheses are held, the more rigorous that evidence must be to convince one side or the other to revise its conclusions. In this case, truly rigorous evidence would require three randomly equivalent groups — one devoted to conjoint analysis, one to SUMM, and one to a choice experiment in which a large number of potential new products are measured systematically to provide the criterion against which the predictions of the two models are judged. Such an experiment would have to be huge, and might be, in view of the intrinsic differences between conjoint analysis and SUMM, difficult to implement properly. Without such an experiment, we must turn to what we have, separate efforts to investigate the validity of conjoint analysis and SUMM.

These efforts have generally made compromises, falling short of "proof" and settling for "substantiating evidence". There are anecdotal success stories for both conjoint analysis (Benbemsty 1983; Page & Rosenbaum 1987) and for SUMM (BonDurant 1991). And there are ample demonstrations of internal consistency. In the case of conjoint analysis, this has often involved demonstrating the ability to predict choices of "hold-out" profiles. In the case of SUMM, it has involved demonstrating that the brand shares obtained from the model have high correlations with first choice share or use share. Evidence of this type is necessary but not sufficient for establishing the validity of a model. A number of formal experiments, mostly based on small samples, have demonstrated predictive power for conjoint analysis (Anderson & Donthu 1988; Leigh, *et al.* 1984; Mohn 1990; Srinivasan & Hartley 1981; Srinivasan & Park 1997; Wright & Kriewall 1980). And while there have been some reports of greater validity for conjoint or hybrid than for self-explicated methods (Green & Krieger 1996), other studies have shown that various, not necessarily optimal, self-explicated methods are as valid or more valid as than conjoint analysis (Wright, *et al.* 1984; Srinivasan 1988; Green, *et al.* 1991; Srinivasan 1996; Srinivasan & Park 1997). Green and Srinivasan summarized the accumulated evidence (1990): "The empirical results to date indicate that the self-explicated approach is likely to yield predictive validities roughly comparable to those of traditional conjoint analysis."

None of these studies dealt with SUMM; their authors were probably not even aware of its existence. We do, however, have a definitive experiment that investigated the predictive power of SUMM (Marder 1997). This experiment involved comparing the share predictions of SUMM with the corresponding shares obtained from a choice experiment STEP. The choice experiment used 32 products (16 different physical products, each at a high price and at a low price), all labelled alike in a decoy package, sent to the respondents as an unfamiliar but otherwise "real" brand. The respondents were divided into 32 randomly equivalent groups of approximately 207 respondents each (final n=6633). Each group received one of the test products as a free sample, followed three weeks later by an ostensibly unrelated questionnaire. The resulting choice shares ranged from a low of 3% to a high of 17%, demonstrating discrimination among the 32 test groups. Of the 496 possible pair-wise comparisons, 58% were significant at better than the 95% level. Thus we know that the "deserved shares" of the 32 products, measured under stringent experimental conditions, were very different.

Part 2 of the interview consisted of a SUMM questionnaire, covering 25 dimensions. This generated a SUMM model of the current eight-brand market (excluding the test brand) based on 6633 respondents. The model was used to test SUMM's ability to predict the results of the choice experiment. To do this, 32 different "new" products were created in the computer, one at a time, each representing one of the 32 test products. Beliefs about each test product were known from the sub-sample that had received that product. It was, therefore, possible to use SUMM precisely the way it is ordinarily used in a practical application, posing "what if" questions of the form: "What share would I get if I could create a product that people will come to believe has characteristics X and Y and Z ... etc.?" This question was posed 32 times, simulating a different new product each time, generating SUMM predictions for 32 different products. These were compared to the shares these products had actually obtained in the choice experiment. The correlation was .88 (Marder 1997). We don't know how conjoint analysis would have performed if a randomly equivalent conjoint sample had been available, and we must allow the possibility that conjoint analysis might have shown higher predictive power than SUMM in such a direct comparison — keeping in mind, however, that there is not too much room for improvement between .88 and 1.0.

Conclusions

We said at the outset that choice models must be evaluated on two criteria — capacity and predictive power. SUMM has greater capacity than conjoint analysis. It is simpler; it is easier and faster to administer; it can handle a larger number of characteristics; it can accommodate both the special case when it is desired to study only objective characteristics, and the general case when "characteristics" are defined as beliefs that vary among respondents.

Conjoint analysis claims greater predictive power on the ground that it offers greater realism. This claim is neither self-evident nor supported by the empirical evidence. To the extent empirical studies have compared conjoint analysis with various (not necessarily optimal) self-explicated methods (other than SUMM), the evidence has shown comparable predictive power for both. In addition, a major experiment has demonstrated high predictive power for SUMM, though without a direct comparison to conjoint analysis. This places the burden of proof on conjoint analysis to demonstrate higher predictive power in an experiment of comparable scope and rigour. In the meantime, there is no compelling reason either conceptual or empirical for giving up SUMM's greater simplicity and capacity for the sake of the "realism" claimed for conjoint analysis.

References

Anderson, J. C. & N. Donthu (1988) A proximate assessment of the external validity of conjoint analysis. *AMA Educators' Proceedings*, G. Frazier, *et al*, eds. Series 54, Chicago: American Marketing Association: 87–91.

Bass, F. M., E. A. Pessemier, & D. R. Lehmann (1972) An experimental study of relationships between attitudes, brand preference, and choice. *Behavioral Science*, 17 (Nov.): 532–41.

Benbemsty, R. L. (1983) Attitude research, conjoint analysis guided Ma Bell's entry into data terminal market. *Marketing News* (May 13): 12.

BonDurant, W. R. (1991) Marketing with choice models. *Second Annual Advanced Research Techniques Forum*, 1-6. American Marketing Association, June 16, Beaver Creek, Colorado.

Carroll, J. D. & P. E. Green (1995) Psychometric methods in marketing research: Part I, Conjoint analysis. *Journal of Marketing Research*, 32(Nov.): 385–91.

Cattin, P. & D. R. Wittink (1982) Commercial use of conjoint analysis: A survey. *Journal of Marketing*, 46 (Summer): 44–53.

Dawes, R. M. (1971) A case study of graduate admissions: Applications of three principles of human decision making. *American Psychologist*, 26.

Dawes, R.M. & B. Corrigan (1974) Linear models in decision making. *Psychology Bulletin*, 81: 95–106.

Edwards, W. & J. R. Newman (1982) *Multiattribute Evaluation*. Beverly Hills: Sage.

Fishbein, M. (1963) An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human Relations*, 16: 233–40.

Green, P.E. (1984) Hybrid models for conjoint analysis: An expository review. *Journal of Marketing Research*, 21 (May): 155–9.

Green, P. E., S. M. Goldberg, & M. Montemayor (1981) A hybrid utility estimation model for conjoint analysis. *Journal of Marketing*, 45 (Winter): 33–41.

Green, P.E., F. J. Carmone, & Y. Wind (1972) Subjective evaluation models and conjoint measurement. *Behavioral Science*, 17: 288–99.

Green, P.E. & A. M. Krieger (1996) Individualized hybrid models for conjoint analysis. *Management Science*, 42(6).

Green, P.E., A. M. Krieger, & M. J. Agarwal (1991) Adaptive conjoint analysis: Some caveats and suggestions. *Journal of Marketing Research*, 28 (May): 215–21.

Green, P.E. & R. Rao (1971) Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8 (Aug.): 355–63.

Green, P.E. & V. Srinivasan (1990) Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54: 3–19.

Green, P.E. & Y. Wind (1973) *Multiattribute Decisions in Marketing: A Measurement Approach*. Hinsdale, IL: The Dryden Press.

——— (1975) New way to measure consumers' judgments. *Harvard Business Review*, 53 (July/Aug.): 107–17.

Huber, G. P. (1974) Multiattribute utility models: A review of field and field-like studies. *Management Science*, 20: 1393–402.

Johnson, R. M. (1974) Trade-off analysis of consumer values. *Journal of Marketing Research*, 11 (May): 121–7.

——— (1991) Comment on "Adaptive conjoint analysis, some caveats and suggestions. *Journal of Marketing Research,* 28 (May): 223–5.

Koestler, A. (1959) *The Sleep Walkers*. New York: Penguin Books.

Lancaster, K. J. (1966) A new approach to consumer theory. *Journal of Political Economy*. 74: 132– 57.

Leigh, T. W., D. MacKay, & J. O. Summers (1984) Reliability and validity of conjoint analysis and self-explicated weights: A comparison. *Journal of Marketing Research,* 21 (Nov.): 456–62.

Lenk, P. J., W. S. DeSarbo, P. E. Green, & M R. Young (1996) Hierarchical Bayes conjoint analysis: Recovery of part-worth heterogeneity from reduced experimental designs. *Marketing Science,* 15(2), 173–91.

Lussier, D. A. & R. W. Olshavsky (1979) Task complexity and contingent processing in brand choice. *Journal of Consumer Research,* 6 (Sept.): 154–65.

Eric Marder Associates, Inc. (1974) *The Finances of the Performing Arts, Vol. II, A Survey of the Characteristics and Attitude of Audiences for Theater, Opera, Symphony and Ballet in 12 U.S. Cities*. New York: The Ford Foundation.

Marder, E. (1968) SUMM (Single Unit Marketing Model). *Advertising Research Foundation Conference,* October.

——— (1973) SUMM (Single Unit Marketing Model). *American Marketing Association Conference, Attitude Research across the Sea,* Winter.

——— (1984) A scale for measuring attitude and opinions. *American Association of Public Opinion Research Conference,* May.

——— (1997) *The Laws of Choice: Predicting Customer Behavior*. New York: The Free Press.

Mohn, N. C. (1990) Simulated purchase chip testing vs. tradeoff (conjoint) analysis — Coca Cola's experience. *Marketing Research,* 2 (Mar.): 49–54.

Page, A. L. & H. F. Rosenbaum (1987) Redesigning product lines with conjoint analysis: How Sunbeam does it. *Journal of Product Innovation Management,* 4: 120–37.

Payne, J. W. (1976) Task complexity and contingent processing in decision making: An information search and protocol analysis *Organizational Behavior and Human Performance,* 16, 366–87.

Ratchford, B. T. (1975) The new economic theory for consumer behavior: An interpretive essay. *The Journal of Consumer Research,* 2: 65–78.

Reik, T. (1948) *Listening with the Third Ear*. New York: Grove Press.

Rosenberg, M. J. (1956) Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology,* 53: 367–72.

Srinivasan, V. (1988) A conjunctive-compensatory approach to the self-explication of multiattributed preferences. *Decision Sciences,* 19 (Spring): 295–305.

——— (1996) Conjoint analysis and the robust performance of simpler models and methods. *American Marketing Association's Annual Marketing Research Conference,* September.

Srinivasan, V. & P. G. Hartley (1981) Forecasting the effectiveness of work trip gasoline conservation politics through conjoint analysis. *Journal of Marketing,* 45 (Summer): 152–72.

Srinivasan, V. & C. S. Park (1997) Surprising robustness of the self-explicated approach to customer preference structure measurement. *Journal of Marketing Research,* 34 (May): 286–91.

Srinivasan, V. & G. A. Wyner (1989) CASEMAP: Computer-assisted self-explication of multi-attributed preference. In Henry, W., M. Menasco, & H. Takanada (eds.) *New Product Development and Testing*. Lexington MA: Lexington Books: 91–111.

Wilkie, W. L. & E.A. Pessemier (1973) Issues in marketing's use of multi-attribute attitude models. *Journal of Marketing Research,* 10: 428–41.

Wright & Kriewall M.A. (1980) State-of-mind effects on the accuracy with which utility functions predict marketplace choice. *Journal of Marketing Research,* 17: 277–93.

About the author

Eric Marder is chairman of Eric Marder Associates, Inc., a marketing research and consulting firm he founded in 1960. The company specializes in "Choice Research," a term Marder uses to distinguish the discipline presented in his book *The Laws of Choice* from traditional marketing research. His client list through the years has included American Home Products, AT&T , Campbell Soup, CBS, General Foods, General Mills, GTE, Hewlett-Packard, Johnson & Johnson, Pfizer, Scott Paper, Xerox, and others.

Acknowledgments

The author is indebted to the editor and to an anonymous reviewer for valuable comments on an earlier version of this article.