Attitude Models Revisited: An Individual Level Analysis

MASAO NAKANISHI
JAMES R. BETTMAN

Models of attitude structure and of attribute importance are examined using regressions for each individual subject for data on toothpaste brands. Results show that beliefs-only and standard evaluation times belief models are virtually indistinguishable, and that importance has different meanings for different subjects.

Much recent theoretical and empirical research on consumer decisions has involved models of cognitive and attitude structure and information processing. The purpose of this article is to report a set of related studies that examine several methodological issues in this area. In particular, several topics related to multiple-attribute attitude models of the Fishbein type (Fishbein, 1972) are examined.

The Fishbein attitude model, adapted to consumer choice, states that

$$A_{act} = \sum_{i=1}^{n} a_i b_{ij},$$

where

- $A_{act}$ = attitude toward the act of purchasing brand $j$,
- $a_i$ = evaluative aspect of consequence $i$ of that act,
- the goodness or badness of the consequence
- $b_{ij}$ = the strength of the individual's belief that the act of buying brand $j$ will lead to consequence $i$,
- $n$ = number of relevant consequences, which has been limited, although it need not be, to attributes of the product class.

Several controversies have arisen regarding this model or the similar Rosenberg (1956) model. It has been suggested that dropping the evaluative aspect ($a_i$) from the model will improve predictions (e.g., Sheth and Talarzyk, 1972). Others have disputed this (Cohen and Athola, 1971). Cohen, Athola, and Fishbein (1972) argue that proper measures of the model's components have not been used. In particular, they object to the use of importance measures as a substitute for the evaluative measure $a_i$. Wilkie and Weinrich (1972) have found that inclusion of all attributes in the model for each individual in the sample does not necessarily lead to the best predictions. Different subjects seem to use different subsets of the attributes or consequences, and taking this into account leads to better predictive accuracy for preferences.

In examining these issues, particularly whether the evaluative term should be dropped, a cross-sectional approach has often been taken to model estimation. That is, all subjects' data points are used to estimate a common model for each brand. Wilkie and Pessemier (1973) quite forcefully point out the inappropriateness of this approach, which ignores individual differences in subjects. Several alternatives have been utilized. Bass and Talarzyk (1972), Bass and Wilkie (1973), Wilkie and Weinrich (1972), and Bass, Pessemier, and Lehmann (1972) use subjects' $a_i$ and $b_{ij}$ ratings to compute values for each brand for $\Sigma_{i-1} a_i b_{ij}$. These values are then ranked to give a predicted brand preference ordering. A brand preference ranking obtained by direct questioning from each subject is used as a comparison point for these predicted values. A confusion matrix is computed, with entries $c_{ij}$, where $c_{ij}$ is the percentage of the subjects for whom the brand actually given preference rank $j$ is predicted to have preference rank $i$. The confusion matrix approach, however, is unwieldy for statistical tests of differences between alternative models.

Beckwith and Lehmann (1973) took a more natural approach, where predicted and stated attitudes were correlated over brands (television shows in their study) for each individual. The individual correlation coefficients were then averaged over the sample to find an overall measure of fit for a model. They then used mean correlation coefficients to compare the $\Sigma a_i b_{ij}$ and $\Sigma b_{ij}$ models. This approach avoids the pitfalls of cross-
sectional analyses, but Beckwith and Lehmann did not perform any statistical tests. In fact, if correlation coefficients between stated and predicted attitudes were computed for more than one model for each individual, they would not be independent observations and hence the usual tests for differences in means would not apply.

The approach taken in this study extends that of Beckwith and Lehmann. First, for each alternative model a correlation coefficient is computed for each individual between stated attitudes toward brands and those predicted by the model. Second, individual correlation coefficients are analyzed using the repeated measures analysis of variance model (Winer, 1962), which is ideally suited for analyzing data consisting of more than one observation per individual. Since this model is capable of handling more than one type of experimental treatment, in addition to studying the effect of dropping the $a_i$ term, as Beckwith and Lehmann did, this study examines how many attributes or consequences give the best degree of fit for each individual subject.

To accomplish this, a mechanism is needed for determining, for each subject, in which order attributes are examined or used by that subject. This perspective assumes a sequential information acquisition model, where attribute information is sought in some order by each subject, and the search is terminated after a satisfactory amount of information has been obtained. The attribute order and amount of information considered satisfactory may vary over individuals. If a subject’s preferred attribute order can be determined, then models using only the attribute accessed first, only the first two attributes, and so on can be estimated for that subject. Tigert (1966) found that a paired comparison measure of attribute importance correlated quite highly with the order in which subjects requested envelopes containing attribute information to help them make brand choices for toddler tops. Hence, such a measure of importance can be used as a surrogate criterion for the order of attribute inclusion. This criterion was used in this study. That is, models using only the attribute rated most important, only the two attributes rated first and second in importance, and so on were estimated.

In summary, the purpose of this study is to examine, by using individual level analysis,

(i) Can the $\sum_{i=1}^{n} b_{ij}$ and $\sum_{i=1}^{n} a_ib_{ij}$ models be differentiated at the individual level with respect to degree of fit?

(ii) How many attributes provide the best degree of fit for subjects’ attitude models?

In examining these two issues it became apparent that the order in which subjects utilize attributes is quite relevant. Hence models of individual’s attribute importance judgments will also be considered in a later section.

METHODOLOGY

The product category used to examine these research questions was toothpaste. This category was chosen because several previous studies attempting to distinguish the $\sum b_{ij}$ and $\sum a_ib_{ij}$ models have used this product class (Bass and Talarzyk, 1972; Bass and Willkie, 1973; Cohen and Ahtola, 1971; Sheth and Talarzyk, 1972). The five attributes used for the study were those found relevant in the earlier work: whitening teeth, preventing cavities and tooth decay, economical in use, freshening breath, and pleasant tasting. Seven brands were used: Ultra Brite, Pepsodent, Macleans, Crest, Close-up, Colgate, and Gleem. Subjects were 121 graduate and undergraduate students.

Special care was taken in obtaining measures for attitude toward the act of buying each brand ($A_{act,j}$); the evaluative aspect of attribute $i$ ($a_i$); beliefs ($b_{ij}$); and importance of attribute $i$ ($I_i$). In response to the criticism of Cohen, Fishbein and Ahtola, (1972) that appropriate measures of model components were not being used, the $AB$ scales developed by Fishbein and Raven (1967) were used to measure $A_{act,i}$, $a_i$, and $b_{ij}$. The $AB$ scales are semantic differential scales, with five adjective pairs utilized in the $A$ scale for measuring $a_i$ and $A_{act,i}$ (good-bad, clean-dirty, healthy-sick, wise-foolish, and beneficial-harmful) and five in the $B$ scale for measuring $b_{ij}$ (possible-impossible, true-false, existent-nonexistent, probably-improbably, and likely-unlikely). The concepts rated are, for example, “Buying Ultra Brite toothpaste for my own use is:” ($A_{act,j}$); “Having white teeth is:” ($a_i$); “The statement ‘Ultra Brite toothpaste whitens teeth’ is:” ($b_{ij}$). Five filler pairs are added for each scale, and the order randomized. Since each pair is scored on a $-3$ to $+3$ bipolar scale, the final measures for $A_{act,i}$, $a_i$, $b_{ij}$, obtained by summing the scores for the five relevant pairs, range from $-15$ to $+15$. Importance was measured by using paired comparisons, in keeping with Tigert’s findings discussed above (Tigert, 1966). Given these measures, ten regressions were run for each subject with seven observations (one for each brand) using $A_{act,i}$ as the dependent variable, with the $\sum b_{ij}$ and $\sum a_ib_{ij}$ models being run for each of five levels of attribute inclusion as determined by the importance scores.\[^1\]

RESULTS

Models of Attitudes

The ten regression equations run to examine research questions (i) and (ii) have the following general form.

\[^1\] In some cases, all the $A_{act,j}$ or all $b_{ij}$ measures were identical for a given number of attributes for a particular subject. In this case, the regression model could not be run. Hence, the number of subjects for each analysis is in general less than 121.
\[ A_{act} = \alpha_1 + \beta_1 \sum_{k=1}^{r} a_{ik}b_{ij} \quad r = 1, 2, 3, 4, 5 \] (2)

\[ A_{act} = \alpha_2 + \beta_2 \sum_{k=1}^{r} b_{ik} \quad r = 1, 2, 3, 4, 5 \] (3)

Here \( i \) denotes the index of the attribute ranked \( k^{th} \) in importance. This was ascertained by examining the rank order of each subject's importance measures. The mean \( R^2 \) across subjects for each of ten regressions is given in Table 1.\(^2\) Also given is the number of subjects for whom

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Source of Variation & Degrees of Freedom & Sum of Squares & Mean Squares & F-ratio \\
\hline
Models & 1 & .016 & .0160 & .358 \\
Models \times Error & 109 & 5.315 & .0488 & .947 \\
No. of Attributes & 4 & 6.350 & .1575 & .944 \\
Attributes \times Error & 436 & 72.542 & .1664 & \\
Interaction & 4 & .034 & .0085 & \\
Interaction \times Error & 436 & 3.916 & .0090 & \\
Between-Subject Error & 109 & 464.376 & 4.2603 & \\
\hline
\end{tabular}
\caption{Analysis of Variance Table for Attitude Models}
\end{table}

the impression that the differences in \( R^2 \)'s between models in Table 1 are so small that it is impossible from this data to tell which model better describes consumer attitude structures. This finding agrees with Beckwith and Lehmann (1973).

This result is not due to the small number of data points for individual regression analysis (i.e., seven brands). The relative magnitude of the error mean squares for between subject variation and within subject variation suggests that the \( R^2 \)'s for each subject for the different models covary highly. Since the degrees of freedom in the ANOVA table are independent of the number of brands, increasing the number of data points for the regressions does not necessarily increase the power of the tests. Also, giving subjects more choices may merely create an unrealistic decision situation where subjects are asked to deal cognitively with more stimuli than they are accustomed to handle.

At this point it is interesting to see what conclusions one might have drawn from the usual cross-sectional analysis. Since there are seven brands involved, regressions were run for each brand across subjects for five levels of attribute inclusion. The order in which attributes were added for each subject was again determined by the rank of the importance rating for that subject. Table 3 gives \( R^2 \) averaged over the seven brands.\(^3\)

The \( \Sigma a_{bi} \) model is slightly, but discernibly more descriptive than the \( \Sigma b_{ij} \) model in this comparison. What caused the better fit of the \( \Sigma a_{bi} \) model in the cross-sectional regression is unknown, but one might have concluded from this result that the \( \Sigma a_{bi} \) model is superior to the \( \Sigma b_{ij} \) model as a descriptive model of \( A_{act} \). On the basis of the individual analysis such a conclusion is clearly untenable.

\(^3\) The \( R^2 \) values could have been improved by the normalization process of Bass and Wilkie (1973), but since our purpose was not to improve on cross sectional results, no normalization was performed. The results are comparable in magnitude to the cross-sectional studies of Cohen and Ahtola (1971), and slightly better than those of Sheth and Talarzyk. Also, the similarity of the two models in terms of degree of fit agrees with the results of previous studies using toothpaste, except for Sheth and Talarzyk (1972).

\(^2\) Since our purpose is to compare the descriptive ability of different models, regression coefficients are not considered in the analysis.
TABLE 3

MEAN $R^2$ VALUES$^a$ FOR CROSS-SECTIONAL ANALYSIS

<table>
<thead>
<tr>
<th>Models</th>
<th>No. of Attributes Included</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\Sigma a_i b_{ij}$</td>
<td>.166</td>
</tr>
<tr>
<td>Mean $R^2$</td>
<td>.155</td>
</tr>
</tbody>
</table>

$^a$ Based on regressions for each of seven brands, each regression using 110 subjects as data points.

but did not improve the fit of either the $\Sigma a_i b_{ij}$ or $\Sigma b_{ij}$ model, since it is impossible to distinguish the two models for a single attribute. But the fact that the single most important attribute explained on the average 39.6 percent of the variance in stated $A_{ort}$ at the individual level is significant. It is clear that, at least for toothpaste, stated $A_{ort}$ is highly correlated with the $b_{ij}$ for the most important attribute. This finding suggests that it is highly useful to examine what determines importance ratings of attributes, since we have argued, following Tseig (1966), that these ratings are related to the order in which attributes are considered by the subjects. We now turn to models of attribute importance.

Models of Importance

Willkie and Weinreich (1972) in a previous study used the concept of determinism as proposed by Myers and Alpert (1968) as a criterion for attribute inclusion. Myers and Alpert argued that the determinism of an attribute was multiplicatively related to its desirability and the amount of perceived differentiation across alternatives. Hence, Willkie and Weinreich, in using this criterion for ordering of the entry of attributes, postulated that the order in which attributes are scanned is the same as the order of the value of

$$D_i = a_i \sigma_i$$

where: $D_i$ = determinism score for attribute $i$,

$\sigma_i$ = standard deviation across brands of the $b_{ij}$ score for attribute $i$.

If the order of attribute inclusion were in fact determined by the importance ratings as we have postulated, then the determinism score $D_i$ and importance ratings should be closely related. Two other possible determinants of importance are simply $a_i$ and $\sigma_i$ alone. The $a_i$ model implies that an attribute is important and accessed early in the search process if it has a high evaluative content, regardless of the degree of uniqueness over brands. A $\sigma_i$ model implies that only attribute uniqueness over brands is relevant for importance judgments, and hence order of attribute usage.$^4$

$^4$ Beckwith and Lehmann (1973) briefly examined this $\sigma_i$.

These three models of importance are examined using individual level analysis, where this time the regressions are carried out across attributes for each subject. The models are formally stated as

$$I_i = \alpha_i + \beta_i a_i \sigma_i$$

$$I_i = \alpha_i + \beta_i a_i$$

$$I_i = \alpha_i + \beta_i \sigma_i$$

where $I_i$ is the importance score for attribute $i$. The $I_i$'s are measured as follows. For each distinct pair of attributes $i$ and $j$ ($5 \times 4/2 = 10$ pairs), the subject was asked to choose which was more important, and then how much more important that attribute was, on a 0 (indifferent) to 9 (very much more important) scale. If this rating is denoted by $d_{ij}$, then the sign of $d_{ij}$ is positive if $i$ is more important than $j$, and negative if $j$ is more important than $i$. Since only half of the pairs are rated, $d_{ij}$ is set equal to $-d_{ji}$. Also, $d_{ii}$ is set equal to 0. Then the importance ratings $I_i$ are given by $I_i = \sum d_{ij}/5$.

The method of comparison was identical to that used for the attitude models. The three regressions were run for each subject, using importance observations for the five attributes.

Table 4 gives the mean $R^2$'s for each model across

TABLE 4

MEAN $R^2$ VALUES$^a$ FOR IMPORTANCE MODELS

<table>
<thead>
<tr>
<th>Models</th>
<th>$a_i$</th>
<th>$a_i \sigma_i$</th>
<th>$\sigma_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $R^2$</td>
<td>.420</td>
<td>.480</td>
<td>.268</td>
</tr>
<tr>
<td>No. sig$^b$</td>
<td>(24)</td>
<td>(27)</td>
<td>(9)</td>
</tr>
</tbody>
</table>

$^a$ Based on regressions from each of 120 subjects, each regression using five attributes as data points.

$^b$ The number of subjects for whom the $R^2$ value is significant at the $\alpha = .05$ level (i.e., $R^2 > .771$).

the sample of subjects and also the number of subjects for whom the $R^2$ value was statistically significant at the $\alpha = .05$ level. In order to test the differences between models, repeated measures analysis of variance was performed using z-transforms of the correlation coefficients. Table 5 presents the ANOVA results.

Since there is only one factor—the model effect—ininvolved, there is only one test to be performed. The $F$-ratio was significant at the $\alpha = .01$ level, indicating that the models differ in their degree of fit. By examining the average $R^2$'s in Table 4, we see that the $a_i$ model is the best, followed closely by the $a_i \sigma_i$ model, with the $\sigma_i$ model performing the worst. The importance of an attribute, and hence desire to utilize that attribute, therefore appears to be primarily related to the evaluative aspect of the attribute.

model, although not in the context of a formal set of models of importance.
Also, the magnitude of within-subject error mean square relative to between-subject error mean square suggests that there are systematic individual differences in the manner in which attribute importance is judged. This is to be expected, as importance is a vague stimulus. Hence for each individual, the best importance model in terms of $R^2$ was determined. These results are shown in Table 6, which gives the number of cases for which the best $R^2$ value is significant at the $\alpha = .05$ level. As expected from Table 5, the $a_i$ model is the best model for by far the largest number of subjects, but the $a_{ij}$ model is the best for about a third of the subjects. For this third of the subjects then, the concept of determinism is most strongly related to the concept of attribute importance. Beckwith and Lehmann (1973) examined only the $\sigma_i$ model, which is the model which performs most poorly in this study.

That $a_i$ was a major determinant of importance judgments is a crucial point. If importance judgments are primarily a function of $a_i$, alone or in combination with $\sigma_i$, and the single most important attribute explains 39.3 percent of variance, then $A_{a_i}$ is also a function of $a_i$ as well as that of $b_{ij}$. The fact that one cannot discriminate between the $\Sigma a_i b_{ij}$ and $\Sigma b_{ij}$ models at the individual level does not imply that $a_i$ is not necessary as a component of $A_{a_i}$. It is more likely that the nature of toothpaste is such that an individual essentially utilizes only one attribute in forming $A_{a_i}$. With more complex products for which several attributes might enter into the formation of $A_{a_i}$, it may be possible to discriminate between the two models.

Not surprisingly, attribute importance means different things to different people. Hence, the use of importance as the equivalent of an $a_i$ component in a multi-attribute model (Bass and Talarzyk, 1972; Sheth and Talarzyk, 1972) only leads to ambiguous interpretation. It would seem that use of a measure with a more common meaning, such as goodness-badness as used by Fishbein would be more appropriate. Even if a Rosen- 

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models</td>
<td>2</td>
<td>21.878</td>
<td>10.9390</td>
<td>40.88*</td>
</tr>
<tr>
<td>Models × Error</td>
<td>238</td>
<td>65.700</td>
<td>2676</td>
<td></td>
</tr>
<tr>
<td>Between Subject Error</td>
<td>119</td>
<td>117.223</td>
<td>.9851</td>
<td></td>
</tr>
</tbody>
</table>

* Significant at the $\alpha = .01$ level.

This study performed individual level analyses to examine some of the issues related to multiple attribute models as descriptive models of consumer attitude structure. As in Beckwith and Lehmann (1973), one finds that at the individual level the explanatory power of the $\Sigma a_i b_{ij}$ model and that of the $\Sigma b_{ij}$ model are virtually identical. Furthermore, no significant improvement in explanatory power resulted from the inclusion of more than the most important attribute. Finally, importance of attributes, and thus indirectly the attribute inclusion order used by subjects, was found to be primarily related to the $a_i$'s.

These findings, of course, cannot be indiscriminately generalized. What we have examined is the structure of consumer attitudes toward a single product class—
toothpaste—and as such we may have found out much more about the nature of attitudes toward toothpaste than about consumer choice processes or attitude structures. In examining this issue, we take a different point of view from Beckwith and Lehmann (1973), who emphasize data-based problems or model specification errors for explaining their results.

This viewpoint has been emphasized by Newell and Simon (1972) with respect to models of human information processing in general. They argue that since humans are adaptive, their behavior is appropriate to their goal, and thus in some sense is behavior demanded by the problem solving environment. If behavior conforms to that demanded by the situation, then a subject's behavior tells us more about the task than about himself. The only way information about psychological factors within the subject can be derived is by examining the subject's behavior in problem solving situations where the subject cannot, due to his limited information processing capacity, be perfectly rational. Newell and Simon summarize these points: "To the extent that the behavior is precisely what is called for by the situation, it will give us information about the task environment. . . . To the extent that the behavior departs from perfect rationality, we gain information about the psychology of the subject, about the nature of the internal mechanisms that are limiting his performance." (Newell and Simon, 1972, p. 55). The task environment of this study, toothpaste, is basically a low involvement product class. Hence the facts that additional attributes beyond the most important did not significantly improve predictions and that the two model types were not distinguishable may

<table>
<thead>
<tr>
<th>TABLE 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEST IMPORTANCE MODELS FOR SUBJECTS</td>
</tr>
<tr>
<td>$a_i, \sigma_i$</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>No. of subjects</td>
</tr>
<tr>
<td>Percentage</td>
</tr>
<tr>
<td>No. significant</td>
</tr>
</tbody>
</table>

* The number of subjects for whom the $R^2$ value is significant at the $\alpha = .05$ level (i.e., $R^2 > .771$).

b Ties between the $a_i, \sigma_i$, and $\sigma_i$ models.
merely reflect the demands of the situation. That is, for a low involvement product class, very simplistic decision processes may be the most rational behavior. Put in another way, the cognitive complexity implied by the linear compensatory multiple attribute models seems to be an overkill for this product class. One suspects that the complexity of the models is not congruent with the complexity of the decision process involved. As argued above, if we wish to study complex models, we must examine complex decision situations, for only in these situations will we learn about psychological properties. This has the important implication that to examine the cognitively complex multiple attribute attitude models, the concentration on low involvement product classes must be abandoned, and models more suited to that type of product developed. Product classes evoking more complex choice rules and more potential individual differences in attribute utilization must be studied. For a similar argument in a slightly different context, see Chaffee and McLeod (1973).

In the light of the findings of this study and the above discussion, there are three methodological requirements for future research in multiple-attribute models: (1) the individual level analysis of this study should be used whenever possible to avoid the problem of cross-sectional analysis, (2) products must be complex enough to match the cognitive complexity of multiple-attribute models, and (3) individual importance judgments on product attributes should be collected, so that the indirect role of might play in the formation of through importance judgments) can be directly studied.

REFERENCES


This point has been made by others (Bass and Wilkie, 1973; Wilkie and Pessemier, 1973), but is worth reiterating given the preponderance of aggregate analyses done in the field.


