INFORMATION INTEGRATION THEORY: AN ALTERNATIVE ATTITUDE MODEL FOR CONSUMER BEHAVIOR

Working Paper No. 187

Cynthia J. Frey and Thomas C. Kinnear

The University of Michigan

FOR DISCUSSION PURPOSES ONLY

None of this material is to be quoted or reproduced without the express permission of the Division of Research.
INFORMATION INTEGRATION THEORY: AN ALTERNATIVE
ATTITUDE MODEL FOR CONSUMER BEHAVIOR

ABSTRACT

This paper presents an alternative approach to modeling consumer attitudes. While information integration theory has found application in social psychology, it has only been recently that a methodology suitable for consumer behavior research has been developed. The following discussion illustrates how this attitude theory may be applied to consumer behavior problems and what conceptual and methodological advantages it has relative to attitude theories currently in use.

INTRODUCTION

As models of consumer behavior have developed, one of the primary variables of interest has been that of attitude. The conceptual proximity of attitude to behavioral intention and choice in these models makes the study of consumer attitude formation and change of great potential value to consumer researchers. A high level of interest in the variable has encouraged the development and application of complex attitude models. In contrast to traditional Thurstone and Likert scale measures of attitude, which provide only summary evaluations of the object, multiattribute attitude models attempt to tap the underlying beliefs which provide the basis for attitude formation and change.

Attitudes and Behavior

The predictive link between measured attitudes and observed behavior has always been tenuous despite the attention it has received from social scientists (Schuman and Johnson, 1976). Attitudes as measured by multiattribute models have shown somewhat the same relational pattern. The problem of poor attitude-behavior prediction appears to be the result of factors other than deficient attitude measurement such as intervening situational stimuli.
Nakanishi's contiguous retrieval model of human information processing addresses the issue of low attitude-behavior correlation (Nakanishi, 1974). Rokeach (1968) and other psychologists visualize attitudes as clusters of ideas and beliefs. Nakanishi theorizes that rather than internally determined sequential processing of information, the actual processing sequence is stimulus dominated. Perceived "stimulus configurations" in the environment are matched against the existing cognitive structure or patterns in long term memory. This serves to identify a particular salient element. Once such an element is identified, contiguously related data sets are accessed. There is no guarantee, however, that all data sets or instructions are executed during the decision process. Similarly, different salient elements and patterns may be triggered by different situational characteristics. Thus, the real value of multiattribute attitude models is in understanding the elements of cognitive structure and their interrelationships rather than predicting actual behavior, since the intervening situational factors are usually beyond the control of consumer researchers.

Multiattribute Attitude Models in Consumer Behavior

As a prerequisite for the advancement of multiattribute models in consumer behavior, researchers should feel confident that the underlying theory is reasonably valid. The Fishbein model has been the most highly researched of the multiattribute models applied to consumer behavior problems. While it rests on a foundation of psychological theory oriented toward attitude formation and change, the underlying theory has not been extensively explored in the context of consumer behavior. The model can be expressed by the following equation:

\[ A_0 = \sum_{i=1}^{n} b_i e_i \]
Where $A_0$ represents the attitude (affect for or against) toward any psychological object,

$b_i$ represents the belief (subjective probability) that object $o$ possesses some attribute $i$,

$e_i$ represents the evaluation of attribute $i$ (expressed on a goodness-badness dimension), and

$n$ represents the number of salient attributes.

Thus, according to Fishbein, one's attitude is really an expression of the sum of the salient beliefs about a specific object, weighted by the value one attaches to each of those beliefs.

In the field of consumer behavior, the construct validity of this concept of attitude has been researched by Bettman, Capon, and Lutz (1975a, 1975b, 1975c, 1975d). Their early research focusing on the multiplicative function within attributes (1975a, 1975b) offers considerable support for the validity of this aspect of the model. Their further research (1975c, 1975d) exploring the adding function across attributes, however, has been inconclusive. While one of their studies (1975c) indicated that adding appeared to be appropriate in most cases, the research design was incapable of distinguishing between adding and equal-weight averaging. An extension of their work (1975d) reveals that averaging of information across attributes appears to be more common than adding. The authors acknowledge that while this finding suggests an important modification of the basic Fishbein model, further investigation using other approaches is necessary for confirmation.

**Cognitive Algebra in Multiattribute Models**

The aforementioned research on adding versus averaging in multiattribute models serves to highlight a long-standing controversy in psychology between Fishbein and Norman Anderson. While each professes to have a theory approximating the cognitive algebra individuals use when combining informational stimuli to form attitudes, the development and operationalization of their
theories differ radically.

Cognitive algebra refers to the mathematical representation of psychological processes or the mathematical function individuals use when combining informational stimuli. Anderson has long maintained the viability of averaging as the most common integration mode for attitude formation (Anderson, 1971). Further, he suggests the possibility of equal-weight, constant-weight, and differential-weight averaging depending on the individual information processor, the psychological object, and the situation. When applying Anderson's information integration theory to experimental situations, one collects summary attitude measures for each object. Computational methods exist for identifying specific attribute parameters.

The Fishbein model is somewhat more restrictive in its formulation. Instead of fitting a model to the data (attitude responses), subjects are required to provide direct estimates of attribute parameters which are then substituted in the prespecified model formula. The summary attitude measure toward each object is then computed on a subject by subject basis. The relative ease of application and operationalization of the Fishbein model has contributed to its widespread acceptance. Meanwhile, speculation as to viable cognitive algebras and individual differences remains as an unsettled issue. In particular the adding versus averaging of weighted attribute evaluations is an important reason for the further study of multiattribute attitude models.

Relevance of the Fishbein $b_i$ Component to Consumer Behavior

Additional support for continued study of attitude models other than the basic Fishbein model concerns the nature of this model's components. The basic Fishbein $b$ variable represents the belief that object $o$ possesses some attribute $i$. This is expressed as a subjective probability. In the case of consumer information used as salient input to brand or product attitude for-
mation, a problem arises. The subjective probability does not appear to deal realistically with objective data or discrete points on a dimension.

An example of this dilemma is suggested by Lutz and Bettman (1977). In the case of discrete attributes such as the number of bedrooms in a house examined by a prospective home purchaser, the object's degree of possession (high-low) of the attribute or likelihood of possession (likely-unlikely) does not seem applicable. The house either has three bedrooms or does not have three bedrooms. The authors suggest that transformation of the bedroom dimension to one where the consumer evaluates amount of bedroom space (high-low) or the belief that the house is spacious (likely-unlikely) is contrived and unrealistic. "It seems more congruent with what consumers probably do to measure the evaluation of three bedrooms directly and then weight that evaluation" (Lutz and Bettman, 1977).

The continually increasing amount of objective data in the marketplace emanating from public policy decisions makes this deficiency in the Fishbein model even more critical. With Federal agencies publicly committed to making more information available to consumers, objective data such as nutritional information and measures of energy usage and product performance become potentially salient elements in the consumer environment. If consumer researchers attempt to replicate the actual stimulus field in an effort to understand cognitive structure and attitude formation and change, then the model which is operationalized should be capable of representing the environment. While the basic Fishbein model was designed for more abstract psychological concepts and deals with such brand attributes as style or status effectively, the model appears to be deficient in its application to attribute dimensions with discrete values. In contrast, the assumptions underlying Anderson's theory of information integration do not directly restrict the format of attribute presentation.
Additional Attributes and Attitude Change

An issue related to the presence of additional types and sources of consumer information concerns attitude change resulting from adding more attributes to the individual's salient cognitive structure. The implications of new information on existing attitudes have yet to be explored in consumer behavior from a modeling perspective. Several studies on this issue have been done in experimental psychology by Fishbein and Anderson. When new information considered slightly less extreme than the existing attitude is incorporated into one's cognitive structure, adding would predict the new attitude to be more extreme. Averaging, however, would indicate that the extremeness of the resulting attitude should be diminished.

Considering the nature of the consumer information environment where individuals have been exposed to many products and situations over time, the most common case will be where new information is incorporated into an existing cognitive structure. The implications of this condition for attitude change are apparent when one considers the resultant attitude differences due to the adding versus averaging model formulation. Unless one can identify the proper cognitive algebra, the effects of a consumer communications program may be misinterpreted.

The dynamic nature of the consumer information environment provides support for the further investigation of Anderson's model. Although only applied to very limited consumer behavior problems to date (Troutman and Shanteau, 1976), this has been at least partially the result of the complexity of its application. Recent methodological developments by Norman (1976a, 1976b) indicate that this drawback has been all but eliminated. The primary purposes of this paper are to describe Anderson's information integration theory approach to attitude modeling and to discuss its possible application in consumer behavior research.
**INFORMATION INTEGRATION THEORY**

Information integration theory has been developed and refined by Norman Anderson since the late 1950's. Rather than being primarily an attitude theory or judgment theory, information integration simply attempts to describe how several coacting stimuli are combined by an individual to produce a response. It is a general theory which evolves from the concept of individuals as active integrators of informational stimuli in their environment. Personal experience, direct observation, written records, and remarks by other persons are just a few examples of the diverse stimuli with which we interact. Areas where the theory has been applied include person perception, evaluation of job offers and job applicants, food taste-testing and meal preferences, and the evaluation of the size-weight illusion of physical objects. Because information integration theory is so broadly defined and widely applied in psychology, responses may be in the form of utilities, preference and difference judgments, or attitudes.

**The Adding Model**

The most easily identified cognitive algebra found in information integration theory is simple adding. The theoretical representation of the model can be written as follows:

\[
R = C + \sum_{i=0}^{n} w_i s_i + \epsilon
\]

Where \(R\) represents the overt response (attitude) measured on a numerical scale,

- \(w\) represents the weight or importance of the stimulus,
- \(s\) represents the evaluative level of the stimulus on some dimension,
- \(C\) represents a constant which allows for an arbitrary zero in the response scale and is usually not explicitly examined,
- \(n\) represents the number of salient stimuli or attributes, and
- \(\epsilon\) refers to the error term.
According to this model, the contribution of stimulus \( i \) to the individual's overall attitude is merely the weight \( w \) multiplied by its associated scale value \( s \). It should be noted that the \( ws \) component allows one to include a representation of the subject's initial opinion in the summary judgment. The logical extension of the adding model which has found far greater support in the social psychology literature is the averaging model. Adding models impose no constraints on the weight parameters, whereas averaging models require the weights to sum to one. Therefore, averaging always implies some degree of stimulus interaction (Bettman, Capon, and Lutz, 1975a).

**Averaging Models**

There are two basic types of averaging models which have been explored in information integration theory. The constant-weight averaging model and the differential-weight averaging model. A third averaging model using equal weights can be considered a special case of the constant-weight model.

The general format of the averaging model is as follows:

\[
R = C + \frac{\sum_{i=0}^{n} w_i s_i}{\sum_{i=0}^{n} w_i} + \epsilon
\]

Note that in this model the effective weight of \( s_i \) is \( \frac{w_i}{\sum w_i} \) in contrast to simply \( w_i \) in the adding model. The constant-weight averaging model assumes that there is equal weighting of the stimuli within factors. If \( A \) represents a stimulus with three levels on the evaluative dimension, then the weights applied to each of the levels should be equal such that \( W_{AI} = W_{A2} = W_{A3} = W_A \). It should be expected that other stimuli will be weighted differently, \( W_A \neq W_B \neq W_C \), with equal weighting within and between stimuli being considered a special case.

Differential-weight averaging implies that weights applied to levels of the same factor will vary in some systematic fashion. Lutz has hypothesized that individuals attach more importance to information perceived as
being negative or low in scale value (Lutz, 1973). It is not clear whether this relationship might be due to the relative scarcity of negative information in the marketplace as Lutz would suggest or whether it might be attributed to some other factor(s) such as perceived economic or social risk involved in the decision. The potential severity of differential weight problems inherent in one's research design deserves careful attention. By relaxing the constant-weight constraint, the degree of complexity associated with parameter estimation, interpretation, and model testing increases considerably.

Parameter Estimation in Information Integration Theory

The cognitive algebra approach to modeling psychological processes involves two basic operations: parameter estimation and the identification of the mathematical function used by the subject to integrate the stimuli. Research efforts to date have been concentrated on the delineation of these combinatorial processes. The application of information integration theory to consumer judgment has been limited to the work of Troutman and Shanteau (1976) and Bettman, Capon, and Lutz (1975d). In both cases the issue of interest is the integration process, adding versus averaging, and no attempt is made at parameter estimation.

The research by Anderson and his colleagues in psychology also reflects a preoccupation with the combinatorial processes of information integration. Efforts appear to have been directed toward developing a typology of processes across tasks and situations. The relatively large number of different functions which have been identified as a result of this research seems to have actually inhibited the development of a general methodology for parameter estimation. Adding, multiplying, and the variety of averaging models discussed in the literature exhibit different parameter assumptions and constraints and require any estimation procedure to be highly flexible.
Functional Measurement Theory

In order to develop estimates of the weight and scale values for each stimulus, one must have an understanding of Anderson's theory of functional measurement. Functional measurement theory consists of solving three measurement problems simultaneously.

Given a k-dimensional stimulus or k pieces of information the subject produces a unidimensional response. This constitutes a mapping, \( f(s_1, s_2, \ldots, s_k) \rightarrow \mathbb{R} \). Underlying this response are the three processes postulated by information integration theory:

1. a **valuation function** maps each objective stimulus or piece of information to a point on a subjective scale, \( V(S) \rightarrow s \),

2. an **integration function** weights and combines the subjective values to give an integrated impression on a subjective response dimension, \( I(s_1, s_2, \ldots, s_k) \rightarrow r \), and

3. a **response function** maps the internal response value to a point on a physical scale, \( M(r) \rightarrow R \).

The data collected in an information integration experiment consist of \( R \), the observed interval scale representation of the internal response. In the case of attitude theory this \( R \) would be the subject's communication of his/her internal attitude \( r \). From this data the integration function (adding, averaging etc.) and weights can be derived which represent the internal response, \( I(s_1, s_2, \ldots, s_k) \rightarrow r \). The lower case \( s \) is the subject's interpretation of the presented stimulus value \( S \). The apparent purpose for the lower case and upper case symbols is to make clear that subjects will perceive and evaluate the same stimuli differently. Also, the same overt response \( R \) may represent different weights or internal processes. To summarize this procedure, from a series of observed \( R \) measures for different levels of stimulus combinations, an integration and weighting process can be identified. Unless the subject's process is unpatternable, once the appropriate model is isolated through tests of fit, the individual weight
and scale values can be estimated. The benefit of this approach according to Anderson (1971) and Norman (1978) is the concurrent validation of the valuation, integration, and response functions.

**Maximum Likelihood Parameter Estimation**

While functional measurement promises to validate the response scale and the integration model provides estimates of subjective scale values, the success of this procedure and the advancement of information integration theory depends on parameter estimation. It has been shown that algebraic manipulation and analysis of variance techniques can be useful, however, their capability to handle complex designs is severely limited. For this reason maximum likelihood solutions have been sought. Given that one is willing to make assumptions about the underlying probability distribution of error, maximum likelihood estimates have several desirable attributes:

1. \( \hat{\theta} \) is a sufficient estimator of \( \theta \) if a sufficient estimator exists, that is \( \hat{\theta} \) contains all the "information" about \( \theta \) that is in the sample,

2. \( \hat{\theta} \) is a consistent estimator of \( \theta \), that is asymptotically unbiased, with variance tending to zero, \( \lim_{n \to \infty} P(\|\hat{\theta} - \theta\| < \varepsilon) = 0 \) for any \( \varepsilon > 0 \), and

3. \( \hat{\theta} \) is efficient, with smaller variance than other estimators for large \( n \).

The maximum likelihood estimator is the hypothetical population value which maximizes the likelihood of the observed sample. If a random variable \( X \) has a probability distribution \( f(x) \) characterized by parameters \( \theta_1, \theta_2, \ldots, \theta_k \) and if we observe a sample \( x_1, x_2, \ldots, x_n \) then the maximum likelihood estimators (MLE) of \( \theta_1, \theta_2, \ldots, \theta_k \) are those values of these parameters that would generate the observed sample most often (Kmenta, 1971).

Maximum likelihood estimates are found by first defining a likelihood function on the data given the model and the parameter values, such that:
\[ L(\theta) = f(x; \theta) \]

Where \( \theta \) refers to a vector of parameter values \( (\theta_1, \theta_2, \ldots, \theta_u) \),
x refers to an array of sample observations, and
\( f \) refers to the joint probability density function \( L(\theta) = \prod_{i=1}^{n} f(x; \theta) \).
A support function for \( \theta \) to aid in computation of the estimate can be defined as:

\[ S(\theta) = \log L(\theta) \]

When applying MLE to information integration problems, the form of the likelihood function is determined by the independence of the observations across stimulus sets in the design. If observations \( R_k \) are independent, then the likelihood function can be written as the product of the probability density function for each stimulus set \( i \) across \( n \) replications.

\[ L(\theta) = \prod_{i=1}^{I} \prod_{n=1}^{N} f_i(x_{i,n}; \theta) \]

Where \( x_{i,n} \) is an observation from a population with a probability density function \( f_i(x_{i,n}; \theta) \). According to Norman (1978), this case is applicable for single subject analyses when responses are assumed to be uncorrelated across replications. It is also considered appropriate for between subject analyses when the model and parameters are assumed to be invariant across subjects.

Given that the probability density function is normal for each stimulus set \( i \), with expected value \( \mu_i \) and variance \( \sigma_i^2 \), the joint likelihood function for a set of observations can be written:

\[ L(\theta) = \prod_{i=1}^{I} \prod_{n=1}^{N} \left( 2\pi \sigma_i^2 \right)^{-\frac{1}{2}} e^{-\frac{1}{2} \left( x_{i,n} - \mu_i / \sigma_i \right)^2} \]

The support function is then:

\[ S(\theta) = -\frac{IN}{2} \log (2\pi) - N \sum_i \log (\sigma_i^2) - \frac{1}{2} \sum_i \left( \sum_n x_{i,n}^2 - \frac{1}{N} \sum_n x_{i,n} \mu_i + N\mu_i^2 \right) / 2\sigma_i^2 \]
Solutions for $\theta$ generally involve the nonlinear optimization of $S$. Two of
the best known methods for solving such systems are the Newton-Raphson method

When observations across stimulus sets in the design are correlated,
they must be considered as samples from a multivariate normal distribution.
The likelihood function can then be written as the product of the multivariate
density functions across $n$ independent samples:

$$L(\theta) = \prod_{n} f(X, \theta)$$

Where $X$ is a vector of observed values $(X_1, X_2, \ldots, X_n)$ from a multi-
variate population with the probability density function $f(X, \theta)$. According
to Norman (1978), this case is appropriate when parameters are estimated
for a group of subjects.

Given the multivariate normal distribution with centroid $\mu$ and variance-
covariance matrix $\Sigma$, the joint distribution for a set of $n$ subjects can be
written:

$$L(\theta) = \prod_{n} (2\pi)^{-\frac{3}{2}} |\Sigma|^{-\frac{1}{2}} e^{-X^2/2}$$

Where $X^2 = (X - \mu)^T \Sigma^{-1} (X - \mu)$

The support function is then:

$$S(\theta) = - \frac{1}{2} \log(2\pi) + \frac{N}{2} \log |\Sigma|^{-1} - \frac{1}{2} \sum_{n} (X - \mu)^T \Sigma^{-1} (X - \mu)$$

In the case of uncorrelated observations the maximum likelihood method
allows for unequal variance among cells in the design. This results in the
squared deviations between the observed data and the theoretical or estimated
values being inversely weighted by the variance. Greater error between
theoretical and observed measures is then concentrated in cells with greater
variance. The pattern of variance across cells should be examined to determine
how large an impact this may have on the resultant parameter estimates. Under
certain conditions, the model may be simplified by allowing equal variance among cells. In this case the MLE are the same as least squares estimates.

Independent Parameter Estimates and Tests of Fit

Tests of fit in information integration have been a cause for concern amongst some researchers because Anderson's parameter estimation procedures derive the independent variable values from the response score. This generally results in discrepancies between observed and predicted values being smaller than when parameters are estimated independently. Due to the following complications, Anderson maintains that independent estimation methods are inferior to his procedure.

From a model analysis perspective, Anderson notes that separate estimation confounds the validity of the model with the validity of the parameter estimates. This problem would arise when the parameter estimates are biased estimators of the true values. Since there is seldom any guarantee that independent parameter estimates are unbiased, discrepancies between predicted and observed response values become uninterpretable. In addition, the combined response variability and parameter estimation variability will decrease the power to detect deviations from the proposed model.

Although Anderson takes a rather negative view of respondent estimation of weight and scale parameters, he does cite evidence where respondent-supplied evaluative measures may be useful (Anderson, 1976). In the case of a single experimental condition, self-estimated parameters may be appropriate depending on the nature of the informational stimuli and the potential interactions. More critical than the estimation of evaluative measures, however, is the estimation of importance or weight parameters which Anderson warns may be very unreliable (Anderson, 1976).
The validity of these measures rests completely on the associated model, but the cited work rarely if ever provides a test of goodness of fit. (Anderson, 1976)

This is in reference to the customary test of the expectancy value model which is a correlational test between the predicted attitude and some other direct attitude measure such as the semantic differential. Refering directly to Fishbein's work, Anderson (1974) comments:

Theories that build on correlation-type statistics may well be building on sand. An adequate test of the model requires a test of the discrepancies from prediction such as is given by the analysis of variance (or likelihood ratio tests).

Regression and Correlation Analysis as a Model Test

There is ample evidence in the literature that frequently in model analysis, high correlations between predicted and observed values are taken as support for a will-fitting model. Contrary to this belief, several experiments have been done which have highly significant interactions among variables yet yield correlations between predicted and observed responses of .90 and higher. One such experiment used as input the width and length of a rectangle for the independent variables and the area of the rectangle as the observed measure of the dependent variable. Using a regression equation, the resultant $r^2$ of approximately .95 clearly shows that while the linear summation model has great predictive ability it is inadequate as a test of goodness of fit and does not reflect the actual underlying model.

Regression analysis has also been used with expectancy value models and attitude studies where the interval scale assumptions have been violated. Again, one is unable to partition the variance into that due to scaling and the portion which actually reflects deviations from the model. One of the advantages of information integration theory is that functional measurement theory provides a means of transforming ordinally scaled data to interval
scales. This is a necessary requirement of the goodness of fit test. The use of regression analysis then can be misleading in testing the appropriateness of theoretical models due to its insensitivity to deviations from the regression model format.

When using maximum likelihood parameter estimation techniques, goodness of fit can be evaluated with a maximum likelihood ratio test. By comparing the hypothesized model with the most general model possible, a ratio is derived which has a value that is distributed approximately as a chi-square random variable for large sample sizes. In addition to using likelihood ratio tests to examine goodness of fit, they may also be used to evaluate such hypotheses as:

1. Whether the factor weights are different,
2. Whether the weight of the initial impression is zero,
3. Whether parameters vary across blocks of trials or different experimental treatments, and
4. Whether the subjective values along different factors are on the same interval scale.

Upon careful inspection, maximum likelihood estimation techniques appear to offer the greatest capability of handling complex research designs, the most flexibility in model application, and the greatest chance of successful parameter estimation for both weight and scale values. With these factors in mind, an experimental design can be developed which incorporates information integration theory with consumer behavior in an information environment.

CONCLUSION

Information integration theory seems to offer much potential for attitude researchers in consumer behavior. This is especially true now that maximum likelihood parameter estimation procedures have been developed.
The purpose of this paper has been to describe this theory and suggest its possible use as an alternative to the popular Fishbein model.
REFERENCES


----- (1978), Maximum Likelihood Estimation for Parameters in Information Integration, College Park, MD: Center for Learning and Cognition, The University of Maryland.

Rokeach, Milton (1968), Beliefs, Attitudes, and Values, San Francisco: Jossey-Bass.