9-599-112 REV: DECEMBER 12, 2001

ROBERT J. DOLAN

Analyzing Consumer Preferences

Introduction

Consumer preferences are at the heart of marketing. When we analyze consumer behavior, we are typically assessing how consumers make purchase decisions (i.e., the process via which they come to value one purchase alternative over another). Understanding of consumer preferences is particularly important for product policy (e.g., what features to have, whether or not to offer a new product) and pricing decisions.

Two procedures with proven utility for the actionable analysis of consumer preferences are

- 1. Concept testing,
- 2. Conjoint analysis.

A concept test is very straightforward: consumers are presented with a product idea and directly asked for their reaction (e.g., how likely would you be to buy this product?). We describe this type of testing and provide examples in Section I. While useful in many situations, the standard concept test has some important limitations. To a large extent, these limitations relate to the diagnostic information provided. If consumers collectively rate a product concept poorly, we know we should not launch the product—but the real question we want input to is how to fix it so that we do have a product that is acceptable to the marketplace. The consumer's reaction to a product (e.g., I am likely to buy it or I'm not) is a reflection of the consumer's underlying preferences.

Conjoint analysis, described in Section II, is a set of procedures developed to overcome a fundamental limitation of concept tests. In conjoint, we reorient our efforts not to look at reactions to a product idea per se but to get insight into the underlying preferences. The development of software to facilitate the consumer questioning and data analysis to achieve this result has been an active area of research—both by market research practitioners and academics. Conjoint is a staple of market research firms' offerings, and several firms specialize in conjoint applications. Thousands of conjoint studies are done each year in product categories ranging from hotels, rural health care systems, and cellular telephones to blue jeans. Section II describes this method and includes example applications. While not focusing on statistical details, we provide some intuition for the data analysis procedures.

Professor Robert J. Dolan prepared this note as the basis for class discussion.

Copyright © 1999 President and Fellows of Harvard College. To order copies or request permission to reproduce materials, call 1-800-545-7685, write Harvard Business School Publishing, Boston, MA 02163, or go to http://www.hbsp.harvard.edu. No part of this publication may be reproduced, stored in a retrieval system, used in a spreadsheet, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the permission of Harvard Business School.

Section I: Concept Testing

Standard Concept Tests are widely used. For example, Colgate-Palmolive¹ faced the issue of whether or not to introduce a new toothbrush and how to position it. Research was conducted presenting mock advertisements conveying various positioning alternatives to consumers. Consumers reacted to the offerings on a five-point "purchase intention" scale, marking one box:



The best of the Colgate options found 87% of consumers rating the concept in one of the "top two boxes." This information helped to indicate the potential for the product.

In a similar vein, when BIOPURE² received FDA approval for a blood substitute for dogs, it conducted a survey in which the product was described to veterinarians who were then asked if they "would try" the product in critical and noncritical cases. At a price of \$100, 95% reported being willing to try for critical cases and 70% for noncritical.

Executing a Concept Test

Designing a concept test presents the usual survey design issues of what sample size to have and how to select respondents. In addition to this, the two key executional decisions in a concept test are

- 1. How to communicate the concept,
- 2. The data to collect from respondents.

On the first issue, Colgate chose to present a "positioning concept" (i.e., the product concept was presented in persuasive form by showing consumers mock advertisements for the Precision toothbrush). An alternative is to state the core idea only without an accompanying marketing message (e.g., "Precision is a new toothbrush with bristles of varying lengths resulting in 35% more plaque removal."). There is no general rule on "core idea" vs. "positioning concept" being the better choice for concept testing. Generally, using a "positioning concept" approach gives better prediction of actual marketplace reaction because there is a closer match of what the respondent sees to what will be seen in the actual purchase situation. The caveat on "positioning concepts" is that the response obtained is a reaction both to the product and the quality of the accompanying presentation

¹ "Colgate-Palmolive Company: The Precision Toothbrush," HBS Nos. 593-064 or 499-082 (condensed version).

² "BIOPURE Corporation," HBS No. 598-150.

The second issue is the data to collect. Colgate and BIOPURE typify concept testing in that some form of purchase intention data were collected. This is often augmented by three other forms yielding the set:

- 1. Intended Purchase Measures
- 2. Overall Product Diagnostics
- 3. Special Attribute Diagnostics
- 4. Respondents Profiling Variables

Data Type #1: Purchase Measures Purchase measures include likelihood of purchase and expected amount. Purchase intention is captured through questions like "Based on this product description, how likely would you be to buy this product if it were conveniently available?"; check one:

- Definitely would buy
- □ Probably would buy
- □ Might or might not buy
- □ Probably would not buy
- □ Definitely would not buy

While this five-point scale is most common, six-, seven-, and eleven- point scales are also used.

For nondurable goods, the frequency of purchase is also key. Purchase intent is a good indicator of trial, but forecasting volume sold requires knowing whether the product will be part of someone's everyday consumption habit or a special-occasion item. The expected purchase incidence question adds this dimension. Again, there is a variety of ways to specify this question but generally it takes a form such as "Which statement best describes how often you think you would buy this product if it were conveniently available to you?"

- Once a week or more often
- Once every two or three weeks
- Once a month
- Once every two to three months
- Once every four to six months
- Less often
- Never

In cases where the product may come in different sizes or is such that multiple units might be purchased at one time, respondents are probed on these issues as well.

In summary, given

Sales volume per potential user in time period = % of potential users in market

who try product

* Expected number of purchases in the period for triers

* Expected number of units per purchase

the purchase measures from a concept test typically are designed to measure the three variables on the right-hand side.

Data Type #2: Overall Product Diagnostics Diagnostic data give insight into why the purchase data turned out the way it did. With respect to the overall concept, tests usually assess the product's perceived uniqueness (e.g., On a 1-5 scale where 1 = very similar and 5 = quite distinct, how would you rate the product relative to ones currently on the market?) and believability (i.e., Does the respondent believe the product can do what it claims?). For example, can the Colgate Precision toothbrush remove 35% more plaque?

Since a high-uniqueness, high-believability concept could still generate low purchase interest, firms usually assess how salient the product is to solving a consumer's problem and its overall interest to the consumer. For example, while a respondent may rate a television permitting the viewing of three channels at once as both unique and believable, purchase interest may be low because the respondent does not view the current channel constraint as a problem.

Data Type #3: Specific Attribute Diagnostics When a concept has a number of attributes or benefits offered, it is useful to probe which attributes/benefits significantly contribute to or distract from the purchase intention. One method is the use of open-ended questions such as "you said that you [state respondent's answer to purchase intention question]. What is it specifically about the product that makes you feel this way?"

A second approach is to collect data on perceptions of specific attributes and their importance to the consumer. For example, we might ask respondents exposed to a new Internet Service concept for data on perception and importance scales as follows:

	Excellent		0	Poor
Entertainment Value				
Educational Value				
Ease of Site Navigation				
	Very Important		2	Not at all important
	Important			important
Entertainment Value				
Entertainment Value Educational Value				

Data Type #4: Respondents Profiling Variables The final set of variables useful in analyzing concepts is the type of consumers who respond in different ways. The most obvious of these is demographics, which help in targeting efforts, but other more innovative data collection can be useful as well, for example, data on

- Current purchase behavior,
- Perception of the category,
- Satisfaction with current brands used,
- Influence in actual purchase decision.

For example, it might be important to understand how satisfied those with high purchase-intent scores are with their current brand. High satisfaction with the current brand makes a switch to a new brand less likely.

Interpreting the Purchase Intent Data

The Purchase Intent score is at the heart of a concept test. How does one best interpret these data? Colgate Precision got an 87% "top-two box" score; is this good or bad? If the brush was introduced, what sales volume could Colgate expect? These are two important, logical questions.

General rules of thumb on "good" purchase-intent scores exist. For example, Taylor, Houlahan, and Gabriel³ claim that, based on their experience with over 100 brands in many product categories, ". . . a concept statement should receive 80% to 90% favorable answers ["I definitely will buy" or "probably will buy"] to encourage subsequent development work." Schwartz⁴ states the following average scores for concept tests across all product categories:



Adding these two gives an 83% favorable rating score—a number not inconsistent with the rule-ofthumb of Taylor, Houlahan, and Gabriel. However, Schwartz also makes the important point that average scores vary appreciably across product categories. For example, he presents data on four categories' average "definitely will buy" scores as shown in **Figure A**.

Figure A Average "definitely will buy" percent—Across all categories and in four specific categories



³ J. Taylor, J. Houlahan, and A. Gabriel, "The Purchase Intention Question in New Product Development: A Field Test," *Journal of Marketing*, January 1975, pp. 90-92.

⁴ D. Schwartz, *Concept Testing*, AMACOM, 1987.

Thus, while Taylor et al.'s "rule-of-thumb" may be a useful first cut in assessing the "goodness" of the purchase intent scores, it is only that. The variation in scores across categories shows the need to have category specific norms or benchmarks. These norms can come from (i) published sources (such as Schwartz), (ii) the company's own files, or (iii) the files of the research company hired to do the concept test. Helpful information from published sources is very limited. The second source may suffice for an active company regularly introducing products into the same categories. Generally, however, there is important value in the benchmarks established by research firms with a broad array of clients participating in many product categories. BASES Worldwide, the largest concept testing firm, has done over 10,000 concept testings and hence has a valuable database to assist in interpreting results.⁵

Section II: Conjoint Analysis

Concept tests have had a long history of use in marketing and continue to be a viable research option in many situations. Testing is relatively inexpensive, the results and methods are easily understood, and the benchmarks developed over time help in interpreting results.

However, in many situations today, the product design question is of a level of complexity that overruns the capabilities of the standard concept test. The issue is not how many people will intend to buy my 233 MHz personal computer, but how many more would be willing if I made it with 300 MHz? Suppose I had to charge \$200 more for it? Suppose to offset the cost of speed, I downgrade the screen size? Made the product heavier? Reduce the warranty length? Conjoint analysis gives us a way to answer these critical questions. It has been used in a wide variety of product categories to deal with these issues of the augmented product design (i.e., product features and other value-adds like brand name and warranty coverage) and pricing. **Table A** presents a representative list.⁶

Of course, we can't get the insights conjoint provides for free. It is more difficult to develop an indepth-enough understanding of the technique to be a responsible, productive user of it. It is more expensive and time consuming to do a conjoint study than a concept test. But, its record of successful use and its taking a place in the repertoire of first-rate marketers make exploration of it more than worthwhile.

The idea behind conjoint is simple. Think about a product category in which the hierarchy of effects is of the Learn \rightarrow Feel \rightarrow Do type (e.g., buying a new PC, enrolling in a health club, subscribing to an information service). In product categories with a high-cognitive front end, we often look at a product as a bundle of attributes. An individual's "value system" is simply how much value the person would put on each level of the attributes. That's what we need to know to dig into the kind of questions noted above on how to trade features off against each other.

The problem is that if we walked up to a consumer and said, "Please tell me your value system," he or she probably could not do it even if they wanted to. In conjoint, we get around this by asking the consumer a series of questions he or she can more easily answer (e.g., which would you prefer: a Dell running at 233 MHz for \$2,000 or a Packard-Bell, running at 300 MHz for \$1,700?) and we let statistic procedures do the hard stuff (i.e., go from these answers to an estimate of the underlying "value system").

Table A

⁵ P. Green, A. Krieger, T. Vavra, "Evaluating New Products," *Marketing Research*, Winter 1997.

⁶ Drawn from R.J. Dolan, "Managing the New Product Development Process," HBS No. 592-011, and P. Green, *et al.*, "Evaluating New Products."

- Automobiles
- Cameras
- Cellular telephones
- Computers
- Condominium design and pricing
- Food processors
- Snowmobiles

Consumer Nondurables

- Blue jeans
- Rug cleaners
- Shampoos

Consumer Services

- Credit cards
- Rural health care systems
- Hotels
- Railway pricing

Industrial Goods

- Lift trucks
- Material requirements planning systems
- Copiers

The trick in conjoint is that, via construction of the value system, we bootstrap ourselves up from asking about preferences on a small subset of products to being able to make predictions about relative preference for *any* products with these attributes. This point will become clearer as we go along. First, we consider how one can calculate a "value system" from some overall judgments.

To get a sense of how it works, let's take an example. Consider a fitness facility interested in optimal design of its locker rooms. To keep things simple, let's say there are only two attributes potentially important to users: (i) whether or not there is a sauna and (ii) the size of available lockers. There are two alternative "levels" for the sauna ("yes" and "no") and three levels for lockers:

- a. Small (20" × 20" × 20") storage lockers permanently assigned plus large hanging ones (72" × 20" × 20") for daily use.
- b. Mid-size only (36" x 20" x 20") permanently assigned.
- c. No permanently assigned locker; hanging locker (72" x 20" x 20") available on daily basis with mirror inside door.

There are thus $2 \times 3 = 6$ different sauna/locker combinations or products. One might in practice ask individuals how important these alternative attributes are. Alternatively, one can simply ask the respondent to rank order the 6 possible combinations from *most* to *least preferred*. The individual might respond as follows:

	Sauna		
	Yes	No	
(SM) Small storage, large daily	Rank 2	Rank 4	
Locker (MED) Medium storage only	Rank 1	Rank 3	
(DAY) Large daily with mirror only	Rank 5	Rank 6	

With these ranks, we can ask the respondent to rate the desirability of the products, ranging from *least desirable* (a score of 0) to *most* (a score of 100). Suppose we are given the following ratings:

Sauna Yes No 599-112



Since each locker size is rated with both levels of the sauna attribute, we can calculate the utility of an attribute level as the average of the score across all choices where it appears. Following this, we would have:

Sauna:) (oo =	
	Yes = 66.7	
	No = 33.3	
Locker:		
	SM = 60	
	MED = 80	
	DAY = 10	

This is the individual's "value system." Note that it recaptures the stated original ranking data:

Product	Value System Score	Value System Score Rank	Stated Original Rank
MED + Sauna	80 + 66.7 = 146.7	1	1
SM + Sauna	60 + 66.7 = 126.7	2	2
MED + No Sauna	80 + 33.3 = 113.1	3	3
SM + No Sauna	60 + 33.3 = 93.3	4	4
DAY + Sauna	10 + 66.7 = 76.7	5	5
DAY + No Sauna	10 + 33.3 = 43.3	6	6

With this value system, we can get an idea of how important the two attributes are to the consumer: the highest-rated locker option has 80 points and the lowest has 10 for a difference of 70. The sauna differential is only 33.3, suggesting the sauna attribute is less important than the locker attribute.

In practice, obviously, things are more complicated. We have lots more attributes and we use a multiple regression-type procedure to go from overall judgments to estimated value system. Conjoint is built on the idea of something called a part-worth model. It says, if a product in a category has n attributes, then the utility of an object i in the category is:

$$U_{i} = U_{i1} + U_{i2} + U_{i3} + \dots U_{in}$$

$$\uparrow \qquad \uparrow \qquad \uparrow$$
Utility of an object *i*
to consumer
$$U_{i1} + U_{i2} + U_{i3} + \dots U_{in}$$

$$\uparrow$$
Utility of object *i*'s
level of attribute #1
to consumer
$$U_{i1} + U_{i2} + U_{i3} + \dots U_{in}$$

That is, to get the utility of an item, we just sum up over all its attributes. The idea of conjoint is that consumers have a pretty good idea of the things on the left-hand side of the equation (i.e., we can ask them about that and use that pretty reliable information to estimate the stuff on the right-hand side).

Now we will go through a real application to see how it works in practice. This is an actual study⁷ for a German automobile company to design and price its new model, code-named LION. LION would be positioned in the marketplace against models from two competitors—another Germanbased company and a Japanese company.

Step 1 Choose Attributes

The first step in a conjoint study is to specify the possibly relevant attributes. Based on past research and its wealth of experience in the category, management specified five key attributes:

- 1. Brand name
- 2. Engine power
- 3. Fuel consumption
- 4. Environmental performance
- 5. Price

(A preliminary research stage is sometimes necessary to elicit possible relevant attributes from consumers, e.g., if this is a new product category for the firm.) Note the important capability that conjoint can handle a mix of hard, tangible features like engine power and fuel consumption, and intangibles like brand.

Step 2 Choose Relevant Levels of Attributes

Determine the relevant levels of the attributes that consumers should be asked to evaluate. In this case, we specified the same number of levels (three) for each of the attributes. This need not be the case, however. Conjoint can accommodate any practical number of attribute levels for an attribute. The following attribute levels were used:

• Brand

- LION, "German" and "Japanese" (in the study, the actual names of the "German" and "Japanese" brands were used; but, for confidentiality reasons, we use those terms here)

- Engine Horsepower
 - 150 HP, 200 HP, and 250 HP
- Fuel Consumption

- 12, 14, and 16 liters per 100 Km

Environmental Performance

- (i) fulfills minimum requirements, (ii) exceeds minimum requirements, and (iii) sets new standards in environmental performance

Price (Deutschmarks/DM)

- 50,000, 60,000, and 70,000

⁷ This study was reported on in R.J. Dolan and H. Simon, *Power Pricing* (New York: Free Press, 1996).

Respondents were given a detailed description of the Environmental Performance variable and the above listing is a shorthand representation.

Step 3 Choose a Sample Size and Respondent Type

Respondents were prescreened for interest in buying an automobile in the 50,000 to 70,000 DM range within a specified time horizon.

Step 4 Choose Response Task and Survey Administration Mode

In this case, a pairwise comparison approach was used. This method describes two alternatives on all five dimensions and asks the consumer for a preference judgment between the pair. The survey was administered on a laptop computer. An example screen is as follows:



If you prefer A, press A; if you prefer B, press B.

The alternatives are set up so the consumer has to trade off one thing to get another. On this screen, LION is markedly better than "Japanese" on Engine Power but performs in an inferior fashion on environmental standards, fuel consumption, and price. Depending on his or her preferences, the respondent makes a choice. The computer software then produces another choice to be made. Because the interview was done on a laptop, the program "learns" the consumer's preferences as it goes and so can adapt the questions to zero in on areas of uncertainty. Usually between 15 and 20 choice comparisons were needed to calibrate the underlying value system. Interview times varied between 30 and 60 minutes.

Step 5 Compute Individual Customer Value Systems

A strong point of conjoint is that value systems are estimated at the level of the individual respondent. There is no presumption that respondents have the same value system. As such, conjoint can be a useful method for defining market segments based on differing attributes of importance.

Step 6 Analyze the Data

(a) Attribute Level Values Often it is useful to look at average values of attributes across the full respondent set to get a general sense of the market.

Table B gives the results for this study. The five panels give the results for each attribute individually. For each attribute, the lowest scale value is 0. Vertical value scales are comparable across attributes.

Table BValues of Attribute Levels for LION Case



We can get a rough indicator of the relative importance of attributes by looking at the spread of the high-to-low values of the attribute. One has to be careful in interpreting this because the value obviously depends on the attribute levels we have chosen for the study. For example, if the Environmental Performance variable included an attribute level "Fails to Meet Requirements After First Year," that would drive up the value difference between lowest and highest attribute levels.

599-112

	(1) Highest Score for any level of this Attribute	% of Importance = Col. (1) entry divided by Total of Col. (1)
Brand	150	150/500 = 30%
Engine Power	140	140/500 = 28%
Price	100	100/500 = 20%
Fuel Consumption	60	60/500 = 12%
Environmental Performance	<u>50</u>	50/500 = 10%
Total	500	

From the values in **Table B** we can derive a rough measure of attributes' importance.

One surprise from this study was the low importance of Environmental Performance, estimated at only 10% of the purchase decision. Previous studies done by the company featured surveys in which people were directly asked the importance of environmental performance. As one might expect, the offered response was typically that it was critical. Here, however, to "get" environmental performance, a respondent had to give up something else, like power or price, and was generally not inclined to do so.

Because price was one of the variables, one can get a rough sense of the dollar value (or actually DM value here) of different performance levels. For example, the "brand value" of LION over "Japanese" was 85 points. The price panel shows a 100-point difference equating to 20,000 DM. Thus, the brand value of LION relative to "Japanese" is roughly:

$$\frac{85}{100} * 20,000 \,\mathrm{DM} = 17,000 \,\mathrm{DM}$$

Similarly, reducing fuel consumption by 25%, from 16 to 12 liters per km, was worth 60 points, or in DM terms:

$$\frac{60}{100}$$
 * 20,000 DM = 12,000 DM

(b) Market Simulations Once we have respondents' value systems, we can predict what automobile they would choose from a given set. For example, suppose a customer had a choice of three automobiles, as follows:

At	tribute	Model A	Model B	Model C
٠	Brand	LION	German	Japanese
•	Engine Power	150	200	250
٠	Fuel Consumption (liters/100km)	12	16	14
٠	Environmental Performance	new standards	meets	meets
			requirements	requirements
٠	Price (DM)	DM 60,000	DM 70,000	DM 50,000

Let's assume this person's value system matched the average market system in **Table B**. Then, we can compute the value he or she would place on each option:

Model A: 230 value points (LION = 85 + Power = 0 + Fuel = 60 + Environmental = 50 + Price = 35)

Model B: 260 value points (by same method)

Model C: 290 value points (by same method)

Two rules are commonly used to translate value points into predictions of share. First is the simple "the consumer buys whatever is the highest point total." This is called the "Maximum Utility Rule." Using that rule, we would predict this person would buy Model C. We can look at part-worths to say this is due to its Engine Power. An alternative rule is the "Share of Utility" rule in which the probability of buying a given model is proportional to its value points, i.e.,

Probability Buy Model A = 230/(230 + 260 + 290) = 29%

Similarly, the probabilities for B and C would be 33% and 37% respectively.

Note that once we have the value system estimates for a representative set of individuals, we can simulate any scenarios we like, e.g.,

- What happens at various price points?
- What happens if LION is offered in two models?
 - a. LION, 150 HP, 12 liters, New Standards, DM 60,000, and
 - b. LION, 200 HP, 14 liters, Meets Requirements, DM 70,000

The first scenario of price changes was fully investigated in this study and the optimal price found to be DM 54,000 vs. DM 60,000 as originally intended, when the thinking was consumers were willing to pay for environmental performance.

General Conjoint Decision Issues

There are a number of different approaches to conjoint, varying mostly in the task which is placed upon respondents. We saw one particular form in the LION study. Respondents were asked *pairwise preferences*. An offshoot of this is to ask the respondent "by how much" is one option preferred, e.g., "Press a number from 1 to 9 to indicate your preference where 1 represents you prefer option A a

great deal, 5 if you are indifferent between the two options, and 9 means you prefer option B a great deal."

Another common approach is to take "full profiles" (i.e., ratings of objects on all the attributes in the study) and simply ask for an absolute rating of desirability rather than pairwise comparisons. Essentially this produces data similar to the pairwise preference methods; the key question is, what response task can a respondent do more reliably?

Finally, there is a hybrid method which uses one of the two methods above in conjunction with the respondents' own estimates—or "self-explicated" ratings. Lilien and Rangaswamy's software from *New Product and Brand Management*,⁸ Chapter 4, follows this approach.

Guidelines for Use

Conjoint is a powerful tool with broad applicability. Necessary assumptions underlying conjoint have been mentioned throughout this discussion. We collect them here to summarize situations wherein conjoint would be most applicable.

1. Product as a Bundle of Attributes

The product must be able to be specified as a collection of attributes. There are some largely image products (e.g., a perfume) for which this is just not possible.

2. Must Know Important Attributes

Conjoint requires that we either know or find out by another method what attributes are salient in the product category.

3. Respondents Can Reasonably Rate Products

The input data we require from respondents are overall preference or purchaselikelihood judgments. This requires a level of respondent familiarity with the product category.

4. Attributes Should be Actionable

The firm should, in most cases, be able to act upon the output of the conjoint by constructing products that deliver the attribute levels used in the analysis.

This note has tried only to communicate the basic principles of conjoint analysis. Many researchers are currently at work expanding the domain of applicability and accuracy of conjoint. Specialist market research firms exist to deal with complicated applications while straightforward ones can be addressed internally. State-of-the-art software is available inexpensively. Complicated or straightforward, effective use of conjoint requires that the manager understand the technique, its vast potential, and its limitations.

⁸ G. Lilien and A. Rangaswamy, *New Product and Brand Management: Marketing Engineering Applications* (Reading, Mass.: Addison Wesley, 1999).