72-9021

BODI, Michael James, 1937-
AN INVESTIGATION OF SHETH'S FACTOR ANALYTICAL MODEL OF BRAND LOYALTY.

The University of Oklahoma, D.B.A., 1971
Business Administration

University Microfilms, A XEROX Company, Ann Arbor, Michigan

72-9021

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN INVESTIGATION OF SHETH'S FACTOR ANALYTICAL MODEL OF BRAND LOYALTY

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF BUSINESS ADMINISTRATION

BY

MICHAEL JAMES BODI

Norman, Oklahoma

1971

AN INVESTIGATION OF SHETH'S FACTOR ANALYTICAL MODEL OF BRAND LOYALTY

APPROVED BY

Dennis M. Crites

Ed F. Crane Jr.

WmR Smith

James McClenderline

William Diemander

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

I express my appreciation to those persons who have helped both in the completion of this dissertation and in my academic work in general.

My greatest appreciation goes to my Committee Chairman, Professor Dennis M. Crites, whose wisdom and leadership demonstrated throughout my graduate training have left their indelible mark. In particular his advice and personal interest in this research have been invaluable.

I thank Professor W. Alan Nicewander for his continuous confidence and prompting on the many technical issues without which this research could not have been accomplished.

My sincere thanks also go to the members of my advising and examining committee: Professors James M. Kenderdine, William R. Southard, Ed F. Crim, and William H. Keown.

I am bound to record my gratitude to my family for sharing with me in this adventure: To my wife, Diane, whose faith and strength have in a very real manner made this all possible. For her cheerful encouragement, understanding, and assistance throughout my graduate studies I express my deepest gratitude and to my children, Marcus and Michele, who have been a constant source of joy.

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

LIST OF APPENDIXES

AN INVESTIGATION OF SHETH'S FACTOR

ANALYTICAL MODEL OF

BRAND LOYALTY

CHAPTER I

INTRODUCTION

The purpose of this study is to further investigate the factor analytical model of brand loyalty. The technique first was introduced by Sheth in 1966 to test the basic tenets of what is now known as the Howard-Sheth theory of buyer behavior.[1] A primary theoretical goal was to trace brand loyalty from a near zero starting point to some new level. The related hypothesis was that acquisition of brand loyalty across purchase trials is a process of learning by repeated purchases of a brand. The point to recognize here and the recurring theme is that learning is a gradual or incremental process over time. In 1968, Sheth reported much of the same data.[2] Here, the emphasis is placed upon the empirical aspects of the factor analytic model of brand loyalty. Howard and Sheth

---

[1]Jagdish N. Sheth, "A Behavioral and Quantitative Investigation of Brand Loyalty" (unpublished doctoral dissertation, University of Pittsburgh), 1966.

[2]Jagdish N. Sheth, "A Factor Analytical Model of Brand Loyalty," Journal of Marketing Research, (November, 1968), pp. 395-404.

in 1969 again reported the same data analyses ". . .on the basis of expediency."[3]

Specifically, in Sheth[4] and Howard and Sheth[5] attempts were made, through analysis of the factor analyzed purchase data, to show that not only is brand loyalty learned but that the process of learning is consistent with Hull's[6] theoretical formulations. Also, brand loyalty scores (BLS's) for each individual buyer are generated from the model. The magnitude of the BLS indicates the degree of brand loyalty. A BLS accounts for both the frequency and pattern of buying a preferred brand. These BLS's are promoted as being an important and valuable aid for market segmentation decisions. The general utility of the model is anticipated by Sheth's statement that "Utilization of the model may very well answer questions as to whether in a specific product category, brand loyalty is a Bernoulli, first-order Markov, linear or non linear learning process."[7]

The primary goals of this study are to observe the factor analytic model both as a theoretically based model, and as an empirical model. The investigation is accomplished by applying the model to simulated consumer purchase data generated from two theoretically diverse statistical learning models. The type of experimental design selected seems

---

[3]John A. Howard and Jagdish N. Sheth, The Theory of Buyer Behavior, (New York: John Wiley and Sons, 1969), p. 398.

[4]Sheth, loc. cit., 1966; Sheth, loc. cit., 1968.

[5]Howard and Sheth, op. cit.

[6]Clark L. Hull, Principles of Behavior: An Introduction to Behavior Theory, (New York: Appleton-Century-Crofts, 1943); Clark L. Hull, A Behavior System, (New Haven: Yale University Press, 1952).

[7]Sheth, op. cit., 1966, p. 232.

appropriate in light of the fact that the central elements of the Howard-Sheth theory of buyer behavior are predicated on the basis of Hullian learning principles.

Relevancy of Study

The relevancy of the present study could be justified on the basis of four separate but related reasons. First, the quote by Business Week that John A. Howard ". . .believes that his theory will be to marketing what John Maynard Keynes' General Theory of Employment, Interest, and Money was to the development of economics.", should not be dismissed.[8] With such mass exposure to businessmen and the general public, and by the explicit suggestion of the theory's potential impact, the need for verification (or refutation) and any subsequent refinements or modifications is of major importance.

Secondly, within the marketing discipline, there is a general plea for the ongoing development of marketing as a science. That is, there is a need for more theoretical and empirical rigor. Engel, Kollat, and Blackwell observe that the majority of published articles reporting "principles" in consumer behavior are the result of one researcher's findings in a single investigation.[9] The lax research atmosphere just described tends to sustain invalid conclusions. One answer is that the marketing researchers establish the required and fruitful tradition of replicating findings now carried on in the behavioral sciences.

---

[8]"A New Theory on Why People Buy," Business Week, (January, 1970), p. 70.

[9]James F. Engel, David T. Kollat and Roger D. Blackwell, Consumer Behavior, (New York: Holt, Rinehart and Winston, Inc., 1968), p. 627.

A third rationale for this study is the recent and current so-called "multivariate revolution" in the field of marketing.[10] There is no question that multivariate measurement techniques are needed, but the apparent quixotic notion that these techniques as a class will provide quick, easy, and accurate solutions to complex problems needs to be dispelled. Identifying appropriate applications of multivariate measures is not obvious, as the outcome of this study may show.

The last characteristic which lends relevance to the study is the increasing acceptance and recognition that the social and behavioral sciences can contribute to the understanding of consumer behavior. The current emphasis on consumer behavior is witnessed by the fact that nearly 50 of the papers and discussions scheduled for the 1971 Fall Conference of the American Marketing Association are directly related to the study of consumer behavior.[11] Green and Tull, whose perspectives are mostly statistical and methodological, foresee more and more research devoted to the investigation of ". . .the social psychology of product and service consumption."[12] Moreover, these sentiments foretell that future marketing men will be more dependent on the social and psychological implications of their overall product marketing strategy.

The existing multivariate techniques, and computing systems available will accentuate the emphasis on the analysis of psychological

---

[10]Jagdish N. Sheth, "The Multivariate Revolution in Marketing Research," Journal of Marketing, (January, 1971), pp. 13-19.

[11]"Fall Conference, American Marketing Association," The Marketing News, First-of-July, 1971, pp. 4-10.

[12]Paul E. Green and Donald S. Tull, Research for Marketing Decisions, (2d ed.; Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1970), p. 462.

variables affecting brand and supplier choice.

Development and Verification of the Howard-Sheth Theory

In the present study the psychological variable is learning, the underlying explanation for the existence of brand loyalty. Why this is so will now be pursued in more detail. Additionally, it is hoped that simultaneously the rationale for the experimental design developed for this study will be revealed.

The general theme of the Howard-Sheth theory of buyer behavior is an explanation of the development of brand loyalty over time. The basic postulate of their theory, as summarized in 1966, is:

> ...the buying behavior of an individual over a period of time is a recursive decision-making process wherein the buyer develops habit mediators (decision rules) which are anchored to the anticipatory motives [goals] of the buyer and the perceived potential of the alternative courses of purchase behavior (with respect to the choice among brands) available to him to satisfy his motives."[13]

At this stage of development the Howard-Sheth theory had been already modified to include more than Hull's learning theory. Osgood's theory of cognitive behavior is provided as a decision-making framework within the Howard-Sheth theory.[14]

An earlier formulation by Howard was based explicitly on Hull's learning model.[15] By contrast, the theory in its present form is quite comprehensive in that it embodies the integration of Hull's learning

---

[13]Sheth, op. cit., 1966, p. 4.

[14]Ibid., p. 47.

[15]John A. Howard, Marketing: Executive and Buyer Behavior, (New York: Columbia University Press, 1963), p. 101.

theory, Osgood's cognitive theory, and Berlyne's theory of exploratory behavior.[16] However, the learning constructs still are ". . .the major components of the theory."[17] Although fascinating, an exposition of the Howard-Sheth theory is beyond the limits of this study. For a complete explanation of the theory refer to Howard and Sheth.[18] A brief overview of Hullian learning theory is presented in Chapter Two of this dissertation.

Others, in an actual field study, have researched the theory of buyer behavior in its entirety with disappointing results.[19] In a series of doctoral dissertations, subsets of the general theory have been examined.[20]

Sheth, in his doctoral dissertation, attempted to empirically observe the learning of brand loyalty. Factor analytic measures of attitude and behavior, the manifest variables, were made. He had partial success in testing:

---

[16]Howard and Sheth, op. cit., p. 32.

[17]Ibid.

[18]Howard and Sheth, op. cit.

[19]John U. Farley and Winston L. Ring, "An Empirical Test of the Howard-Sheth Model of Buyer Behavior," Journal of Marketing Research, (November, 1970), pp. 427-438.

[20]Sheth, op. cit., 1966; George S. Day, "Buyer Attitudes and Brand Choice Behavior" (unpublished doctoral dissertation, Columbia University, 1967); Schlomo, Lamport, "Word-of-Mouth Activity During the Introduction of a New Product" (unpublished doctoral dissertation, Columbia University, 1969); Terrence U. O'Brien, "Information Sensitivity and the Sequence of Psychological States in Brand Choice Process" (unpublished doctoral dissertation, Columbia University, 1969); Michael Perry, "Computer Simulation of Consumer Brand Choice" (unpublished doctoral dissertation, Columbia University, 1968).

> ...whether learning by experience as the theory of buyer behavior suggests is relevant in the development of brand loyalty. The analysis of the purchase data utilizing the reference curves [factors] suggests that this is so.[21]

The present investigation deals only with learning behavior as reflected in purchase protocols. A purchase protocol is defined as a finite series of ones, for a purchase of a favorite brand, and zeros, for the purchase of any other brand. The measurement of attitude as reaction tendency is not considered.

Learning Implications

Howard and Sheth take the position that if satisfaction is achieved from repeat purchases of a brand, then brand loyalty is likely to occur within the buyer. Further, if the level of satisfaction derived is equal to or greater than that expected, the perceived brand potential of that brand relative to competing brands is increased. Thus, an important incremental effect occurs, that is, the purchase probability for the particular brand will be greater the next time the product is bought. The buyer's ongoing development (over purchase opportunities) of his habit mediators or decision rules is defined as a learning process. In the theory of buyer behavior, learning from experience (measured by number of purchases) to develop the habit mediators is a necessary and dominant aspect of developing brand loyalty over time (n purchase trials).

In their recent conceptualization of buyer behavior Howard and Sheth do not identify the shape of what could be called brand

---

[21] Sheth, op. cit., 1966, p.214.

loyalty (learning curve) over time. In 1963, however, Howard[22] did illustrate an individual's learning curve for brand A (versus all others). With the horizontal axis showing the number of purchase trials, and the vertical axis showing the brand A purchase probability, the traditional learning curve is depicted. That is, the learning curve is negatively accelerated: meaning that the buyer's incremental amount of learning across each successive trial becomes less and less. Additionally, Howard assumes, in reference to his learning curve illustration, "...that 'not buying' is not an alternative; the remaining probibilities apply to other brands..."[23] In this study, the same assumption is adopted when individual purchase protocols are generated. As far as the theory is concerned, Sheth makes this statement:

> However, in an experimental design, if the other influences are held constant or eliminated, and the effect of purchase behavior is the only influence over time, it is suggested that the behavior curve may be the exponential learning curve for the individual buyer.[24]

Ideally, for measurement purposes, a group of adult buyers is needed who, in actual purchase situations, are completely unfamiliar with existing brands of various product categories. The sample group then is confronted with brand choice decisions from several brands with no prior experience or generalizations from past experiences with similar brands.

---

[22]John A. Howard, Marketing Management: Analysis and Planning, (Homewood, Illinois: Richard D. Irwin, Inc., 1963), p. 36.

[23]Ibid.

[24]Sheth, op. cit., 1966, p. 43.

A requisite condition for operationally measuring a behavioral variable such as learning brand preference from purchase experience is near zero or negligible previous exposure. Already stabilized brand preferences seriously inhibit the possibility of measuring new learning. The fact that the American consumer in general has high product and brand awareness be it from experience, word of mouth, imitative or from whatever source is well documented in the literature.[25]

Simply on an individual rational basis, certain existing brand loyalties (e.g. preference for a particular brand of cigarettes) might be traced back to a personal referral as opposed to the loyalty developing from repeat purchase situations. Observe the remarkable ability of children to recite from memory T.V. commercials. The consequence of all this is that much of the learning has already occurred.

Accordingly, the available, standard panel data, although valuable for many types of studies, are not desirable for attempting to explain the nature of learning starting from an unlearned state at some specified probability of purchase up to some asymptotic level. The level or extent of overall learning is measurable, but how learning came about is not determinable. Starting with the initial purchase situation up to some arbitrary asymptote is precisely the portion of the learning curve, both in the aggregate but primarily for individuals,

---

[25]Herbert E. Krugman, "Processes Underlying Exposure to Advertising," American Psychologist, Vol. 23, (1968) pp. 245-253; Herbert E. Krugman, "The Impact of T.V. Advertising: Learning without Involvement," Public Opinion Quarterly Vol.29 (Fall, 1965) pp. 349-356; William D. Wells, "Children as Consumers," in Joseph Newman (ed.) On Knowing the Consumer (New York: John Wiley & Sons, 1966) pp. 138-145; Eugene Gilbert; Advertising and Marketing Young People, (Pleasantville, New York: Printer's Ink Books, 1957). Lester Guest, "Genesis of Brand Awareness," Journal of Applied Psychology, Vol. 26, (1942) pp. 800-808.

that needs measurement if the important learning aspects of the Howard-Sheth theory are to be empirically verified.

Sheth's Experimental Design

Extraneous influences were reduced in the Sheth study by the use of a small number of "just arrived," foreign national, graduate students who, for pay, volunteered to maintain purchase records of price, location, brands, products, etc.[26] Sheth, in a separate article, provides an exhaustive discussion of the panel make-up and the control for generalization and imitative behavior.[27]

The advantages Sheth may have accrued by eliminating or holding constant a number of external influences in his desire to test the theory in a naturalistic environment are counterbalanced by the disadvantage of small sample sizes. Factor analysis was applied to two samples who purchased different brands of rice. The first sample consisted of 17 subjects and 6 purchase trials. The second rice sample, a modification of the first, resulted from dropping three members and adding one more trial for a total sample of 14 members measured over 7 purchase trials. Also, the development of brand loyalty for bread was observed on samples of 26 subjects and 8 trials and 14 subjects and 12 trials.

In order to maintain a sufficient degree of reliability in the calculation of principal components analysis the ratio between obser-

---

[26]Sheth, op. cit., 1966, p. 113.

[27]Jagdish N. Sheth, "Learning of Brand Preference at the Adult Age," Journal of Advertising Research, (September, 1968), pp. 25-36.

vations (buyers) and variables (purchase trials) is estimated to be around 10 to 1.[28] Sheth is not close to meeting this criterion.

Necessarily, before elaborating more specifically on how the investigation of the factor analytic model is to be conducted, attention is turned to a counter-theory of learning. Then, following a presentation of implications of Sheth's model for applied market research, the research plan for this study is described.

A Counter Theory of Learning

In psychology, a counter theory of learning is usually referred to as one trial learning or all-or-none learning. One of the earliest advocates of this learning concept was Lashley.[29] Putting this theory in a marketing context; when a consumer is in a buying situation, she tries one cue or hypothesis after another. When one hypothesis does not work, she discards it and adopts another. In this way, the consumer continues trying one strategy after another until she hits one that satisfies her. For an uncertain number of purchase trials she goes from one wrong hypothesis to another. Suddenly, without any previous gradual improvement, she hits upon the correct hypothesis, and at once has "learned" her favorite brand. Therefore, her behavior reflects "all-or-none learning." Probably, the most noted proponent of all-or-none learning is Guthrie.[30] Chapter two provides a brief description of his learning theory.

---

[28]Green and Tull, op. cit., p. 462.

[29]K. S. Lashley, Brain Mechanisms and Behavior, (Chicago: University of Chicago Press, 1928).

[30]Edwin R. Guthrie, "Psychological Facts and Psychological Theory," Psychological Bulletin, Vol. 43 (1946). pp. 1-20.

It is curious that in the last decade the great majority of market researchers have implicitly utilized the basic postulates of Guthrie's learning theory (in the form of brand preference or loyalty) under the guise of Markov chain analysis.[31] First-order Markov chain analysis does fit the all-or-none precepts well in that the probability of switching states (say in a two-state system of unlearned or learned) on purchase trial n is dependent only upon the probability of switching states on trial n - 1. Such a learning situation is independent of the history of purchases and, so, is time-independent. The above discussion of incremental learning dictates time-dependency because a given probability of purchase at time t represents the cumulative effect of all previous purchases. Another assumption of Markov chains is that the transitional or switching probabilities are stationary for the specified period of study.

To return to the issue of investigating a truly psychological construct such as learning which is manifested in buyer behavior, it

---

[31] Jean E. Draper and Lassy H. Nolin, "A Markov Chain Analysis of Brand References," Journal of Advertising Research (September, 1964), Vol. 4, pp. 33-39; Solomon Dukta and Lester Frankel, Markov Chain Analysis: A New Tool of Marketers, (New York: Audits and Surveys Company, Inc., 1962); Louis A. Fourt, Applying Markov Chain Analysis to NCP Brand-Switching Data, (Chicago: Market Research Corporation of America, 1960); Frank Harary and Benjamin Lipstein, "The Dynamics of Brand Loyalty: A Markovian Approach," Operations Research (January-February 1962), Vol. 10, pp. 19-40; Jerome D. Herniter and John F. Magee, "Customer Behavior as a Markov Process," Operations Research (January-February, 1961), Vol. 9, pp. 105-122; Richard B. Maffei, "Brand Preference and Simple Markov Processes," Operations Research (March-April, 1960), Vol. 8, pp. 210-218; George Styan and Harry Smith, Jr., "Markov Chains Applied to Marketing," Journal of Marketing Research (February, 1964), Vol. 1, pp. 50-55; J. W. Weedlock, "A Clue to Purchase Patterns - Markov's Mathemagic," Sales Management (September 18, 1964), Vol. 93, pp. 71-72.

is not natural to organize thinking around Markov chains, but instead around the all-or-none question of learning.

Up to this point, the introduction of this study has portrayed mostly theoretical implications for the factor analytic model of brand loyalty. Comments follow on managerial implications.

Implications of Sheth's Model for Applied Marketing Research

A series of quotations about the use and value of the computed brand loyalty scores demonstrate the expected utility of the factor analytic method of measuring brand loyalty. Sheth states:

> More interesting and valuable is the model's capability to generate brand loyalty scores separately for each person in the sample. The distribution of such scores then becomes the starting point for further analysis for market segmentation.[32]

A more direct statement comes from Howard and Sheth:

> The model provides a single number for each buyer, which summarizes his pattern of purchase behavior over time. This is a much more desirable measure of the dependent variable than some measure of average tendency, such as the proportion of purchases of a brand in the total purchases of the buyer. Utilizing the brand-loyalty scores, we can discriminate between two or more groups having some explanatory or controlling variables.[33]

Concerning the use of BLS's as dependent variables in a multiple discriminant analysis, Howard and Sheth relate:

> ...since the brand-loyalty score takes into account time-lag effects and since most firms have more than one marketing strategy

---

[32]Sheth, 1968, *op. cit.*, p. 395.

[33]Howard and Sheth, *op. cit.*, p. 271.

> to manipulate in nonexperimental markets, a multiple discriminant analysis having several marketing strategies as the independent variables and the brand-loyalty scores as the dependent, classifactory variable may enable us to understand the complexity of the real-world situation.[34]

The data analysis of this study is intended to help evaluate the merits of using BLS's produced from the factor analytic model for the specified applications. How the data of this study will come about is considered next.

Simulation and Stochastic Learning Models

To obtain the basic data for this study, Monte Carlo procedures are applied to the operations of two stochastic (probabilistic) learning models resulting in simulated, individual purchase protocols. The quest for maximum control of the experimental environment prompted the use of synthetic data. Whereas Sheth's desire was to verify a theory in a natural setting, and therefore, was compelled to use small samples of purchase panel information with their attendant environmental influences, the present concern is the observation of the measuring instrument itself. That is, no theoretical position is taken on how brand loyalty is learned; the effects of the data, generated to reflect the two different learning theories, upon the factor analytic results, however, are a major concern.

In reference to the two learning models, one allows the generation of incremental learning data. The model, also called a linear model:

---

[34] *Ibid.*, p. 272.

> ...deals directly with the probability of a correct response on trial n, P_n and assumes that all subjects in an ideal experiment have the same probability of correct response on trial n, and that the successive responses are statistically independent.[35]

Accordingly then, operation of the other model provides all-or-none learning data:

> ...the assumption here is that on each trial the subject either learns completely, so that the probability of a correct response goes to one, or learns nothing, so that the probability of response remains the same.[36]

Restle and Greeno reveal their preference between the two models in the statement, "...the all-or-none model is the most interesting, and we think it is the most deserving of future work."[37]

The various samples generated from each model have 300 statistical consumers and 30 purchase trials; both are of sufficient size to insure needed reliability. From these series of purchase patterns, the proportion of ones within each trial is computed. Remember that a one represents a purchase of a preferred brand, say brand A, and a zero represents either the purchase of any other brand or no purchase al all. The proportion of brand A purchases, then, is the aggregate probability of buying brand A on that trial. These probabilities, plotted across purchase trials, result in a brand loyalty or mean learning curve.

---

[35]Frank Restle and James G. Greeno, Introduction to Mathematical Psychology, (Reading, Mass.: Addison-Wesley Publishing Company, 1970), p. 42.

[36]Ibid., p. 41.

[37]Ibid. p. 79.

Use of Simulated Learning Curves in a Brand Loyalty Experiment

Five matched pairs of aggregate brand loyalty curves, for a total of ten individual runs, were simulated. Each pair or set is specified by selected parameters, respectively, of the two learning models.[38] Effectively, then, there are ten different aggregate curves. The two aggregate curves in each pair, however, turn out to be "identical;" the underlying psychological processes on which they are predicated, however, are conceptually different.

There are two principal reasons for presenting the five sets of learning curves: (1) as a vehicle for hypothesis testing and; (2) as a practical substitute for reality.

Hypothesis Testing

Factor analytic results of the two components (incremental versus all-or-none data) of each set may permit conclusions about the stated hypotheses. Additionally, the composite outcome may define appropriate modifications to Sheth's suggested measurement technique.

The major advantage in subjecting pseudo-data to measurement may be apparent at this point. Generating the five different sets of curves on an *a priori* basis allows an unambiguous test of factor analysis as a technique for differentiating between types of learning. That is, by eliminating the contaminating environmental influences in real purchase data, *only* the *a priori* representations of the learning theories

---

[38]For the remainder of the study a matched pair of aggregate learning curves will be referred to as being a single curve or as a set number (e.g., set 5 is a negative learning curve).

are forged in the specified operations of the respective models.

With the nature of the purchase data already known, it may be reasonable to expect that the factor analytic procedure, applied to the data from either model, does accurately identify individual brand loyalty scores. However, when theory confirmation is the issue, the burden on Sheth's suggestions is compounded. Now, the factor analytic results must somehow isolate, whether through factor configuration or factor score make up, the inherent difference between gradual and all-or-none learning.

Forgetting for awhile the matched mean curves (which disguise the differences), an appreciation of the markedly diverse learning conceptions might be obtained by observing on the same graph four individual purchase probability curves derived from each model. In Fig. 1 the two sets of curves are easily discriminable, and quickly dramatize the "hidden" features of how consumers may become brand loyal.

It may be tempting at this juncture to support one of the two major learning theories, which is a legitimate prerogative. The principal purpose of the brand loyalty experiment, however, is to develop a rigorous test of whether Sheth's factor analytic technique can discriminate between the two types of learning processes and <u>not</u> to accept or reject either underlying learning process.

Substitute for Reality

To lend credence to the pseudo-data, and to illustrate possible relationships between the learning curves and marketing realities, five contextual situations were created. With imagination, other marketing phenomenon rationalities can be associated with each hypothetical curve.

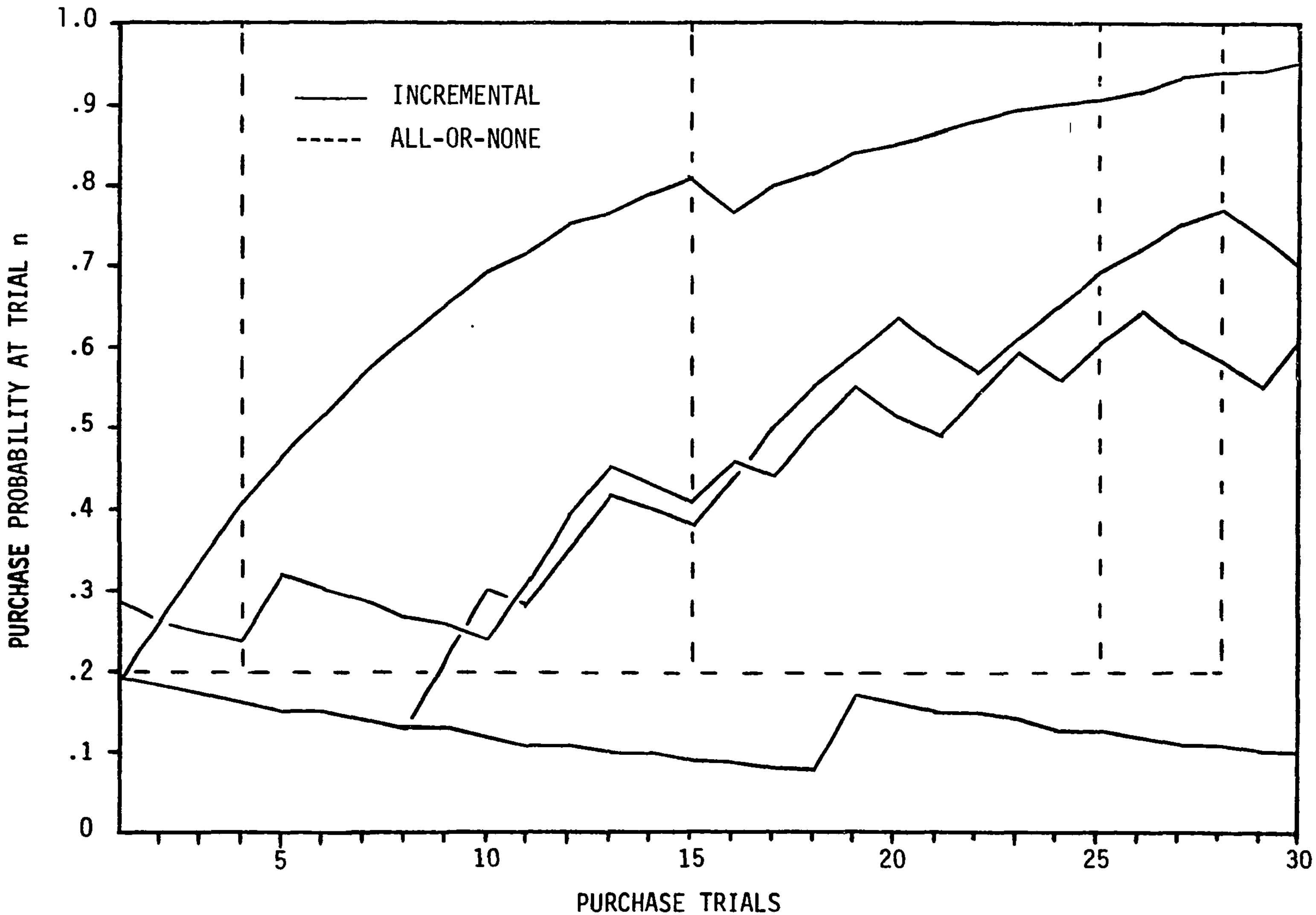

Fig. 1.--Illustration of four individual purchase probability curves derived from each learning model. Data taken from Set 3; See Table 3, p. 61.

There is nothing venerable about the situations depicted. Variations in interpretation among the five market place settings are warranted. There is flexibility and latitude in the structuring of a particular situation to the extent that it is most meaningful to either the individual market researcher or student of marketing. In a particular time period, place, and market segment, any given number of hypothesized situations could exist and probably do.

The constraints, obviously, are that one must abide by the selected learning parameters and the resultant curve in hypothesizing about a market condition.

Hypotheses

The principal hypothesis, the nature of which has been most directly addressed in Chapter I, is stated below:

Hypothesis 1: That Sheth's proposed factor analytic measure of brand loyalty does differentiate between incremental learning data and all-or-none learning data.

Related to hypothesis 1, but more specifically associated with an expected outcome of the model itself, two additional hypotheses are offered.

When the factor analytic model is applied to incremental learning data hypothesis 2 is proposed.

Hypothesis 2: That there will be one dominant factor (reflecting the mean learning curve), and correspondingly each individual will have one brand loyalty (factor) score; other factors and factor scores will be residuals or disturbance terms due to measurement error.

When the factor analytic model is applied to all-or-none learning data hypothesis 3 is proposed.

Hypothesis 3: That there will be two or more factors that account for the mean curve, and correspondingly individuals will have an equivalent number of brand loyalty (factor) scores.

If hypotheses 2 and 3 are true, then hypothesis 1 is true. Given the just stated condition, it may be concluded that the factor analytic model is a viable tool for determining how brand loyalty is learned. If factor analysis does not reveal distinguishing features of the two learning models, then at best it reduces to an empirical technique for isolating other aspects of brand loyalty data.

CHAPTER II

BACKGROUND

The appropriate starting point for providing a logical flow of supporting background information is a brief explanation of the two learning theories.

The extracted segments presented are an injustice to the elegance and comprehensiveness of the complete theories. The particular aspects are included to acquaint the reader, not familiar with the study of learning, with the theoretical positions underlying the applied stochastic learning models.

Hull's Behavior System

The primary sources used here are Hilgard and Bower[1] and Bugelski.[2] The system is a behaviorism, and as such falls into the family of theories which also includes those of Guthrie and Skinner. In general, the basic logic of the system operates within the broad framework of the classic S-O-R formula. Hull's system is a massive elaboration of the formula: where the stimulus (S) affects the organism

---

[1]Ernest R. Hilgard and Gordon H. Bower, Theories of Learning, 3rd ed., (New York: Appleton-Century-Crofts, 1966).

[2]B. R. Bugelski, Psychology of Learning, (New York: Henry Holt and Company, Inc., 1956).

(O), and what happens as a consequence, the response (R), depends upon O as well as upon S.

Hull proposed a reinforcement theory as the primary condition for habit formation, while rejecting contiguity (S-R associations) alone, or a dual theory. In later years, although some modifications were made in interpretation, he held to the reinforcement principle as central to learning. The reinforcement theory of Hull requires, in terms of a primary reinforcing situation, either drive reduction, as in need satisfaction, or drive-stimulus reduction, as in the satisfaction of a craving rather than a need. Hull pictured the learning process and his principle of learning by reinforcement as serving individual survival just as Darwin's laws of natural selection are presumed to operate in the survival and evolution of species.[3]

The evocation of a learned response implies the involvement of six major processes. Some of them are nurtured by environmental events; mostly, however, the processes are inferred intervening variables (connecting links between conceptualizations and observable behavior). These processes are reinforcement, generalization, motivation, inhibition, oscillation, and response evocation. It is not the purpose here to explain the interrelationships of the six processes. A condensed excerpt of the reinforcement--habit strength relationship, however, is revealing and germane. Habit strength ($_{S}H_{R}$) is the result of a reinforcement of S-R connections in accordance with their proximity to need reduction.

---

[3] Ibid., pp. 156-157.

Concerning the reinforcement and habit strength relation and the developments growing out of it, the essence of the learning theory is observed: (1) the operation of rewards, (2) the effects of repetition, (3) and the gradients of reinforcement. The implications, as stated by Hilgard and Bower, follow:

> 1. Learning depends upon contiguity of stimulus and response closely associated with reinforcement defined as need reduction....
> 2. The course of learning described as a simple growth function is based on the implied assumption that the increment of habit strength with each reinforcement is a constant fraction of the amount remaining to be learned. Because more remains to be acquired early in learning and little remains late in learning, the result is a curve of decreasing gains, very familiar in laboratory studies of learning.
> 3. The upper limit of learning tends to be at a maximum when need reduction is great, when delay between response and reinforcement is short, and when there is little separation between the conditioned stimulus and the response to be acquired.[4]

As one might foresee, when modifications are added to handle multiple S-R learning, the theory becomes unwieldy. Despite such complexities, the theoretical system has been used in practice as a qualitative device for explaining and ordering experimental results. The traditional theory is rich in concepts but (in its attempt to be comprehensive) has become so complex that quantitative predictions are virtually impossible to obtain.

Dissatisfied with this state of affairs mathematical psychologists in formulating models for learning curtailed drastically the scope of the theorizing. The focus of attention is on firm control of only a few variables, and the experimental arrangements are constructed to

---

[4]Hilgard and Bower, op. cit., p. 158.

to minimize the influence of other factors not under consideration. The general approach is known as mathematical learning theory and it is explored after Guthrie's contiguity theory is discussed.

Guthrie's Contiguity Theory

The basic principle of "association by contiguity" appears to be universally adopted and acceptable to all learning psychologists. Without exception, contiguity in time is presupposed as a basis for learning. It was given above that Hull assumes contiguity of response and reinforcement; Guthrie assumes contiguity of stimulus and response. While they differ on what events are contiguous, on contiguity they agree.

Guthrie's one law of learning (association), which forms the basis for all learning phenomenon, was stated by Guthrie as follows:

> "A combination of stimuli which has accompanied a movement will on its recurrence tend to be followed by that movement."[5]

There is a notable simplicity about the statement in that it avoids mention of drives, of successive repetitions, of rewards or of punishments. Stimuli and movements in associative pairs: that is all. A second statement is needed to complete the basic postulate about learning: "A stimulus pattern gains its full associative strength on the occasion of its first pairing with a response."[6]

The statement is suggestive of a recency principle.

---

[5]Edwin R. Guthrie, The Psychology of Learning, (New York: Harper, 1935), p. 26.

[6]Ibid., (1st rev. ed., 1952), p. 30.

Consider, if learning is complete in one trial, then the last response to a particular stimulus combination will be the same response when the stimulus combination next recurs. For Guthrie, the recency principle becomes a primary, if not _the_ primary principle of learning.

The meaning Guthrie attaches to the term requires some elucidation. It does not refer to recent events, as one might gather. Suppose that some trial-and-error situation is being considered where a person might try various responses, say A, B, C, D, in succession. Assume D is the correct response. Since D is the last response made it terminates the trial. On the next occasion D might occur as the first attempt at solution. In any series of learning trials that terminate with a correct response, the correct response is obviously, and must necessarily be, the most recent response made in the situation.

Guthrie holds that this function is a sufficient and adequate explanation of learning. He states that all learning situations are basically the same in the respect that they have to terminate with some response. How can an individual ever learn if he once begins to respond incorrectly? The only answer is that for one reason or another, due to changes in the situation which the experimenter or teacher is not keen enough to note, the individual does happen to make the correct response. This response will then be retained, says Guthrie, barring some change in the stimulus situation which would invite the former or another response.

From Hull's point of view, recency only works because it normally characterizes the response which was also rewarded, or reinforced. By itself recency would have no effect on learning. Since it is difficult to separate recency from reinforcement in many trial-and-error learning experiments, the argument can be carried on with as much fervor as the debators can muster.

Mathematical Learning Theory

Before the incremental and all-or-none learning models are described, a broad overview of mathematical learning theory is presented. The references cited are a rich source for detailed exposure.

Definition

The label "mathematical learning theory", in a sense is misleading if not inaccurate. Mathematical learning theory is not proposed as a substitute for other theories of learning. In fact, it is not a theory at all. More appropriately, the term embodies an orientation or method in theory construction without specific reference to the substantive postulates expressed in particular theories. Atkinson, Bower, and Crothers succinctly define mathematical learning theory as the use of explicit mathematical means for theory construction and research on learning.[7] The following discussion embellishes certain features embodied in the definition.

Stochastic Models

Almost all current learning models characterize behavior as a probabilistic rather than as a deterministic process; the underlying assumption is that of a stochastic process. Simply defined, a stochastic

---

[7]Richard C. Atkinson, Gordon H. Bower and Edward J. Crothers, An Introduction to Mathematical Learning Theory, (New York: John Wiley and Sons, Inc., 1965), p. 1.

process is a sequence of outcomes that can be analyzed by using probability theory.

Since 1950 the modern point of view in mathematical learning theory has extended considerably the probabilistic approach to behavior theory. The psychological orientations condoning such an approach dictate that variability in behavior be built into the structure of these types of models.

Two logical interpretations allow regarding behavior as a stochastic process. The first is that, in fact, behavior is inherently probabilistic. The other interpretation is that behavior is completely deterministic, but that minute fluctuations in the individual and the stimulus conditions over time produce the observed probabilistic behavior. Bush and Mosteller favor the notion that behavior is intrinsically probabilistic.[8] However, they go on to point out that such an assumption is not a necessary condition in their stochastic models for learning.

Parameter Invariance: A major advantage of stochastic models is that they require only the classification of potential responses into response categories. The experimenter then needs only to count the frequency of occurrence for each response category. The relationships of these proportions are analyzed as appropriate. The point is that the theorist is not "defeated" before he even gets started due to measurement problems regarding dependent variables, or inferentially, his independent

---

[8] Robert R. Bush and Frederick Mosteller, Stochastic Models For Learning, (New York: John Wiley and Sons, Inc., 1955). p. 3.

variables. Traditionally, the emphasis in learning theory construction was a predisposition toward "theoretical" dependent variables such as response strength, excitation-inhibition, cognitive maps, and so on. Obviously, such theories are testable only to the extent that the variables themselves are measurable. The stochastic models utilized in this study avoid these several traditional problems in that the response categories are preclassified.

For example, the reader familiar with the traditional theories of Hull and Guthrie may be perplexed at first exposure to mathematical models for learning. Usually, the mathematical model is specified quite simply, without attaching a list of postulates about drive level, amount and delay of reinforcement, stimulus intensity, and so on. Generally, in the use of the models, these classical variables have not been the topic of the investigation and are presumed fixed throughout the experimental situation. It is true, however, that such variables may well influence the estimated parameter values of a given model, i.e., the probability of learning or the rate of learning. The selected parameters are like initial-state descriptions that are used but not derived from the theory.

A parameter used in the operation of a model should reflect the constant features of the individual and the stimulus condition, which hopefully remain constant in any one experimental run under constant environmental conditions. Surely, it is unreasonable to require that all parameters remain unchanged from one experimental situation to another.

<u>The Theory/Model Relationship</u>: Usually it is not necessary to

draw a distinction between the terms theory and model, but in this study it is useful to distinguish between the purely formal structure (the model) and the underlying conceptualization (theory) on which the structure is based. The reason for this distinction is that, in this study, the same formal structure arises from quite different conceptualizations. For example, the same mathematical expression for the mean learning curve is derived from the all-or-none learning model and the linear operator model.[9]

The all-or-none theory is quite different from gradual learning theory both in concept and in various details of the data to be expected. Yet, they make the same prediction regarding the aggregate learning curve. Thus two consequences must be treated. If the interest is only in the average learning curves (dependent variable), then it does not matter which theory is used. However, if the requirement is to distinquish between the theories, then other aspects of the data besides the mean learning curve must be examined.

In the use of stochastic models the comprehensiveness of the traditional theorizing is foregone: an attempt is made to focus on only a few variables, and the experimental arrangements are constructed to minimize the influence of other factors not under consideration. This is the price to be paid for tractable theories.

---

[9]Frank Restle and James G. Greeno, Introduction to Mathematical Psychology, (Reading, Massachusetts: Addison-Wesley Publishing Company, 1970), pp. 34-36; Clyde H. Coombs, Robyn M. Dawes, and Amos Tversky, Mathematical Psychology, (Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1970). p. 294.

Underlying Psychological Theory

Certainly, a mathematical learning model selected for investigation must be an abstraction of the basic axioms of the given classical learning theory. Around 1950 two related branches of development in learning models emerged. The class of models, called operator models, which permits the stages of learning to be infinite in number, was initiated by Bush and Mosteller.[10] The other class of models, called state models, where only a finite number of learning stages are possible, was initiated and developed by Estes in 1950.[11] His basic theoretical ideas are collectively known as stimulus sampling theory. Stimulus sampling theory is briefly described on subsequent pages in the State Models: All-or-none Learning section.

These two approaches are parallel in structuring the probablistic nature of learning. The learning process is conceived of as a sequence of discrete trials: each trial consists of a stimulus situation to which the individual responds (selects one) from a set of alternative responses in accordance with an associated set of probabilities; the response is followed by an outcome, which may induce changes in the probability values before the next trial. Of course, how and why this probability flow from trial to trial and the resulting distributions come about constitute their similarities and differences.

---

[10]Robert R. Bush and Frederick Mosteller, "A Mathematical Model for Simple Learning," Psychological Review, Vol. 58 (1951), pp. 313-323; Bush and Mosteller, op. cit.

[11]W. K. Estes, "Toward a Statistical Theory of Learning," Psychological Review, Vol. 57 (1950), pp. 94-107.

The distinction of infinite and finite states in the two classes of models is of no particular theoretical significance; however, the respective underlying conceptions of the learning process that augur the distinction is paramount.

Before delineating in some detail the operational aspects of the two "simple" models, each representing the two broad classes of stochastic models, a comment on the rationale for their selection is offered. The two-operator linear model and the all-or-none model are appropriate candidates for comparison because these two models differ sharply in their assumption about how an individual stimulus-response association is learned. According to the linear model each reinforcement is assumed to bring about a change in the probability of a "correct" response. Learning, which occurs on each positively reinforced presentation, is gradual because a number of increments in p_n (probability of a correct response) are needed before this quantity is increased to say, a near unity value. In this sense, the linear model is a probabilistic analog of the Hullian incremental learning theory.

By contrast, in the all-or-none model the effect of a single reinforcement is either to produce complete learning of the individual stimulus-response association, or no learning at all. The model reflects the stimulus-response association theory in the Lashley[12] and Guthrie tradition.[13]

---

[12]K. S. Lashley, Brain Mechanisms and Behavior, (Chicago: University of Chicago Press, 1928).

[13]E. R. Guthrie, "Psychological Facts and Psychological Theory," Psychological Bulletin, Vol. 43 (1946), pp. 1-20.

As a starting point, it may be sufficient to discover to what extent the data permit a choice between these diverse formulations. The expectation is that there would be less differentiation among models that are more similar. Mathematically, the all-or-none model is a prototype of Markov models, whereas the two-operator model is basic among linear models. Processes of both general types have played prominent roles in quantitative theorizing in other sciences, and it is not surprising that analogous processes, e. g., purchase behavior, are extant in marketing.

Operator Models: Incremental Learning

A plausible experimental situation is presented and a model constructed for it in order to identify and describe the concept of an operator.

The Concept of an Operator

Consider the two brand experimental situation in which a consumer learns to buy brand A. On the first trial it is reasonable to assume that she has a probability of .50 of purchasing brand A, i. e., $p_1 = .50$. Suppose she buys A and is satisfied. The expectation might be that her probability of buying A next time, p_2, to be greater than .50, perhaps as high as .70.

Suppose on the next trial the consumer buys brand X and the outcome is less than expected, i. e., no satisfaction or even some dissatisfaction. Again, it would be expected that the probability of this consumer buying A on the third trial, p_3, to be greater than p_2, perhaps .85.

Now suppose on the third trial she buys A and is satisfied again. When this happened on the first trial the probability of buying A increased from .5 to .7, an increment of .2, but clearly it cannot increase the same amount again. The issue then is how to describe the effect of buying brand A and being satisfied on the probability of buying brand A.

An alternative is that the increment in probability is a constant proportion or fraction of the amount of change that is still possible. By making this assumption, the effect of buying A and being satisfied on the nth trial results in an increment to p_n that is proportional to the amount to be learned, $1-p_n$, say $B(1-p_n)$. In other words, B, the <u>rate</u> of learning is constant while the proportional amount to be learned, $1-p_n$, may be diminishing. In this case the probability on the n + 1st trial is:

$$p_{n+1} = p_n + B(1-p_n) \quad (1a)$$

$$= (1-B)p_n + B; \quad (1b)$$

i. e., the probability on the nth trial is operated on by this linear transformation and becomes the probability on the next trial.[14] That is p_{n+1} is a linear function of p_n. Hence, this expression is called a <u>linear operator</u> and would be applied on any trial in which the consumer bought brand A and was satisfied.

What happens to p when the consumer buys brand X and is not satisfied? Should the same operator apply? Probably not. The equal

[14]The unconventional notation shown here and in the remainder of the study is used so that the computer program notation is paralleled as much as possible.

learning rate condition suggests that consumer satisfaction or dissatisfaction have "equal but opposite" effects on subsequent purchase behavior; seldom is this notion justified, either by psychological theory or by data.[15] Thus, the first operator is designated Gp, with the letter B identifying the parameters, and the second operator Lp is:

$$Lp_n = (1-B1)p_n + B1,$$

where the parameter B1 represents the portion of $(1-p_n)$ that is added to p_n as a result of buying brand X and not being satisfied. Given the consumer's sequence of purchases through the first n trials, then, one could calculate the probability of her buying brand A on the n + 1st trial.

Necessary Specifications

The preceding gives an introductory conception of what a stochastic model looks like and what an operator is, so attention now is turned to those conditions that need to be specified in order to render the basic model experimentally operational.

Assumptions: First, the necessary assumptions are few and straightforward. As already identified, and accounted for within the mathematical properties of the model itself, is the notion that learning is incremental; a basic postulate in the Hullian system. Two further assumptions are that all subjects have the same probability, p_n of a "correct" response on trial n, and that the successive responses are statistically independent. The latter independence-of-path assumption says that the probability of any response on this next trial is de-

---

[15]Bush and Mosteller, *op. cit.*, 1955, p. 108.

pendent only on its probability on the preceding trial and the event that occured, i. e., reward or punishment. If, for example, two consumers have the same probability for a given brand purchase, it does not matter if their detailed histories are different. Any differences in pattern and frequency of their respective sequences are irrelevant.

The Two-Operator Linear Model: The Fixed-Point Form: A very useful form of the two-operator linear model is known as the fixed-point form. A new parameter, the asymptote, needs to be introduced and is designated as A_i. In order to conform to the probability system, the limiting values of the asymptote which may be chosen or predicted for each operator range from equal to or greater than zero to an eventual limit of unity.

The asymptotic value specified in the respective operators identifies the maximum and the minimum portion of total learning that can be accomplished in a particular experimental setting. The two operators take the form:

$$Gp_n = (1-B)p_n + B(A_1)$$

$$Lp_n = (1-B1)p_n + B(A_2).$$

This form of the linear operators gets its name because when $p = A_i$ the operator becomes the identity operator, i.e., $Gp = p$, or $Lp = p$. The name fixed-point form comes from the fact that A is a limit point, a fixed value that p approaches when the operator is applied repeatedly. When p reaches its limiting value, it stops changing.

The nature of the purchase protocols to be generated for this study is similar to many learning experiments where the task consists of acquiring a particular response so that the probability of that response approaches unity and the probability of other responses approaches zero. Hence, the most commonly occurring linear operators are of the form:

$$Gp = (1-B)p_n + B \qquad (A_1 = 1) \qquad (2a)$$

$$Lp = (1-B1)p_n. \qquad (A_2 = 0) \qquad (2b)$$

The first is an acquisition or gain operator with the limit point $A_1 = 1$. The other operator, Lp, is an extinction or loss operator which, with repeated applications, moves p to zero.

Restrictions on Parameter Values: Each operator has two parameters: B, A_1 and B1, A_2, respectively. The restrictions are: $0 \leq B_i, A_i \leq 1$.

For the brand loyalty curves generated here, the asymptotic values will always be one and zero, unless otherwise specified. The learning rates may be, $B \lesseqgtr B1$. Generally, $B > B1$ is used because the outcome is a positive aggregate learning curve; if $B = B1$ a Bernoulli function occurs; and if $B < B1$ is used, then an extinction (negative learning) curve develops. These three relationships hold only if $A_1 = 1$, and $A_2 = 0$.

As a guideline, for most experiments, it is found that estimates of B_i tend to run below .20.[16]

State Models: All-Or-None Learning

Up until now, hopefully for the purpose of minimizing ambiguity,

---

[16]Coombs, Dawes and Tversky, op. cit., p. 264.

the all-or-none learning model representing contiguity theory has been undifferentiated. In reality, a variety of all-or-none models exist and are being investigated. The one used in this study, which is explained below, is called the one-element model. "One-element" and "all-or-none" model may be used interchangeably. Before the one-element model is described, the general operational framework, stimulus sampling theory is explicated.

Stimulus Sampling Theory

As in mathematical learning theory in general, stimulus sampling theory is not a learning model per se, but is to be used as a scheme for model building. Because of the very general structure of the theory it serves as a conceptual framework for generating a class of stochastic processes.

The following brief description is borrowed heavily from various sources.[17] Of course, the original work on stimulus sampling theory by Estes[18] and Estes and Burke[19] may be referenced.

Stimulus sampling theory views behavior as being elicited by antecedent stimulus events that are "associated," "connected," or "conditioned" to responses. The primitive terms of the theory are stimulus, response, association, and reinforcement. The association term refers to a functional connection between a stimulus and a response

---

[17]Atkinson, Bower and Crothers, op. cit., p. 344; Hilgard and Bower, op. cit., pp. 338-345

[18]Estes, loc. cit.

[19]W. K. Estes and C. J. Burke, "A Theory of Stimulus Variability in Learning," Psychological Review, Vol. 60 (1953), pp. 276-286.

i.e., that response A_i is conditioned to stimulus S_k.

The term "reinforcement" refers to a theorectical event that can produce changes in associations of stimuli present at the moment the reinforcing event occurs. Reinforcing events are classified according to the stimulus-response associations that are left unchanged by that event. Thus, if response classes A_1, A_2..., A_r have been identified then there will correspond r reinforcing events, labeled E_1, E_2..., E_r. If stimulus S_k is already associated to response A_i, then E_i is that reinforcing event that leaves the S_k - A_i association unchanged; if stimulus S_k is associated to some other response A_j , then the occurrence of an E_i reinforcing event when S_k is present will tend to change the association of S_k so that it is now associated with A_i instead of the previous A_j.

The change in association of a stimulus is assumed to be a discrete all-or-none switch; a previous association is erased (counter-conditioned, unlearned, etc.) when a new response is associated with a stimulus. This notion is formalized in the theory by saying that at any point in time each stimulus is associated with exactly one response. The reinforcement postulates of the system are responsible, then, for making it a dynamic, changing process.

There has been a proliferation of model types derived from the basic notions of stimulus sampling theory e.g., component models, pattern models, small-element models, and others.[20] The only model that

[20]Consideration of the principal aspects of these various types goes beyond the scope and needs of this study. For an overall view of the state models see (Coombs, Dawes, and Tversky, op. cit., pp. 279-306).

is discussed in this study--the one-element model--is of the class, small-element models.

One-Element Model

The one-element model was selected for this study for two reasons; (1) this model has had an important role in the recent development of mathematical learning theories and, (2) the relatively simple mathematical and conceptual structure make it preferred.

Assumptions: Stated in the context of a purchasing process, for a single consumer and a single stimulus-response association the assumptions of the all-or-none model are as follows:

1. The stimulus member (a brand in a product category) of each S-R association (buy or not buy a particular brand) in the product category can be represented mathematically by a single hypothetical stimulus element. The marketer may question this assumption on the grounds that a brand as a stimulus has several components e.g., color, size, price, etc, and therefore, should be identified with several stimulus elements. Obviously for ease of presentation, and also, when the products are reasonably differentiated, there is justification for assuming only one stimulus element per brand: hence, the name, one-element model. In essence, the diversity among brands enables the purchaser to associate her response to the stimulus characteristics as a whole.

2. Each time a particular brand is exposed to the purchaser, she samples the stimulus element. The word "samples" means that the purchaser considers the brand-purchase--not-purchase association each

time it is presented.

3. On each purchase trial the stimulus element (Brand A) is in exactly one of two learning states. Either the element is in state L ("learned" as associated to the purchase of brand A) or the element is in state U ("unlearned" and not associated to the purchase of brand A).

4. On each reinforced (satisfied or dissatisfied) purchase trial, the probability of a change of the element from U to L is c; the probability of a transition from L to L is one. That is, if the purchaser has not formed the association (brand A with purchase brand A) prior to the current reinforcement of this association, she now forms it with probability c. If the association was formed prior to the current reinforcement, it is preserved with probability one.

5. If brand A is in state L when the purchaser responds, she purchases brand A with probability one. If brand A is in state U when the purchaser responds, her purchase of brand A response is a guess with probability g.

6. At the beginning of the brand loyalty learning experiment, the brand is in the unconditioned state U.

<u>The Learning Process as a Markov Chain</u>: Assumptions three, four, and five provide the all-or-none feature of the learning process. The buyer is assumed to have learned (perfectly) the stimulus (brand A) --response (purchase brand A) association (i.e., the association has become conditioned and the probability of buying brand A is 1) or she knows nothing and is just guessing.

These assumptions allow a representation of the learning process as a Markov chain, with the following structure:

$$
\begin{array}{ccccc}
 & & \multicolumn{2}{c}{\text{trial } n+1} & \\
 & & L_{n+1} & U_{n+1} & \Pr(\text{buy brand A} \mid \text{row state}) \\
\hline
\text{trial } n & L_n & 1 & 0 & 1 \\
 & U_n & c & 1-c & g \\
\hline
\end{array}
\qquad (3)
$$

On the left is the matrix of transition probabilities between states from trial to trial. The bottom row of this matrix represents the first statement in assumption number four, and the top row expresses the second part of that assumption. It is given, then, that each row of the transition matrix must sum to one. Additionally, every trial gives the consumer the same opportunity to learn. In the linear model this led to an assumption that the learning rate B_i is the same for all trials. In all-or-none theory, it is the probability of learning c that is constant on all trials. On the right is a probability vector which indicates, for each state on trial n, the probability of buying brand A on that trial; as specified by assumption number five.

A sequential process is termed a Markov process (also referred to as Markovian) when the probability that state S_i occurs on trial n following any sequence of states s is independent of all states in that sequence except the one that occurs on trial n-1. A common way of characterizing Markov processes is to state that they are path-independent. The only important consideration is the state of the process on trial n-1. A sequential process is called a stationary process when the probability that state S_i follows a given sequence s is independent of the trial numbers on which s occurs. That is, the probability of moving from the unlearned to the learned state on any particular trial is assumed to be a constant. When a process is both stationary and Markovian it is called a Markov chain. For more detail see Kemeny and Snell.[21]

The Markov chain described by Eq. (3) is an absorbing Markov chain. An absorbing Markov chain has at least one absorbing state and it can be reached from any state (not necessarily in one trial). Note that state L is absorbing because it cannot be left and it can be reached from state U.

In general, terminal learning may not be perfect i.e., the probability of switching from state L on trial n to state U on trial n + 1 is >0, or alternatively, the probability of staying in state L on repeated trials is <1. In other words no state in the Markov chain is now absorbing. Markov chains with the property that it is possible to go from every state to every other state (not necessarily in one trial), are called ergodic. The consequences of such a condition are explained at the time actual data sets are formulated.

It is mentioned again that, in spite of an underlying theory of learning as an all-or-none process, the mean learning curve (with its averaging effect) is gradual and obscures the existence of such a process.

With the conceptual and operating features of the two models now accounted for, the procedures required for simulating these mecha-

---

[21]J. G. Kemeny and J. L. Snell, "Markov Processes in Learning Theory," Psychometrika, Vol. 22, pp. 221-230; J. G. Kemeny and J. L. Snell, Finite Markov Chains, (Princeton, New Jersey: D. Van Nostrand Company, 1957).

nisms are documented in the following section.

Procedure For Simulating Individual Protocols

With computer applications, one can estimate parameter values by trial-and error techniques: try a value of each, compute the whole theoretical learning curve and compare it with an obtained curve, continue modifying parameters until a good fit is made. Since real data are not used in this study, the only requirement is that logical parameters be "plugged" into the model. Of course, once a curve is generated for either learning model, a parameter search process must be carried out in order to match it with the remaining model.

The reason for using a computer in behavioral simulation is one of speed and accuracy. Speed is particularily important in this experiment because the programs are developed for repeated application of rather simple rules. For other implications and applications of computer simulation see Laughery and Gregg,[22] and Newell and Simon.[23]

Before discussing the use of simulated learning curves in this experimental setting, it may be instructive to graphically portray the simulation procedure that is occurring in the operation of the respective learning models.

The use of a Monte Carlo procedure involves generating pseudo-data by following precisely the rules specified by the model and accessing a random number generator whenever an event occurs in the

---

[22]K. R. Laughery and L. W. Gregg, "Simulation of Human Problem-Solving Behavior," Psychometrika, Vol. 27 (1962), pp. 265-282.

[23]A. Newell and H. A. Simon, "Computers in Psychology," in R. D. Luce, R. R. Bush, and E. Galanter, eds., Handbook of Mathematical Psychology, (New York: Wiley, 1963).

process which is probabilistically determined. The output from a Monte Carlo run then is a representation of how real data would look if the learning model were correct and if the parameter(s) had the value(s) used in the Monte Carlo computations. Bush and Mosteller,[24] who were among the first to use Monte Carlo methods in testing psychological models, coined the term "stat-rat" to describe the hypothetical organism of a Monte Carlo run. Presumably then for this application the term "stat-con" is an appropriate substitution for hypothetical consumer.

Statistical Consumer: Two-Operator Linear Model

In making a set of Monte Carlo computations, first set up the desired constants for the operators and a starting probability in numerical form. The rules of the illustrated model are: initial p_n (probability of buying brand A) = .20 for trial 0; B (rate of learning in Gp) = .10; B1 (rate of learning in Lp) = .05.

Gp = (1-.10).20 + .10 (if consumer is positively reinforced)

Lp = (1-.05).20 (if consumer is negatively reinforced)

The stat-con now can be carried mechanically through a sequence of purchase trials. Of interest is: (1) recording a one for purchase of brand A; and a zero for not-brand A; (2) which operator is to be applied on each trial, applying it; and (3) recording the p value at each stage.

In Table 1 the second column gives the p value for each trial.

---

[24]Bush and Mosteller, op. cit., pp. 128-131.

TABLE 1

ILLUSTRATION OF OUTCOME FOR 30 PURCHASE TRIALS OF ONE STAT-CON. THE OPERATIONS ARE $Gp = (1-.10)p+.10$ AND $Lp = (1-.05)p$, AND $p = .20$.

Trial Number	p Value	Random Number	Operator	Purchase Protocol
0	.20	25	Lp	-
1	.19	69	Lp	0
2	.18	92	Lp	0
3	.17	08	Gp	0
4	.25	03	Gp	1
5	.33	31	Gp	1
6	.40	38	Gp	1
7	.46	37	Gp	1
8	.51	72	Lp	1
9	.49	80	Lp	0
10	.46	15	Gp	0
11	.51	01	Gp	1
12	.56	64	Lp	1
13	.54	97	Lp	0
14	.51	74	Lp	0
15	.48	75	Lp	0
16	.46	19	Gp	0
17	.51	49	Gp	1
18	.56	29	Gp	1
19	.53	72	Lp	1
20	.58	11	Gp	0
21	.62	17	Gp	1
22	.66	15	Gp	1
23	.69	56	Gp	1
24	.72	57	Gp	1
25	.75	72	Gp	1
26	.78	41	Gp	1
27	.80	43	Gp	1
28	.82	87	Lp	1
29	.78	95	Lp	0
30	.74	--	--	0

Notes:

A 1 or 0, the result of the p value-random number match, appears on the next succeeding row of column 5.

The Gp or Lp is on the same row as the p value to which it is applied. The result of the applied operator, the new p value, is on the next succeeding row of column 2.

The third column is a set of 2-digit random numbers. Considering the uniformly distributed random numbers from 01, 02, ..., 99, 00, those numbers $\leq p_n$ dictate applying Gp, and those numbers $> p_n$ on a particular trial go to applying Lp. The fourth column gives the particular operator to be applied on each trial, and column five gives a one for the purchase of brand A and a zero for the purchase of not-brand A for each trial.

Inspection of Table 1 clarifies the procedure used. A p value of .20 was chosen for trial 0; the probability of buying brand A. The first random number chosen was 25, which is above 20, (p value decimal points are omitted) and so the Lp (row 0) was applied to the initial p value of .20(row 0) and this gave a new value of .19 (row 1). Pertaining to the purchase protocol, a dissatisfying reinforcing even occurred prior to trial one (row 2) which caused not-brand A to be purchased; thereby giving a 0 (row 2) for the first element in the protocol. Then another random number 69, was selected and it was above 19 (a 0 in row 3, another not-brand A purchase), indicating again the Lp should be applied. On trial number 3 (row 4) the random number was 08 which is less than the p value of 17. The consequence was the first purchase of brand A (row 5). Hence Gp (row 4) was applied to .17, giving a new p value of .25 (row 5).

The calculations in Table 1 demonstrate the random element inherent in any probabilistic model. A second stat-con would have produced a protocol different from that in Table 1, because her sequence of random numbers would differ from that of the first consumer. After running a large group of stat-cons, all with p_n = .20, B = .10, and B1 = .05, the mean proportion of brand A purchases on each trial may be calculated. Such mean learning curves are provided in Chapter 3. First, however, simulating the operations of the all-or-none model is demonstrated.

Statistical Consumer: All-Or-None Model

The purchase data, simulation procedure is basically the same for the all-or-none model. However, now, two random numbers must be selected for each purchase trial: first, to determine whether brand A, or not-brand A was purchased; and secondly, to determine whether on that same trial learning has occurred.

Again, the desired probabilities for the model are set up, remembering that both the purchase probabilities for the respective states, and the transition probabilities are fixed throughout the given experimental run.

The conditions in the example below are: g(probability of purchasing brand A in unlearned state) = .20; c(probability of switching from unlearned to learned) = .01; the .00 in the transition matrix demonstrates that the probability of switching from L to U; and the 1.0 in the vector is the probability of purchasing brand A once in the learned state. It is assumed that the buyer starts in the unlearned state.

		trial n+1		
		L_{n+1}	U_{n+1}	Pr(buy brand A|row state)
trial n	L_n	1.0	.00	1.0
	U_n	.01	.99	.20

In Table 2 the second column gives the brand A purchase probability on each trial. The third column is a set of 2-digit random numbers. The fourth column gives a one for the purchase of brand A, and a zero for the purchase of not-brand A. The probability of switching states, and the corresponding random numbers for each trial are given in columns five and six respectively. Column seven shows, trial by trial, whether the brand A--purchase brand A association has been learned by the purchaser.

Observe in Table 2 that the first random number, 65, is greater than 20, therefore, not-brand A was purchased on the first trial (authorizing a 0 in the second row of column 4). Note also, that the second random number (second row of column 6), 05 is greater than .01, the probability of learning the brand A--purchase brand A association, hence no learning occurred on the first trial. The stat-con made brand A purchases on trials 6, 9, 18, 21, and 23 demonstrating that she picked brand A about 20% of the time without learning taking place. On purchase trial 23 the stat-con did learn to buy brand A. The consequence was that she bought brand A inclusive of purchase trial 24 through 30. It is not necessary that a purchase of brand A immediately precede the switch from U to L as it did in this case.

To duplicate the data simulation procedures manually for any large number of stat-cons is obviously laborious and the opportunity for error is great. The only feasible way to generate a large number of purchase protocols is to develop computer programs that accomplish the task quickly and with perfect accuracy. This is exactly what was done. The two programs are provided at Appendix A and B.

TABLE 2

ILLUSTRATION OF OUTCOME FOR 30 TRIALS OF ONE STAT-CON. THE PURCHASE PROBABILITIES ARE STATE U = .2, STATE L = 1.0, AND TRANSITION PROBABILITIES ARE; U TO L = .01, L TO U = 0.0.

Trial Number	Probability of Purchase	Random Number	Purchase Protocol	Transition Probability	Random Number	Learn State
0	.20	65	-	--	--	U
1	.20	60	0	.01	05	U
2	.20	41	0	.01	39	U
3	.20	51	0	.01	88	U
4	.20	53	0	.01	75	U
5	.20	15	0	.01	46	U
6	.20	26	1	.01	91	U
7	.20	45	0	.01	60	U
8	.20	03	0	.01	69	U
9	.20	91	1	.01	84	U
10	.20	54	0	.01	10	U
11	.20	91	0	.01	48	U
12	.20	41	0	.01	92	U
13	.20	91	0	.01	25	U
14	.20	25	0	.01	53	U
15	.20	96	0	.01	83	U
16	.20	77	0	.01	99	U
17	.20	12	0	.01	32	U
18	.20	75	1	.01	63	U
19	.20	85	0	.01	41	U
20	.20	09	0	.01	80	U
21	.20	47	1	.01	08	U
22	.20	10	0	.01	57	U
23	.20	12	1	.01	01	L
24	1.00	35	1	.00	50	L
25	1.00	36	1	.00	37	L
26	1.00	83	1	.00	22	L
27	1.00	97	1	.00	27	L
28	1.00	10	1	.00	97	L
29	1.00	42	1	.00	84	L
30	1.00	--	1	.00	78	L

Notes:

A 1 or 0, the result of the purchase probability-random number match appears on the next succeeding row of column 4.

For a given trial, the resulting learn state is on the same row as the transition probability-random number match.

The next and last step in the experimental development highlights the main features and uses of factor analysis (principal components analysis) including the necessary variations engendered when applied to learning data. Specifically, how principal components analysis, based on Sheth's suggestions, is used in this study is further developed in Chapter 3--Method.

Basic Concepts of Factor Analysis: Principal Components Analysis

Until now the term, factor analysis, has been used generically in this study. Subsumed within the general classification are several variations of the same basic approach to the analysis of complex data. For example, principal components as a factor analytic technique was used by Sheth and so is described here. There are other sources available.[25]

First, is presented a general description and purpose of principal components analysis, followed by two examples of usage. Finally, attention is turned to the application of principal components analysis to learning data.

A general note on terminology: in this study the terms, "factors," "components" or "principal components" and "reference learning curves" or "reference curves" may be interpreted as synonomous. Also,

---

[25]H. H. Harmon, Modern Factor Analysis, 2d ed., (Chicago: University of Chicago Press, 1960); W. W. Cooley and P. R. Lohnes, Multivariate Procedures for Behavioral Sciences, (New York: John Wiley and Sons, Inc., 1962); M. G. Kendall, A Course in Multivariate Analysis, 2d ed. (New York: Hartner Publishing Company, 1965); Paul Horst, Factor Analysis of Data Matrices, (New York: Holt, Rinehart and Winston, Inc., 1965); Benjamin Fruchter, Introduction to Factor Analysis, (Princeton, New Jersey: D. Van Nostrand Company, Inc., 1954); L. L. Thurstone, Multiple Factor Analysis, (Chicago: University of Chicago Press, 1947).

the terms, "factor scores," "component scores," and "brand loyalty scores" may be used interchangeably.

Principal Components Analysis

This section briefly will detail the algebra of components analysis as it is traditionally used and, in the following section, in the context of Sheth's and Tucker's specialized application to learning data.

Components analysis, as traditionally used, has two basic purposes: (1) the parsimonious description of multivariate data, and (2) the identification of the structure underlying multivariate data.[26] "Multivariate data" means multiple observations or measurements on the same person or entity being investigated. Let Y be the basic data matrix having N rows and n columns. N is the number of persons or entities being measured and n is the number of dependent variables on which each person is observed. For example, Y might be the summary data matrix for N different products evaluated on n different criteria - say, price, quality, use, and various psychological and physical product variables.

In general the correlations of the n dependent variables are non-zero, indicating the dependent variables are over-lapping and measuring redundant information. The components analysis model expresses the observed data in terms of weighted composites of uncorrelated components. The basic equation for components analysis is,

---

[26] Paul E. Green and Donald S. Tull, Research for Marketing Decisions, (2d ed.; Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1970), pp. 402-413

$Y_{(N,n)} = F_{(N,r)} \quad A'_{(r,n)}$ $\quad r \leq n$, where, F is an (N,r) matrix of uncorrelated component scores.

A' is a (r,n) matrix of weights. In application, r is generally less than n.

The basic equation indicates the assumption that the n observed variables (usually correlated) are weighted composites of a few uncorrelated components (or component scores). The model provides for parsimonious data description in that the components are uncorrelated and there are fewer components than original variables. For simplicity most presentations of components analysis work with both the original data and the components scores in standard score form; i.e., each dependent variable and each component has mean zero and standard deviation one. Given that the data have this scaling, A, the transpose of the matrix of weights is given by,

$$A = QD^{\frac{1}{2}},$$

where Q contains the columnwise eigenvectors of the n X n correlation matrix, R, of the n dependent variables. D is the diagonal matrix of eigenvalues for R, the correlation matrix. F, the matrix of components scores for each person is given by the equation,

$$F = YQD^{-\frac{1}{2}}$$

$$F = YAD^{-1}$$

Unless there are linear dependencies in the basic data matrix Y, the number of components necessary to exactly reproduce Y in the equation $Y = FA'$ is n, indicating that the n-dependent variables are weighted composites of n uncorrelated component scores. In practice $r < n$ components are extracted on the basis that <u>most</u> of the variation in the dependent variables is accounted for by a certain number of components which is less than the number of original variables.

Examples of Principal Components Analysis

Suppose the basic data matrix in standard score form is,

$$Y = \begin{bmatrix} 0.41 & 0.83 \\ -0.95 & -1.08 \\ -0.27 & 0.03 \\ -0.61 & -0.76 \\ 0.41 & 0.51 \\ 0.75 & 0.83 \\ -1.29 & -1.24 \\ 0.41 & -0.29 \\ 1.09 & 1.47 \\ -1.29 & -1.56 \\ 0.75 & 0.35 \\ -1.29 & -1.08 \\ 1.09 & 1.15 \\ 1.43 & 0.83 \\ 1.43 & 0.51 \\ -0.95 & -0.76 \\ -1.29 & -1.08 \\ 0.41 & 1.15 \\ -1.29 & -1.24 \\ 1.09 & 1.47 \end{bmatrix}$$

There are two observations (independent variables) on each of twenty individuals; hence, the data matrix is 20 X 2. The correlation matrix for the two dependent variables is,

$$R= \begin{bmatrix} 1.00 & .92 \\ .92 & 1.00 \end{bmatrix} = Y'Y\ 1/N \ .$$

The matrix, Q, of columnwise eigenvectors of R is,

$$Q= \begin{bmatrix} .71 & .71 \\ .71 & -.71 \end{bmatrix},$$

and finally the diagonal matrix of eigenvalues for R is,

$$D= \begin{bmatrix} 1.92 & .00 \\ .00 & .08 \end{bmatrix} .$$

The matrix of weights for the components which reproduce the original data matrix is,

$$A = \overset{Q}{\begin{bmatrix} .71 & .71 \\ .71 & -.71 \end{bmatrix}} \cdot \overset{D^{\frac{1}{2}}}{\begin{bmatrix} 1.39 & .00 \\ .00 & .28 \end{bmatrix}}$$

$$= \begin{bmatrix} .99 & .20 \\ .99 & -.20 \end{bmatrix}$$

The matrix A is generally referred to as the matrix of factor loadings. Finally the matrix of component scores for each individual is,

$$F = YQD^{-\frac{1}{2}} .$$

$$F = Y \cdot \begin{bmatrix} .71 & .71 \\ .71 & -.71 \end{bmatrix} \cdot \begin{bmatrix} .72 & .00 \\ .00 & 3.57 \end{bmatrix}$$

$$F = Y \cdot \begin{bmatrix} .51 & 2.53 \\ .51 & -2.53 \end{bmatrix}$$

$$F = \begin{bmatrix} 0.63 & -1.06 \\ -1.04 & 0.33 \\ -0.12 & -0.76 \\ -0.70 & .38 \\ 0.47 & -0.25 \\ 0.81 & -0.20 \\ -1.29 & -0.13 \\ 0.06 & 1.77 \\ 1.31 & -0.96 \\ -1.45 & 0.68 \\ 0.56 & 1.01 \\ -1.21 & -0.53 \\ 1.14 & -0.15 \\ 1.15 & 1.52 \\ 0.99 & 2.33 \\ -0.87 & -0.48 \\ -1.21 & -0.53 \\ 0.80 & -1.87 \\ -1.29 & -0.13 \\ 1.31 & -0.96 \end{bmatrix}$$

For the above data, the basic question is, "Are both components necesary, or can Y be adequately summarized with one component?" This question is the hoary "number of components" problem for which factor analysts have no definite answer. The most commonly used procedure for determining the number of important components for a set of data involves the eigenvalues and trace of R, the correlation matrix. Specifically, the procedure involves summing the eigenvalues corresponding to the components one wishes to retain and dividing this sum by the trace of R (written TR (R)). The trace of R is the sum of the main diagonal elements of R. This ratio is interpreted as the variance in Y accounted for by a certain number of components.

In the above example the eigenvalue corresponding to the first component is 1.92, and TR (R) is 2.0. The percent of variance in Y accounted for by the first component is then 1.92/2.0 = 96%; therefore, one would most likely conclude that only one component is necessary to summarize the responses of the twenty people on the two dependent measures. In other words, the one set of component scores contains nearly all the information contained in the two original scores. This is not surprising since the two sets of original scores correlate .92. The above example illustrates the use of components analysis in reducing the complexity of multivariate data; i.e., several measures can be reduced to a small number of component scores which contain nearly all the information in the original scores.

Another important use of components analysis is the identification of the structure underlying multivariate data. The structure of the observed data is provided by the weight matrix, A. If the variables and components are in standard score form, it can be shown that the weights in A are also correlations. Remembering that A is generally an $n \times r$ matrix, A provides the correlations of the original variables and the component scores. Suppose for example, one administers six tests to N people. Further suppose that three of the tests involve verbal reasoning (V_1, V_2, and V_3) and three involve mechanical comprehension (M_1, M_2, and M_3). Assume the weight matrix A is found to be,

$$A = \begin{array}{c} \\ V_1 \\ V_2 \\ V_3 \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{c} \begin{array}{cc} \mathrm{I} & \mathrm{II} \end{array} \\ \begin{bmatrix} .80 & .11 \\ .60 & .16 \\ .50 & .05 \\ .16 & .67 \\ .03 & .86 \\ .01 & .92 \end{bmatrix} \end{array}$$

$(n=6,\ r=2)$

In other words, two components (I and II) are found to be a sufficient solution. A contains the correlations of the original tests with the components. Note from the correlations that component I is primarily defined by the three verbal tests and component II is a mechanical comprehension factor. The six original test scores for people have been reduced to two components; the matrix A provides information about the nature of the two-component structure underlying the six tests.

Principal Components Analysis of Learning Data

The preceding section dealt with the traditional - psychometric usage of components analysis. Applications of components analysis to the study of learning data originated with the work of Tucker[27] and

---

[27]Ledyard R. Tucker, "Determination of Parameters of a Functional Relation by Factor Analysis," Psychometrika, Vol. 23 (1958), pp. 19-23.

has been extended by Weitzman,[28] Ross,[29] and Sheth.[30] These specialized applications to learning data are sufficiently distinct from traditional components analysis as to merit a separate discussion.

The following material will examine the basic components analysis matrices - Y, F, Q, D, and A - when used in the analysis of learning data. Again Y will be the basic N x n raw data matrix; however, in the case of learning data, Y will contain observations for N people on n discrete learning trials.

In some learning studies, as well as the present investigation, the data composing Y is binary - "ones" indicating correct responses and "zeros" incorrect responses. In this study a "one" indicates purchase of Brand A on a given trial (purchase opportunity) and a "zero" references the purchase of $\bar{A}$, some other brand. On the other hand, there are many learning studies where the elements of Y are continuous variables - for example, the number of items recalled during a learning trial.

The usual summary data for a binary learning matrix Y consists of summing each of the n columns of Y and dividing these sums by N, the

---

[28]R. A. Weitzman, "A Factor Analytic Method for Investigating Differences Between Groups of Individual Learning Curves," Psychometrika, Vol. 28, (1963), pp. 69-80).

[29]John Ross, "Mean Performance and the Factor Analysis of Learning Data," Psychometrika, Vol. 29 (1964), pp. 67-73.

[30]Jagdish N. Sheth, "A Behavioral and Quantitative Investigation of Brand Loyalty" (unpublished doctoral dissertation, University of Pittsburgh), 1966; Jagdish N. Sheth, "A Factor Analytical Model of Brand Loyalty," Journal of Marketing Research, (November, 1968), pp. 395-404.

number of people. These ratios are interpreted as the probability of a correct response on a given trial, and when plotted across trials generally yield the familiar negatively accelerated exponential learning curve. A question which has been raised in the psychological literature is, "How well does the average or mean learning curve fit the individual learning data?"[31]

Tucker[32] was evidently the first to suggest the use of components analysis for the study of individual differences in learning curves. Tucker's basic approach, which was adopted by Sheth, is now presented. Consider the basic equation of component analysis:

$$Y_{(N,n)} = F_{(N,r)}\, A'_{(r,n)} \qquad r \leq n.$$

In this traditional approach this equation specifies that the observed scores in Y are a weighted composite of the component scores, F. A' is the matrix of weights. In Tucker's and Sheth's usage of components analysis, the basic matrices are conceptualized differently; the matrix A' is considered to be a series of r reference learning curves and F, the component scores, are weights. The statement of the model for learning data is as follows. The series of scores in Y for individuals are weighted sums of reference learning curves. Components analysis is used, then, to answer questions as to the number and forms of reference curves which are needed to account for the observed learning data. If

[31]M. Merrill, "The Relationship of Individual Growth to Average Growth," Human Biology, Vol. 3 (1931), pp. 37-40; C. L. Hayes, "The Backward Curve; A Method for the Study of Learning," Psychological Review, Vol. 60 (1953), pp. 269-275.

[32]Tucker, loc. cit.

there is one dominant component, all individual learning curves are mulitples of one another and have the same shape. If there are a few dominant principal components, then only a corresponding number of reference learning curves are needed to account for the individual learning data.[33]

In reference to factor scores (BLS's), their prime use is that they provide an index of degree of brand loyalty for each buyer. If there is one dominant component, then the functional relation between brand loyalty and purchase trials dictates only one BLS for each buyer. If there is more than one dominant component, signaling that groups of individual learning curves have different form and shape, then there is an equivalent number of BLS's for each buyer. A given buyer based upon the respective magnitudes of her BLS's, could belong to one dominant component or any combination of dominant components.

---

[33]Ledyard R. Tucker, "Learning Theory and Multivariate Experiment: Illustration by Determination of Generalized Learning Curves," in R. B. Cattell, ed., Handbook of Multivariate Experimental Psychology, Rand McNally, (1966), pp. 476-501.

CHAPTER III

METHOD

Three aspects of this experimental investigation of the merits of factor analysis as a measure of brand loyalty are provided in this section. First detailed, is the use of the simulated learning curves in the measurement of the origins of brand loyalty. Contextually, five different hypothetical market situations are created and reviewed. The spectrum of curves reflecting these market situations, in effect, provides the test for the factor analytic technique. Second, the method of components analysis applied to the generated learning data is provided. This presentation should be considered in conjunction with the just explicated background material on principal components analysis. Third, a concise summary is outlined of the computational steps accomplished in this research.

Use of Simulated Learning Curves in a Brand Loyalty Experiment

Five matched pairs of aggregate brand loyalty curves, for a total of ten individual runs, were simulated. Each set is specified by the appropriate parameters, respectively, of the two learning models. Effectively, then, there are but five different aggregate curves.

The two learning models in Table 3 are shown with the parameters used to generate the curves in Fig. 2 through Fig. 6. The proportion of buyers (N = 300) purchasing brand A on each trial is also the probability of purchasing brand A on that trial. Thus, by summing the ones and calculating the proportion on each trial the brand loyalty curves easily are plotted. Observe in Fig. 2 through Fig. 6 that the nth trial difference between the matched curves within a set are so small as to be attributable to chance differences. The consequence is shown: the two curves in each set are virtually "identical," but the underlying psychological processes on which they are predicated are conceptually different.

In two sets (Fig. 2 and Fig. 3), in order to match an aggregate learning curve it was necessary to use extreme parameters. Parameters were used that are not usually associated with findings in more conventional learning experiments. Coombs, Dawes, and Tversky[1] indicate that, in general, learning rates are less than .20; for purposes of this study such extreme learning rates as B = .98, B1 = .01, and B = .80, B1 = .01 are used. These magnitudes may distort reality to some degree but concurrently provide a stringent test for the factor analytic outcome; these will be explained later in more detail.

Since there is no strict requirement for using matched pairs, brand loyalty curves could be generated simply from either learning model and in turn factor analyzed. Such a procedure allows use of more

---

[1]Clyde H. Coombs, Robyn M. Dawes, and Amos Tversky, Mathematical Psychology, (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970), p. 264.

TABLE 3

LEARNING PARAMETERS FOR THE FIVE MATCHED SETS
OF SIMULATED LEARNING CURVES

	Incremental	All-Or-None	Pr(buy A| row state)
	$Gp = (1-B)p_n + BA$ $Lp = (1-B1)p_n + B1(A1)$	$\begin{array}{cc} & \begin{array}{cc} L & U \end{array} \\ \begin{array}{c} L \\ U \end{array} & \begin{bmatrix} 1.00 & 0.00 \\ c & 1-c \end{bmatrix} \end{array}$	$\begin{bmatrix} 1.00 \\ g \end{bmatrix}$
Set 1	$Gp = (1-.98).20 + .98$ $Lp = (1-.01).20$	$\begin{array}{cc} & \begin{array}{cc} L & U \end{array} \\ \begin{array}{c} L \\ U \end{array} & \begin{bmatrix} 1.00 & .00 \\ .20 & .80 \end{bmatrix} \end{array}$	$\begin{bmatrix} 1.00 \\ .20 \end{bmatrix}$
Set 2	$Gp = (1-.80).10 + .80(.90)$ $Lp = (1-.01).10$	$\begin{array}{cc} & \begin{array}{cc} L & U \end{array} \\ \begin{array}{c} L \\ U \end{array} & \begin{bmatrix} 1.00 & .00 \\ .10 & .90 \end{bmatrix} \end{array}$	$\begin{bmatrix} .90 \\ .10 \end{bmatrix}$
Set 3	$Gp = (1-.10).20 + .10$ $Lp = (1-.05).20$	$\begin{array}{cc} & \begin{array}{cc} L & U \end{array} \\ \begin{array}{c} L \\ U \end{array} & \begin{bmatrix} 1.00 & .00 \\ .01 & .99 \end{bmatrix} \end{array}$	$\begin{bmatrix} 1.00 \\ .20 \end{bmatrix}$
Set 4	$Gp = (1-.10).40 + .10$ $Lp = (1-.10).40$	$\begin{array}{cc} & \begin{array}{cc} L & U \end{array} \\ \begin{array}{c} L \\ U \end{array} & \begin{bmatrix} 1.00 & .00 \\ .60 & .40 \end{bmatrix} \end{array}$	$\begin{bmatrix} .4 \\ .4 \end{bmatrix}$
Set 5	$Gp = (1-.12).60 + .12$ $Lp = (1-.22).60$	$\begin{array}{cc} & \begin{array}{cc} U & L \end{array} \\ \begin{array}{c} U \\ L \end{array} & \begin{bmatrix} 1.00 & .00 \\ .05 & .95 \end{bmatrix} \end{array}$	$\begin{bmatrix} .01 \\ .60 \end{bmatrix}$

"legitimate" learning parameters to produce curves which are the same in form but different in shape.

Brand Loyalty Interpretations of Learning Curves

Presented in this section is a description of possible market place circumstances and a corollary discussion of the implications of the selected learning parameters. For convenience, it may be assumed in all cases: (1) that the initial probability of purchasing brand A represents the current market share and, (2) that each purchase trial represents one week.

Rapid and Complete Learning: Set Number One

Of the five learning curves generated, the rapid and complete learning depicted in set 1, Fig. 2, is probably the most difficult to justify in the real-world. It is chosen, however, as an extreme case primarily for the purpose of examining the resultant factor analytic outcome.

A Conceived Market Situation: There is a very rapid increase in the probability of purchasing brand A: the mean curve shows a radically high rate of learning. By the 16th trial, for all practical purposes, both complete learning and 100% purchase of brand A has occurred. Rarely would it be expected that a sample with a 20% market share, at time 0, just 16 weeks later represents 100% brand loyalty.

The most loyal buyer or sample of buyers occasionally will not buy brand A because of price, availability, time, or any combination of reasons even though complete learning of brand A has occurred. Thus preventing this probability of purchase curve as an expected outcome. A situation where such a curve might emerge is when brand A represents a totally new product for which there is no close substitute. Examples might be the introduction of the Wilkinson stainless steel razor blade, or a new Polaroid film, for a given market segment.

Learning Parameters: The two sets of parameters needed to generate the curve are unusual. As just alluded to, both learning and the probability of purchasing brand A is near one at about the 16th trial.

To achieve this condition, the all-or-none model must be absorbing (probability of staying learned = 1) and the probability of buying once learned must be one; otherwise the aggregate curve could never reach an asymptote of one. Clearly then, with two probabilities fixed, the shape of the curve is further manipulated only by the combination of the starting probability of purchase, .20, in the unlearned state, and the probability of learning (switching from U to L) .20.

The incremental learning parameters needed to reproduce the curve are so extreme that the model becomes mathematically almost equivalent to the all-or-none model. The very high rate of learning, B = .98, in the gain operator and negligible rate of learning, B1 = .01, in the loss operator should provide individual protocols approaching those of all-or-none learners.

For support of this notion observe what happens when B = 1.0, B = 0.0, and p_n = .20.

$$Gp = (1-B)p_n + B$$

$$Lp = (1-B1)p_n$$

$$Gp = (1-1).2 + 1 = 1.0$$

$$Lp = (1-0).2 \quad = 0.20$$

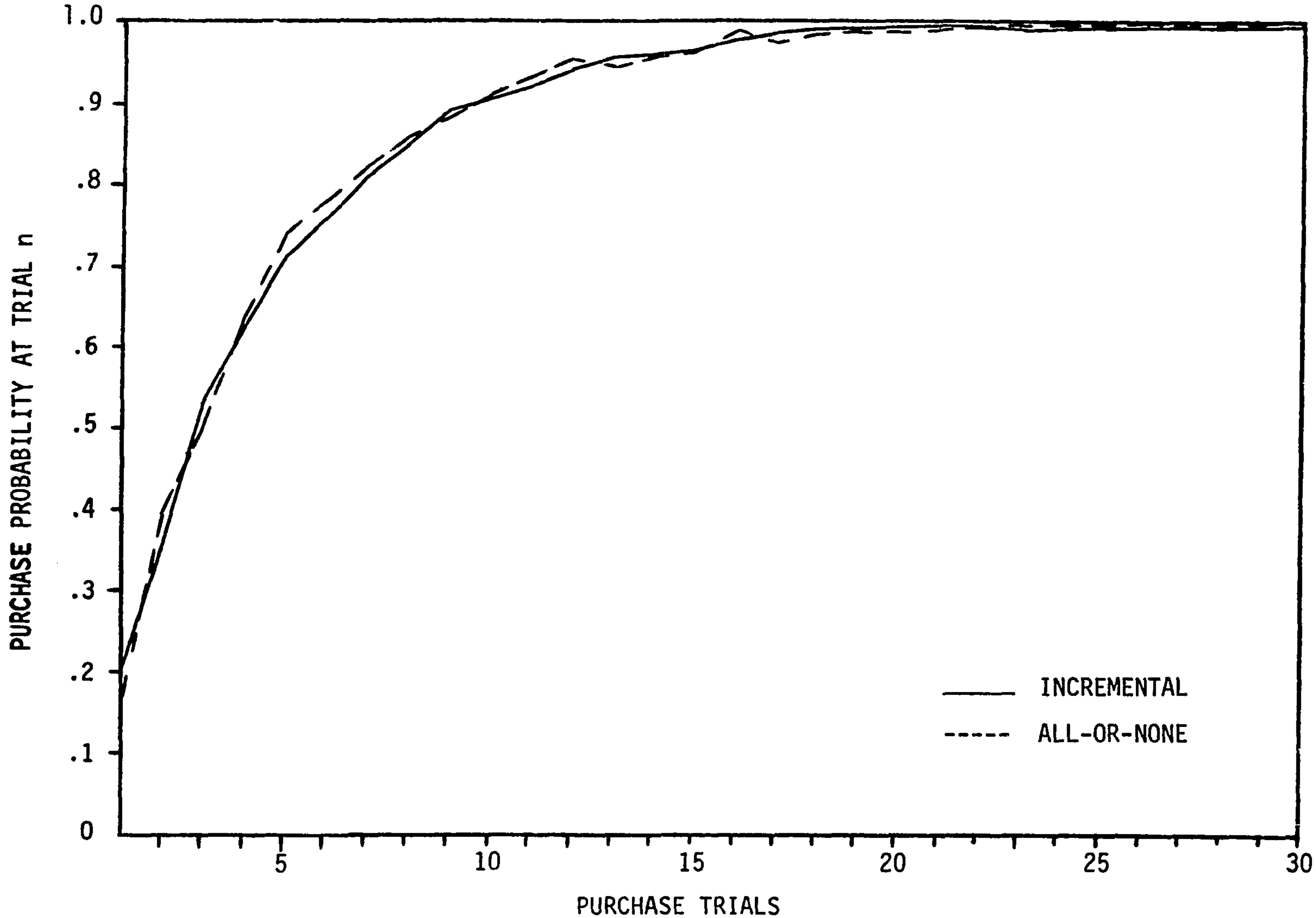

Fig. 2.--Simulated mean learning curves for set number 1.

situation where such a curve might emerge is when brand A represents a totally new product for which there is no close substitute. Examples might be the introduction of the Wilkinson stainless steel razor blade, or a new Polaroid film, for a given market segment.

Learning Parameters: The two sets of parameters needed to generate the curve are unusual. As just alluded to, both learning and the probability of purchasing brand A is near one at about the 16th trial.

To achieve this condition, the all-or-none model must be absorbing (probability of staying learned = 1) and the probability of buying once learned must be one; otherwise the aggregate curve could never reach an asymptote of one. Clearly then, with two probabilities fixed, the shape of the curve is further manipulated only by the combination of the starting probability of purchase, .20, in the unlearned state, and the probability of learning (switching from U to L) .20.

The incremental learning parameters needed to reproduce the curve are so extreme that the model becomes mathematically almost equivalent to the all-or-none model. The very high rate of learning, B = .98, in the gain operator and negligible rate of learning, B1 = .01, in the loss operator should provide individual protocols approaching those of all-or-none learners.

For support of this notion observe what happens when B = 1.0, B = 0.0, and p_n = .20.

$$Gp = (1-B)p_n + B$$

$$Lp = (1-B1)p_n$$

$$Gp = (1-1).2 + 1 = 1.0$$

$$Lp = (1-0).2 \quad = 0.20$$

Repeated trials produce stationary probabilities of repurchase. If the first purchase is brand A, all subsequent purchases will be the same since in one trial complete learning has occurred. If the first purchase is not-brand A, the probability of buying brand A stays at .20 until that nth trial when brand A is purchased. Then, of course, the purchase probability remains at one.

This special case is elaborated because, of the five sets presented for measurement, this one, and to a lesser extent set 2, provide the most rigorous test for Sheth's factor analytic technique. The implication is that the raw data generated from two conceptually different frameworks are so similar that the ability of the factor analytic technique to differentially identify type of learning is suspect.

Rapid and Incomplete Learning: Set Number Two

Observe in Fig. 3 that set 2, in appearance, is quite similar to the curve in set 1. However, changes in specified parameters alter significantly the interpretations.

A Conceived Market Situation: Again, the increase in the probability of purchasing brand A is unusually quick. The increase in market share, now beginning at 10%, through the life of the market period is still greater than "normally" expected. The situation once more may be characterized by brand A representing a new and highly differentiated product. An illustration might be the first firm out with a suitable "sweetner" after the cyclamate scare. In general, the chances up to a point of capturing 100% of the market look good. Notice that around trial 20 the increase in probability of purchase of brand A "tails off"

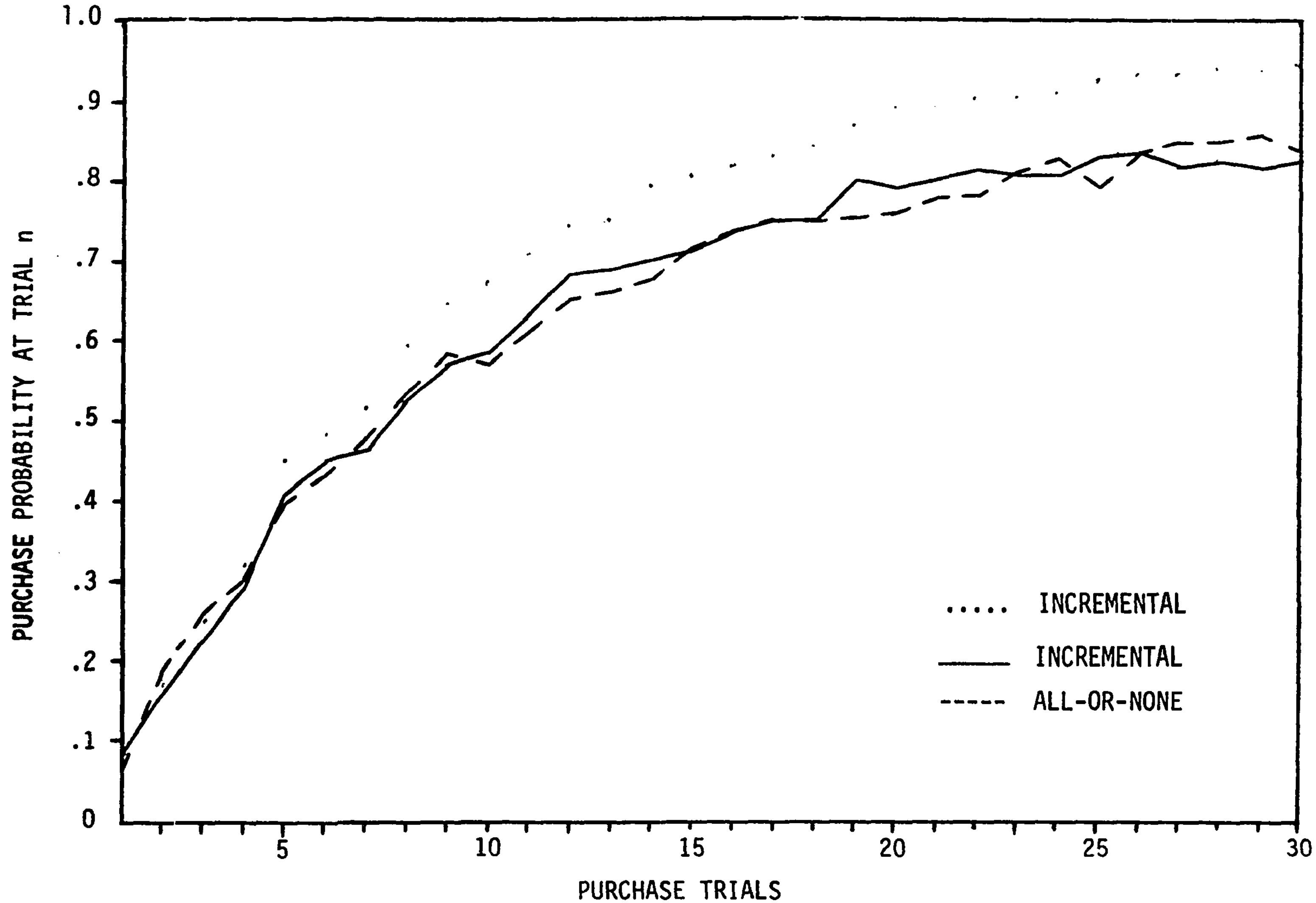

Fig. 3.--Simulated mean learning curves for set number 2.

and never reaches one. This issue is discussed below.

Learning Parameters: In a sense the term "incomplete learning" is a misnomer at least as related to the all-or-none conception. The consumer still learns completely on one nth trial, however, now her probability of purchasing brand A in the learned state is .9. Alternatively, for easily imagined reasons--out-of-stock-- 10% of the time she purchases not-brand A. The impact is that, in the aggregate, the curve at some nth plus 30th trial can reflect a maximum purchase probability of .90.

The effect of the learning rates, B = .80 and B1 = .01, in the incremental model is substantially the same as in set 1. For example, after three consecutive brand A purchases the p value is .8936, which is very close to .90 the maximum amount of learning allowed in this case. Obviously, this is not one-trial-learning but with the combination of learning rates used the outcome is close to one-trial-learning.

The point to recognize in set 2 is that the asymptotic limit in the Gp is set at .9 instead of 1.0. Therefore, the individual consumer never can completely learn to buy brand A. In the aggregate then, the .9 probability of purchase eventually will be reached, say in 10 to 20 more trials, and stay at that level as long as the current market conditions prevail. The plotted data points above the curve show the effect when eventual complete learning is allowed. That is, if the asymptotic value is changed from .9 to 1.0 in the incremental model.

Moderate Learning: Set Number Three

The main features in Fig. 4 are the absence of the traditional

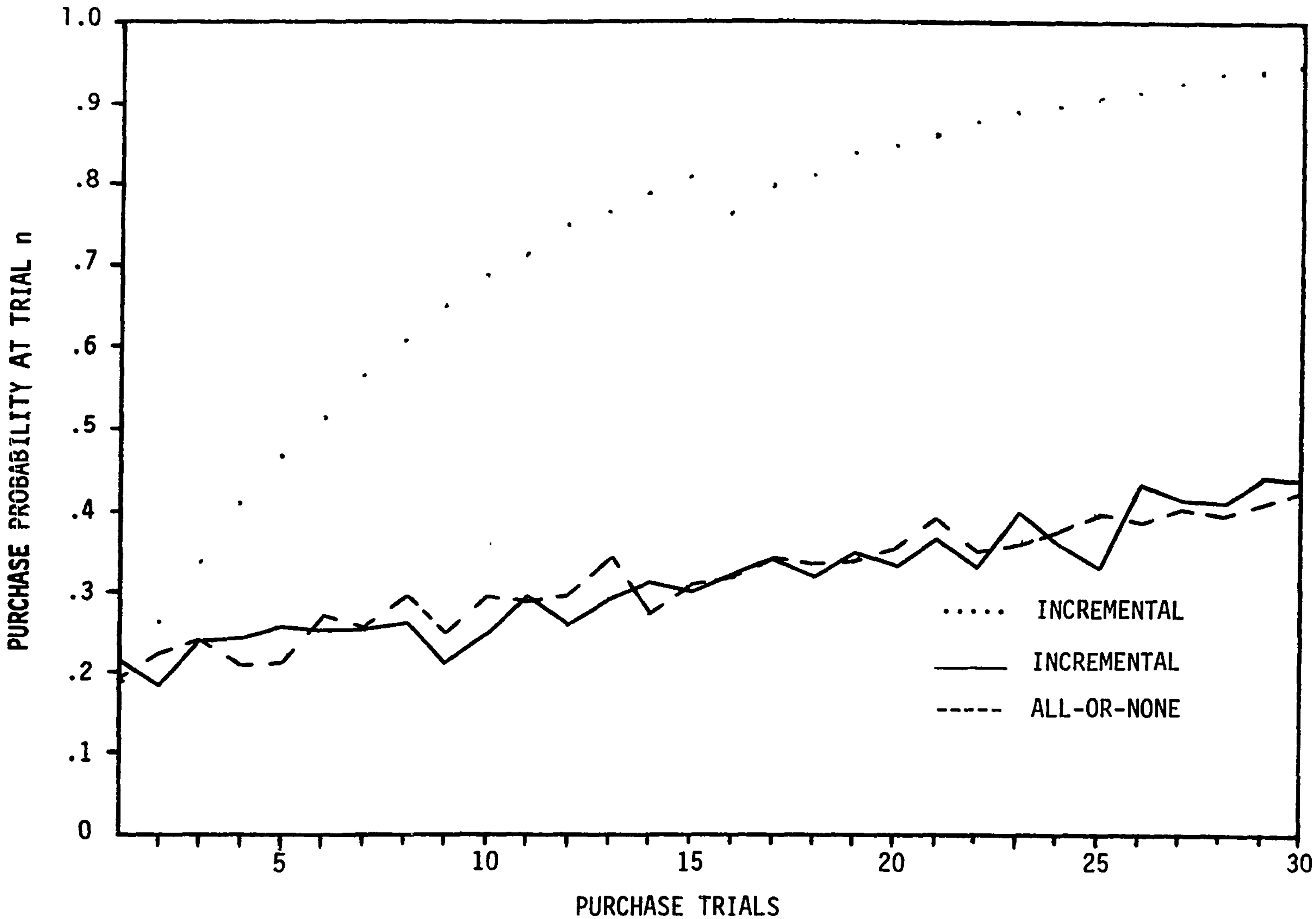

Fig. 4.--Simulated mean learning curves for set number 3.

negatively accelerated group learning curve, and more gradual, reasonable, increases in the probability of purchase across trials.

A Conceived Market Situation: Going from 20% to approximately 40% market share in 30 weeks is still a "healthy" increase, but surely more reasonable than the situations presented in sets 1 and 2. The consumers are displaying much more variability in their buying patterns. The reason may be that a host of market factors are in flux, or that the brand A product is in a somewhat homogeneous product group negating the possibility of dramatic early learning.

Learning Parameters: These parameters are fixed such that mathematically the steady rise in the curve will continue for many more purchase trials until the upper asymptote of one is reached.

The reason that the "linear" function is observed is due to the probabilistic effects of the combination of selected parameters. The all-or-none model compared with set 1 is identical except for one sharp variation. The purchase probability in the unlearned state, the probability of staying learned, and the purchase probability once learned are all the same. The effect of changing the probability of learning to .01 from .20, however, is profound. It is known that stochastically the consumer buys 20% of the time when in state U, but the probability is only 1 in 100 trials that she will learn (switch from U to L) to buy brand A 100% of the time. The small increase in the purchase probability from trial to trial shown in the mean curve is the consequence of this low probability of ever learning to purchase brand A. The attempts of the "teeth whiteners" to capture some of the Colgate, Crest, Gleem market share may be a fitting example here.

The same type of thing is occurring in the incremental model but takes on a different form. The Gp learning rates decreased from .98 and .80, respectively, to .10 while at the same time the Lp learning rate increased from .01 to .05. Thus, the associated reinforcements are nearly uniform yet maintaining a net positive effect. Translating into marketing terms, the assumption is that the stimulus (brand A) is not well defined (or perceived) in the market place causing the much trial-to-trial variability in purchase response. Consequently, increase in purchase probability by trial is slight. When a consumer buys on almost every trial the conventional negatively accelerated curve, shown above the mean curves, is clearly evident.

Learning as a Constant: Set Number Four

Clearly in Fig. 5 the purchase probability is the "same" on trial one as on trial 30. In other words, in total, the probability of buying brand A is a constant function of purchase trials. The degree of acquisition is easily observed, but how that level of brand loyalty came about is not evidenced.

A Conceived Market Situation: Logically, the market conditions are, or have been stable throughout the period examined. There has been no special promotion, nor has there been any product innovation on brand A. Brand A is well established at a fixed market share. In a local market, such common items as bread, milk, and coffee would fit this description. The consumers had to have had developed collectively brand loyalty up to its current level prior to the existing market situation. Remember that, given a constant 40% market share, the variability in

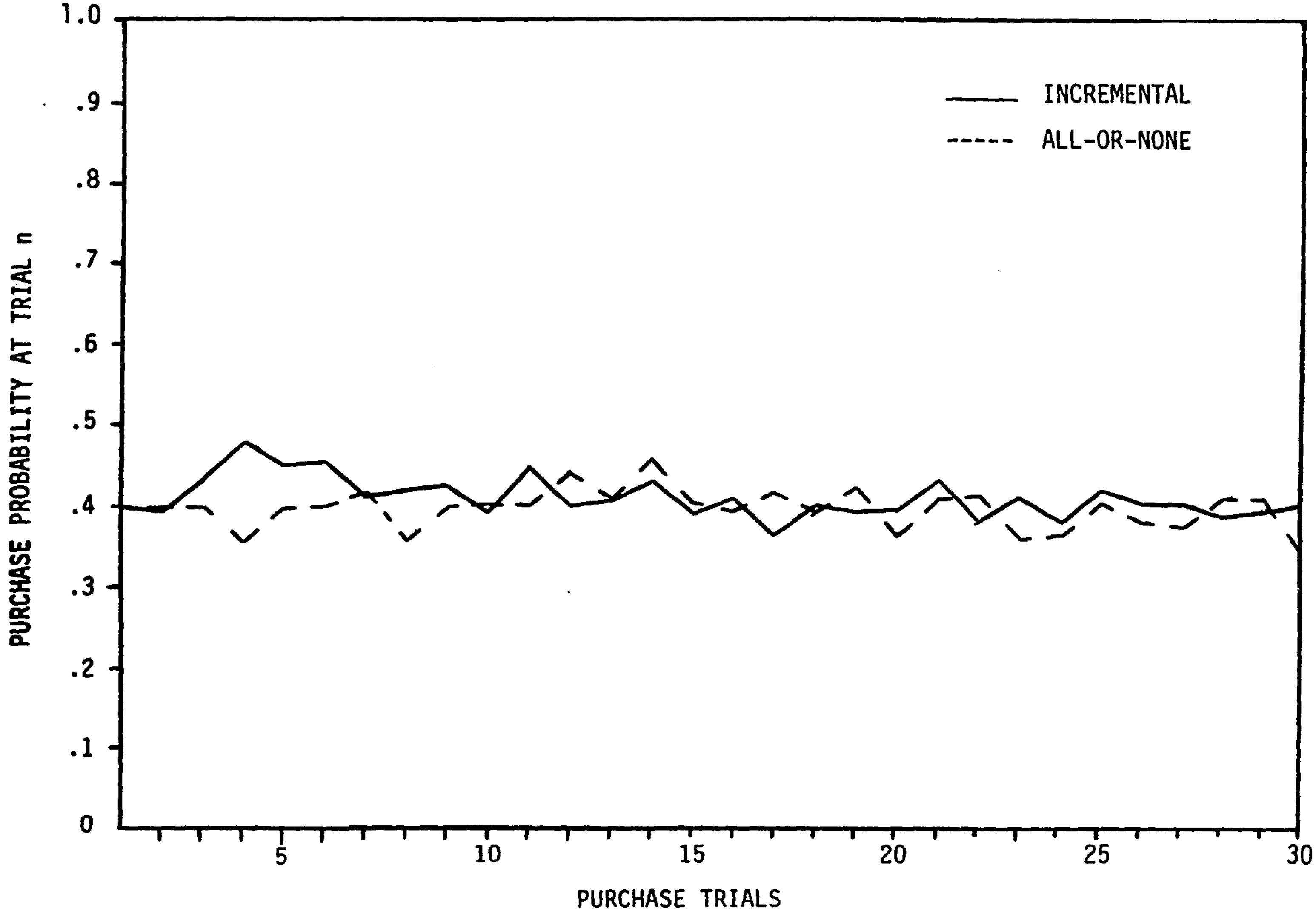

Fig. 5.--Simulated mean learning curves for set number 4.

purchase response within and across consumers may be considerable.

Learning Parameters: As long as the transition matrix is absorbing in the L state, the buyers stay learned. The probability of purchase in state L is adjustable, however, as was done in set 2. By equating the purchase probabilities for both the U and the L state the straight-line function was obtained. A real world interpretation may be difficult since now the probability of purchasing brand A is the same regardless of whether learning has occurred. It could be that price, availability, etc. counterbalance the effect of learning.

Another explanation more in line with theory (depending upon how one views permanency of learning) is that the transition matrix is not absorbing. If given an ergodic Markov chain (defined above), then the consumer, with a stated probability, can unlearn. The implications of this might be best pointed out by an example which hopefully will suffice for a real-world explanation of the constant purchase probability, yet eliminate the equal purchase probability whether in the U or L state.

In general, and without the use of specified probabilities, when a consumer unlearns the association between brand A and purchase of brand A at some fixed probability (switching from L to U), the overall effect is that now she is allowed to buy brand A a greater proportion of the time while in the learned state. That is, her brand A purchase pattern is more consistent with learned behavior. An additional effect is that the consumer is now in the U state more often than before allowing her purchase probability in the U state to remain the same or decrease.

The final outcome is more tenable because the purchase probabilities in the respective states are not equal and are intuitively consistent with human behavior. The reason that actual probabilities were not cited in the example is that the various combinations of parameters needed to maintain a constant function are not easily determined. Several Monte Carlo runs may be necessary before true constancy is achieved.

The incremental model parameters required for maintaining learning at a constant level are readily explained. When using the two-operator linear model with rates of learning in the Gp and Lp equal, the level of the average learning always remains precisely at the initial brand A purchase probability (if and only if the upper and lower bounds are 1 and 0 respectively).

In application, the consumer by her purchase behavior is positively or negatively reinforced in the same magnitude. The effects on the p_n then are equal but in opposite directions maintaining a constant probability of purchase.

Unlearning: Set Number Five

The main reason for inclusion of the negative learning in Fig. 6 is to provide a broad range of curves for study. Actually the curve is quite similar to set 3 but in the opposite direction. At first, the extinction process may be confusing because now practically everything talked about must be conceptualized in reverse.

<u>A conceived Market Situation</u>: This curve might describe several circumstances. The declining stage in a product life cycle; relatedly,

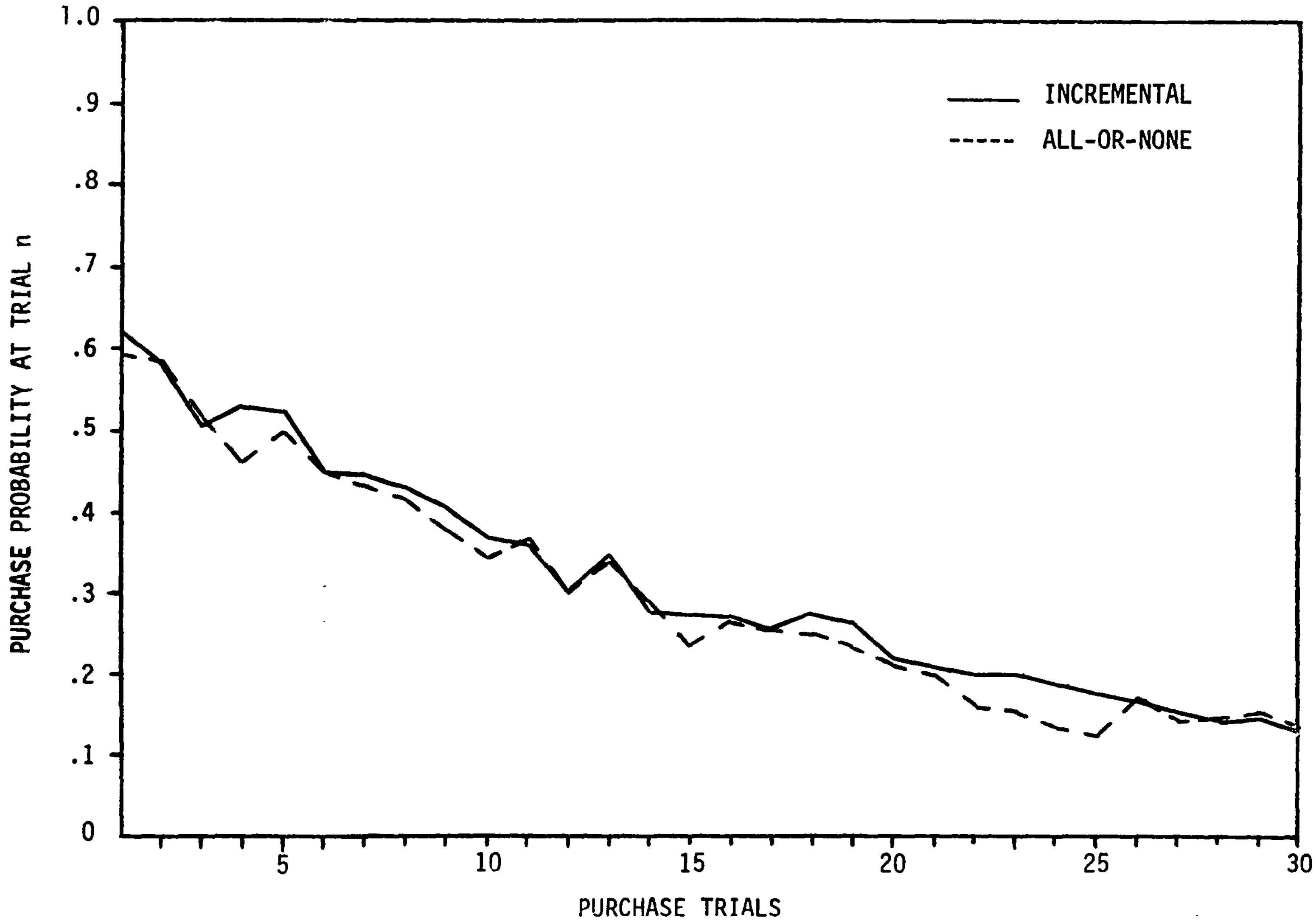

Fig. 6.--Simulated mean learning curves for set number 5.

obsolescence; a strong promotion campaign from other brands; or the dissipating effects of an initial promotion introducing brand A, all of these could be relevant. At the beginning of the time period the probability is .60 that brand A will be bought and gradually declines to about .15. Over many more purchase attempts the probability of purchase will decline to less than 1% because the gradual "unlearners" eventually will never buy brand A while the all-or-none "unlearners" though completely unlearned will still buy 1% of the time in state U.

<u>Learning Parameters</u>: The row states are reversed in the all-or-none model in order to reflect the situation. The assumption is made that the consumer starts out in the learned state with a purchase probability of .60. She has only a 5% chance of switching from L to U, but once she is unlearned that condition prevails and from then on she buys brand A only 1% of the time.

The only thing new in the incremental model is that for the first time the rate of learning in the Gp is less than the rate of learning in the Lp, i.e., $B < B1$. The net effect across trials is that the consumer derives relatively more satisfaction from other brands than from buying brand A. The same outcome could be described in terms of dissatisfaction.

Certainly there are many combinations of parameters in either model that are not examined here. The purpose in presenting these five sets was for reasonably representative exposure. Especially with the all-or-none model, when the absorbing state constraint is ignored, significantly greater flexibility is obtained. However, the decision was made to keep the theoretical and mechanical implications as simple as is practical. The computer program (Appendix B) written for the two-state all-or-none model is a general program and will handle all probability combinations in the transition matrix.

Explanation of Computational Procedure Involved in the Factor Analytic Model

The algebra of components analysis as applied to learning data is slightly different from the traditional approach in which standardized variables are used (see Principal Components Analysis in Background --Chapter Two). Let Y be the N x n basic data matrix containing the learning scores for N people observed across n trials. Let y_{ij} be the purchase score for the ith person on the jth trial or jth purchase opportunity:

$$y_{ij} = 1 \text{ if brand A purchased,}$$

$$y_{ij} = 0 \text{ if brand } \bar{A} \text{ purchased.}$$

Note that Y is left in raw score form. $Y'Y$ is an n x n matrix of trial sums of squares and sums of trial cross products. $Y'Y$ is symmetric with the sums of squares for trials on the diagonal and sums of cross products on the off-diagonal. Sheth, instead of working with the cross-products matrix, rescales the matrix into a matrix designated C.

$$C = D^{-\frac{1}{2}}_{Y'Y} Y' Y D^{-\frac{1}{2}}_{Y'Y} ,$$

where $D^{-\frac{1}{2}}_{Y'Y}$ is a diagonal matrix containing the reciprocals of the square roots of the diagonal elements of $Y'Y$. C, the normalized cross products matrix, has a unit diagonal and off-diagonals ranging between zero and one. Since C is symmetric, a powerful theorem of matrix algebra states that:

$$C(n,n) = Q(n,n)D(n,n)Q'(n,n),$$

where Q is a matrix of columnwise eigenvectors of C, and D is a diagonal matrix of eigenvalues for C. Once the eigenvalues and eigenvectors are determined, the basic matrices F and A for principal components can be determined. As was the case for traditional components analysis:

$$F = YQD^{-\frac{1}{2}}$$

$$F = YAD^{-1}$$

$$\text{and } A = QD^{\frac{1}{2}}.$$

However, in the present data analyses the component scores F were scaled so that their sums of squares were equal to one: i.e., so that $F'F = I$, where I is an identity matrix. The following equation:

$$F = YD_{Y'Y}^{-\frac{1}{2}}QD^{-\frac{1}{2}}$$

$$F = YD_{Y'Y}^{-\frac{1}{2}}AD^{-1},$$

yields component scores whose sums of squares are all equal to one. The basic components model can then be constructed from these matrices A and F.

In general the same number of components as there are trials is necessary to exactly reproduce the original data from $Y = FA'$. However, if only the important or dominant components are retained, one may write:

$$\hat{Y}_{(N,n)} = F_{(N,r)}A'_{(r,n)} = FA'_{(r)} \qquad r < n,$$

and $F'A_{(r)}$ will be the best least squares estimate of Y.[2] $\hat{Y} = FA'_{(r)}$ is a least squares estimate of Y based on r principal components, where r is less than n, the number of trials.

Once again, following Sheth's work, the dominant principal components were selected by computing the percent of variability accounted for by each component. The percent of variability accounted for by the pth principal component (% pth component) is,

$$\text{\% pth component} = \frac{\lambda_p}{TR(C)} ,$$

where λ_p is the eigenvalue associated with the pth principal component, and TR (C) is the trace of the normalized cross products matrix. In this research all C matrices were 30 x 30 and had unit diagonals; therefore, TR (C) is 30 for each set of data. Suppose, for example, the eigenvalue associated with the first component was 24. Then,

$$\text{\% 1st component} = \frac{\lambda_1}{TR(C)} = \frac{24}{30} = 80\%.$$

This result would be strongly indicative of a one-component solution, since 80 percent of the variation in Y is accounted for by the first principal component of the data.

Once the number of dominant principal components is determined for a set of learning data, the elements of A, the factor loadings, can be plotted across trials to produce the shape of the dominant reference learning curves underlying individual performance. If only one dominant factor is found, the shape of the reference curve (plotted factor

---

[2]Ledyard R. Tucker, "Learning Theory and Multivariate Experiment: Illustration by Determination of Generalized Learning Curves," in R. B. Cattell, ed., Handbook of Multivariate Experimental Psychology, Rand McNally, (1966), pp. 476-501.

loadings) should be nearly identical to the mean learning curve computed on the original data.

Summary of the Computational Steps Involved in This Research

1. Y, the purchase behavior matrix (learning data) was generated from the two mathematical learning models by the use of Monte Carlo techniques. All matrices Y were of dimension 300 x 30. The computer programs written for this purpose are found in Appendix A and B.

2. $Y'Y$ was computed and normalized into $C = D_{Y'Y}^{-1/2} Y'YD_{Y'Y}^{-1/2}$ All matrices C were of dimension 30 x 30.

3. C was resolved into eigenvectors and eigenvalues through use of a standard computer subroutine.

4. Using the matrix Q of eigenvectors and D, the diagonal matrix of eigenvalues, the first 13 sets of factor loadings and factor scores were computed, respectively;

$$A = QD^{1/2}$$

$$F = YD_{Y'Y}^{-1/2} AD^{-1} .$$

The computer program used to accomplish steps 2, 3, and 4 is in Appendix C.

5. The eigenvalues corresponding to each principal component were used to determine the number of dominant reference learning curves underlying each set of observed data.

CHAPTER IV

RESULTS

With all the necessary methodological conditions described, the summarized results now are presented. In Table 4 through Table 13 are presented outcomes resulting from use of Sheth's normalized cross-products matrix as input for standard principal components analysis. The three columns of data in each Table represent the first three reference curves (factors) computed. The elements, in this case representing the 30 purchase trials, comprising each reference curve are called factor loadings. When the factor loadings are plotted, the resulting curves identify the independent dimensions which underlie the learning data. If only one dominant reference curve is found, its shape should be nearly identical to the mean learning curve computed on the original data. Notice that in all cases the first reference curve is composed of all postitive values. The remaining reference curves, usually have negative values since they act as correction terms for the first reference curve.

The percent of variation explained and the eigenvalue for each reference curye are presented in the last two rows of each Table.

The first reference curve for each of the five sets were plotted and are shown in Fig. 7 through Fig. 11. The reference curves are shown directly with their corresponding aggregate learning curves for ease of comparison and interpretation.

Since there is neither a real benefit nor an efficient way of presenting the 10 groups of 300 brand loyalty scores (factor scores), only selected BLS's and purchase protocols are shown for analysis and evaluation at the time BLS's are discussed.

TABLE 4

REFERENCE CURVES (FACTOR LOADINGS) FOR INCREMENTAL LEARNING DATA FROM SET 1
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.50775	-.64866	.48858
2	.67292	-.62133	.15806
3	.79815	-.48647	-.13024
4	.85776	-.38145	-.22513
5	.90370	-.26960	-.25093
6	.92542	-.20357	-.23741
7	.94737	-.12509	-.19718
8	.96067	-.07008	-.15573
9	.97143	-.02083	-.11116
10	.97789	.01312	-.07521
11	.98184	.03667	-.04782
12	.98436	.05471	-.02513
13	.98704	.07655	.00374
14	.98799	.08493	.01487
15	.98828	.08886	.02041
16	.98895	.10364	.04186
17	.98910	.11392	.05705
18	.98903	.11707	.06176
19	.98903	.11707	.06176
20	.98903	.11707	.06176
21	.98860	.11947	.06540
22	.98860	.11947	.06540
23	.98860	.11947	.06540
24	.98860	.11947	.06540
25	.98768	.12083	.06758
26	.98768	.12083	.06758
27	.98768	.12083	.06758
28	.98768	.12083	.06758
29	.98768	.12083	.06757
30	.98768	.12083	.06758
% of Variation Explained	89.97	5.20	1.99
Eigenvalues	26.9907	1.5591	0.5963

TABLE 5

REFERENCE CURVES (FACTOR LOADINGS) FOR ALL-OR-NONE
LEARNING DATA FROM SET 1
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.41111	-.51271	-.73996
2	.66224	.42000	-.38363
3	.75128	.42564	-.27467
4	.83625	.36835	-.08832
5	.89448	.21359	-.04677
6	.91657	.17806	-.04767
7	.93960	.13160	-.00438
8	.95203	.06861	.03787
9	.96208	.03517	.00822
10	.97444	.00447	.04995
11	.98098	-.00230	.04096
12	.98593	-.03510	.03335
13	.98247	-.01540	.05176
14	.98750	-.04221	.06217
15	.98898	-.05062	.04817
16	.99065	-.06791	.03893
17	.99019	-.05224	.05885
18	.99117	-.07268	.04140
19	.99147	-.07301	.04227
20	.99214	-.07553	.04433
21	.99147	-.07301	.04227
22	.99193	-.08027	.03693
23	.99177	-.08194	.03862
24	.99177	-.08194	.03862
25	.99177	-.08194	.03862
26	.99112	-.07942	.03657
27	.99177	-.08194	.03862
28	.99177	-.08194	.03862
29	.99112	-.07942	.03657
30	.99112	-.07942	.03657
% Of Variation Explained	89.242	3.17	2.74
Eigenvalues	26.7726	0.9505	0.8232

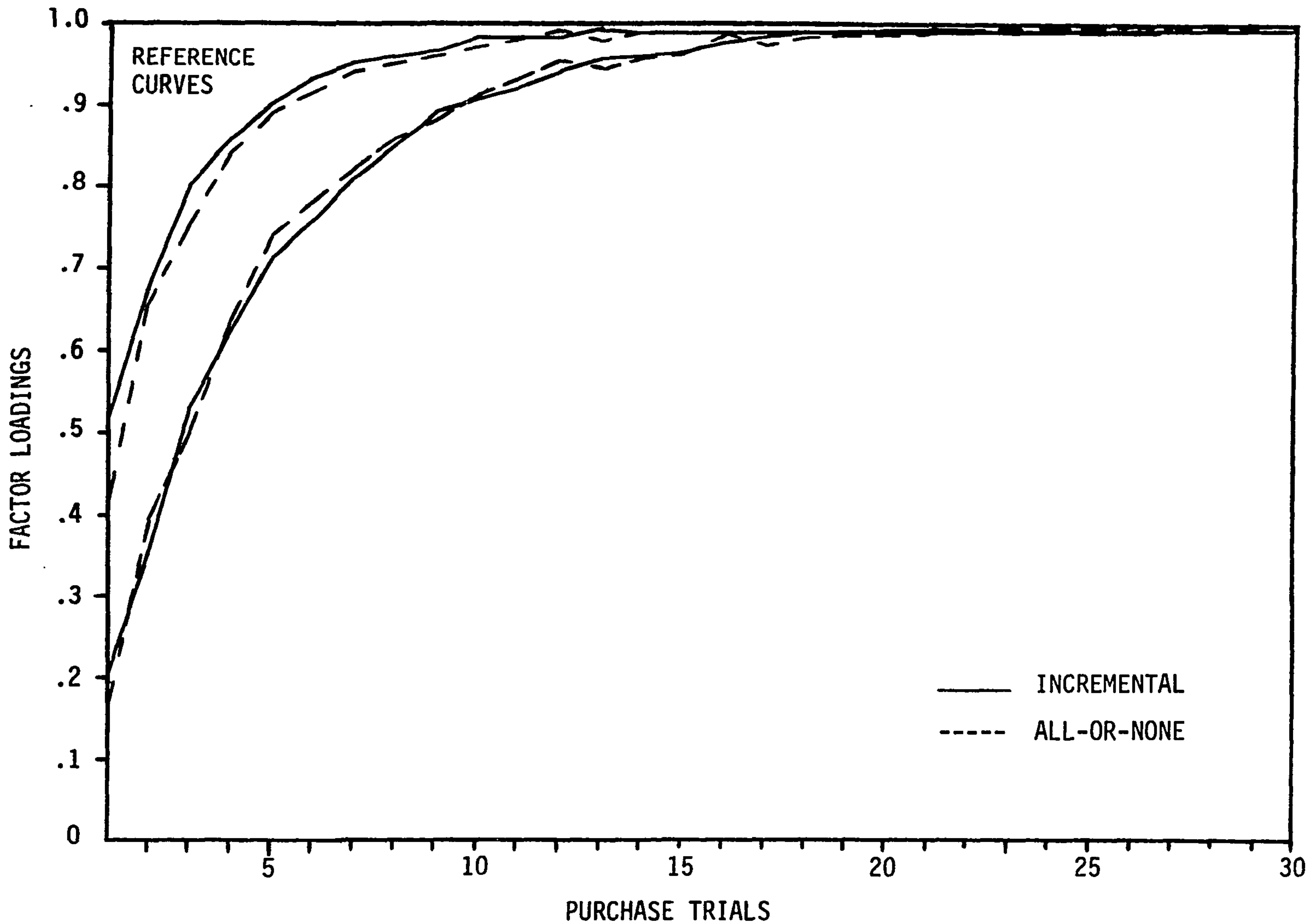

Fig. 7.--Reference curves with corresponding mean learning curves for set number 1.

TABLE 6

REFERENCE CURVES (FACTOR LOADINGS) FOR INCREMENTAL LEARNING DATA FROM SET 2
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.38390	-.55892	.44229
2	.52706	-.60362	.31750
3	.60520	-.57367	.16814
4	.67873	-.55940	.03769
5	.78452	-.40659	-.15909
6	.80831	-.39522	-.19687
7	.81238	-.32466	-.25447
8	.84861	-.19405	-.27359
9	.87095	-.15600	-.24719
10	.87201	-.09421	-.21028
11	.89826	-.01487	-.19198
12	.91823	.01050	-.11422
13	.91672	.05480	-.11620
14	.91502	.08533	-.09143
15	.92334	.08890	-.03054
16	.92048	.14164	-.05142
17	.92251	.13860	-.01825
18	.92014	.11887	.02051
19	.94386	.15902	.05172
20	.92066	.15392	.11122
21	.92745	.16298	.07873
22	.92762	.18704	.07864
23	.91962	.19345	.11593
24	.91551	.18989	.11866
25	.92588	.15849	.10438
26	.91660	.17982	.13484
27	.91157	.19634	.14047
28	.90760	.20307	.10737
29	.89994	.20019	.15332
30	.90598	.16214	.16303
% Of Variation Explained	74.24	7.55	2.97
Eigenvalue	22.2716	2.2761	0.8907

TABLE 7

REFERENCE CURVES (FACTOR LOADINGS) FOR ALL-OR-NONE LEARNING DATA FROM SET 2
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.25199	-.13056	-.84749
2	.49902	.46104	-.26629
3	.59912	.39633	-.25534
4	.64151	.52588	-.08128
5	.72930	.44624	.01995
6	.76591	.32854	.03555
7	.80138	.33217	.04908
8	.84112	.21115	.09829
9	.86674	.15939	.13796
10	.87043	.19889	.10814
11	.88777	.08501	.10794
12	.89410	.05973	.09353
13	.89023	.01081	.06338
14	.92312	.00033	.08878
15	.92077	-.07241	.05563
16	.89958	-.13805	.08168
17	.91744	-.11759	.04010
18	.92291	-.15009	.05492
19	.91540	-.13174	-.01141
20	.91935	-.16643	.08474
21	.92358	-.14717	-.02729
22	.92258	-.14626	-.01120
23	.93465	-.14946	-.02915
24	.94256	-.12691	-.04116
25	.89118	-.19554	-.09120
26	.91625	-.19253	-.01993
27	.91987	-.16616	-.05550
28	.90690	-.17788	-.07106
29	.92529	-.16601	-.05993
30	.90130	-.20349	-.05697
% Of Variation Explained	72.55	5.31	3.29
Eigenvalue	21.7635	1.5918	0.9865

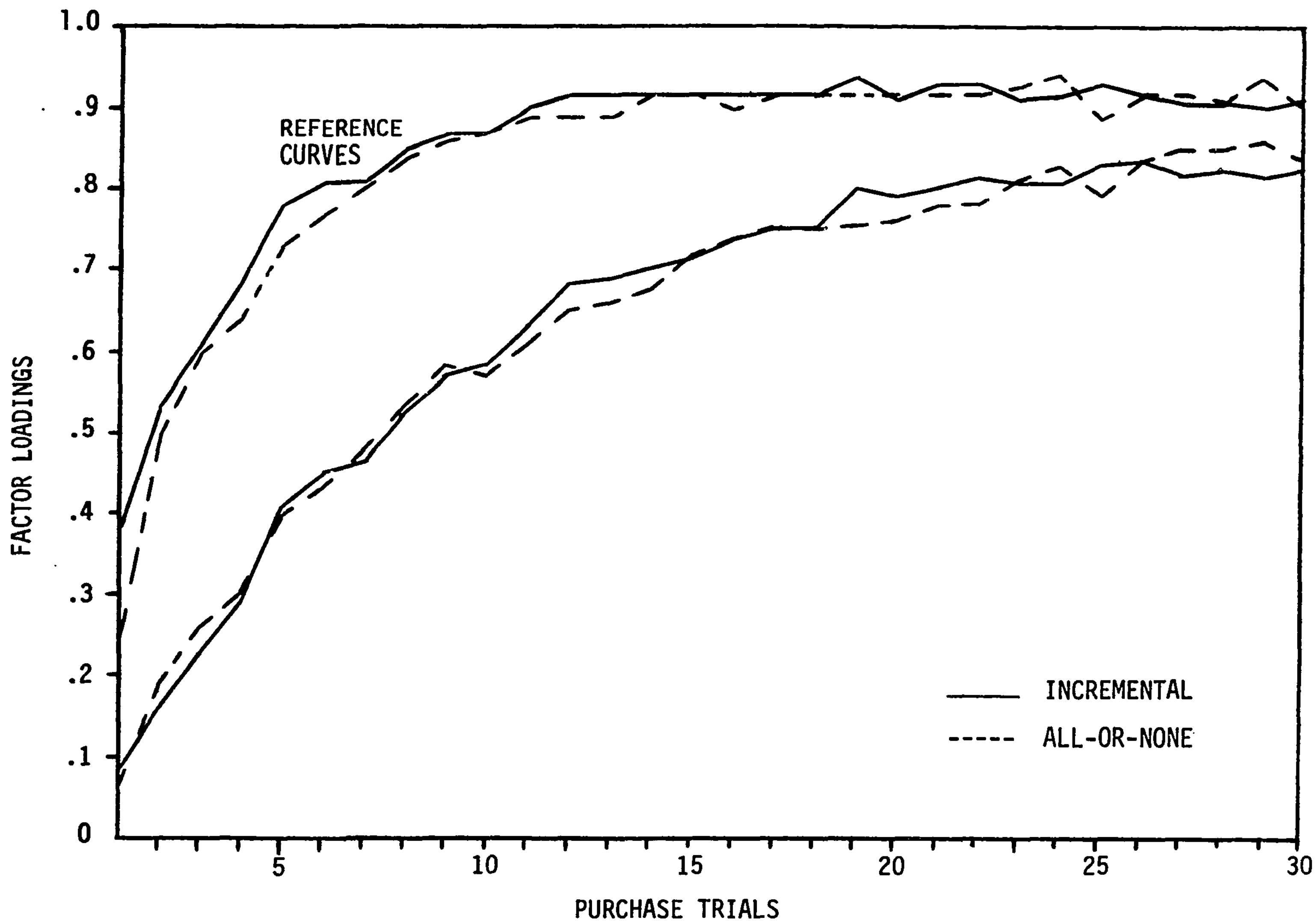

Fig. 8.--Reference curves with corresponding mean learning curves for set number 2.

TABLE 8

REFERENCE CURVES (FACTOR LOADINGS) FOR INCREMENTAL LEARNING DATA FROM SET 3
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.59777	-.15542	-.16168
2	.53448	.57188	-.09145
3	.63858	-.12040	.28624
4	.58367	.12534	.29024
5	.60229	.06802	.52867
6	.61991	.33692	.15653
7	.60826	.24099	.05353
8	.61503	-.05713	.05387
9	.60422	.26627	-.19300
10	.60833	.29226	-.14749
11	.61481	.02534	-.02805
12	.62070	-.08119	.32897
13	.66116	-.05650	-.06044
14	.68881	-.04645	.14905
15	.66886	-.22533	.09207
16	.67462	-.08427	-.10069
17	.73397	-.05720	.03333
18	.68250	-.13378	.04535
19	.71396	.04468	.11875
20	.70339	.10267	-.23394
21	.70664	.02755	-.05424
22	.71446	-.02036	.03768
23	.71109	.07054	-.07476
24	.71278	-.06457	-.16152
25	.66729	-.05233	-.22118
26	.74203	-.25705	-.09988
27	.75295	-.04265	-.13864
28	.73007	-.15772	.02573
29	.72939	-.02352	-.12330
30	.70700	-.26460	-.17463
% Of Variation Explained	44.53	3.33	3.19
Eigenvalue	13.3598	0.9978	0.9555

TABLE 9

REFERENCE CURVES (FACTOR LOADINGS) FOR ALL-OR-NONE
LEARNING DATA FROM SET 3
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.32751	-.13382	.79691
2	.46266	-.29549	.26453
3	.48430	-.37467	.24813
4	.53313	-.33004	-.05268
5	.55163	-.42241	-.04694
6	.59978	-.27429	.00585
7	.56583	-.35020	-.07761
8	.62005	-.25939	-.08546
9	.63548	.-13253	-.29200
10	.66591	-.14861	-.13733
11	.67099	-.18248	-.09204
12	.70858	-.06052	-.12896
13	.73100	-.10393	.00619
14	.74137	.03451	-.07117
15	.74724	.01297	-.11197
16	.71210	-.20245	-.09558
17	.71553	.03955	-.08025
18	.77497	.11745	-.02033
19	.75431	.02486	-.04028
20	.78398	.15337	-.05501
21	.78357	.17871	-.04525
22	.79686	.24579	-.05011
23	.78665	.18675	-.03741
24	.78204	.15582	-.01035
25	.79663	.18738	.07432
26	.78731	.16700	.10125
27	.79331	.22490	.13635
28	.79790	.20759	.07124
29	.79632	.19621	.15115
30	.78764	.23336	.13869
% Of Variation Explained	49.03	4.56	3.47
Eigenvalue	14.7088	1.3685	1.0400

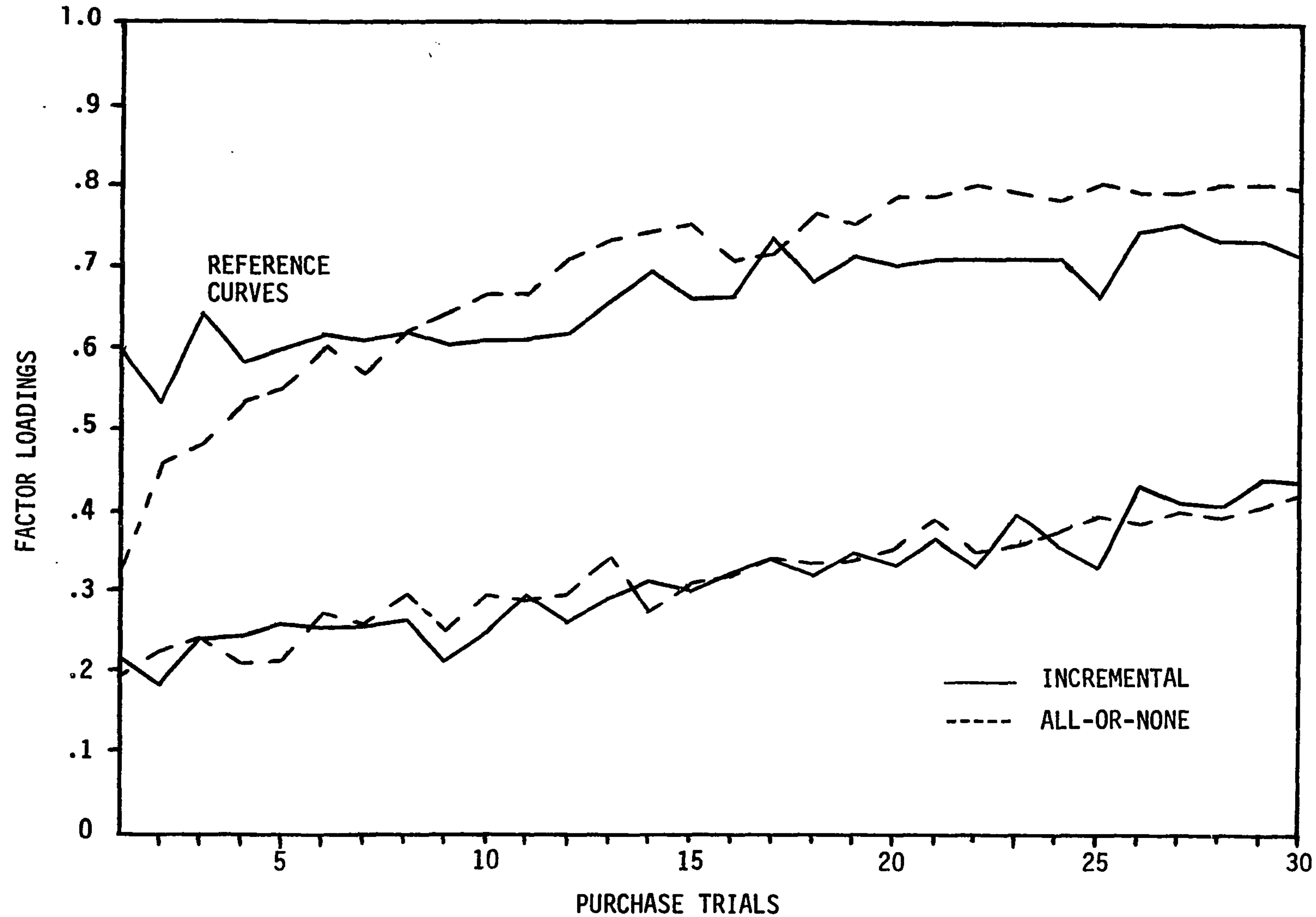

Fig. 9.--Reference curves with corresponding mean learning curves for set number 3.

TABLE 10

REFERENCE CURVES (FACTOR LOADINGS) FOR INCREMENTAL LEARNING DATA FROM SET 4
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.62218	-.33238	.00953
2	.62693	-.38635	.21808
3	.68725	-.25673	-.25177
4	.70769	-.14219	.05914
5	.71775	-.28789	.09455
6	.73493	-.24189	.18197
7	.71282	-.17917	-.26283
8	.73920	-.10489	.13283
9	.70326	-.08007	.08071
10	.66548	-.00125	.33020
11	.70062	-.18247	-.14335
12	.71520	-.07455	-.18929
13	.71580	.01095	-.33851
14	.76359	-.05802	.04127
15	.72461	-.05328	-.18903
16	.72437	.10141	-.14146
17	.71733	.13808	-.25073
18	.75011	.04265	.07273
19	.73791	.14666	-.12135
20	.72128	.20328	.03645
21	.76516	.24398	-.12453
22	.72248	.15371	-.01757
23	.74975	.14931	-.00600
24	.72924	.21476	.19007
25	.74113	.11263	.02819
26	.75219	.19989	.00293
27	.75168	.12130	.09580
28	.74436	.13946	.04340
29	.72781	.18326	.28240
30	.76433	.05403	.15063
% Of Variation Explained	52.13	3.17	2.79
Eigenvalue	15.6399	0.9509	0.8378

TABLE 11

REFERENCE CURVES (FACTOR LOADINGS) FOR ALL-OR-NONE LEARNING DATA FROM SET 4 (n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.65464	-.10116	-.05979
2	.63601	.28178	-.10376
3	.64156	-.20903	.03168
4	.62083	-.04585	-.11656
5	.63197	-.20066	.16590
6	.64610	.09700	.19320
7	.64336	-.35257	.04084
8	.60843	.03569	.36297
9	.65956	.14647	-.18031
10	.64678	.15710	-.12582
11	.65794	-.22461	-.15002
12	.66724	.08257	-.13702
13	.66063	.34086	-.01513
14	.70009	.18880	.04297
15	.65128	.16378	.23377
16	.64542	-.12500	.10295
17	.65727	.23453	.07986
18	.65447	.07057	.07189
19	.67178	-.03428	-.35771
20	.60892	.08285	.37293
21	.65230	-.27263	-.18333
22	.66286	-.16425	.08169
23	.62805	.16768	.08898
24	.61777	.03846	-.12448
25	.64640	.05333	.02899
26	.63750	-.03443	-.00608
27	.61484	.23725	-.26296
28	.66357	-.16918	.11184
29	.66209	-.16651	-.28851
30	.60376	-.29160	.15866
% Of Variation Explained	41.66	3.36	3.06
Eigenvalues	12.4988	1.0090	0.9174

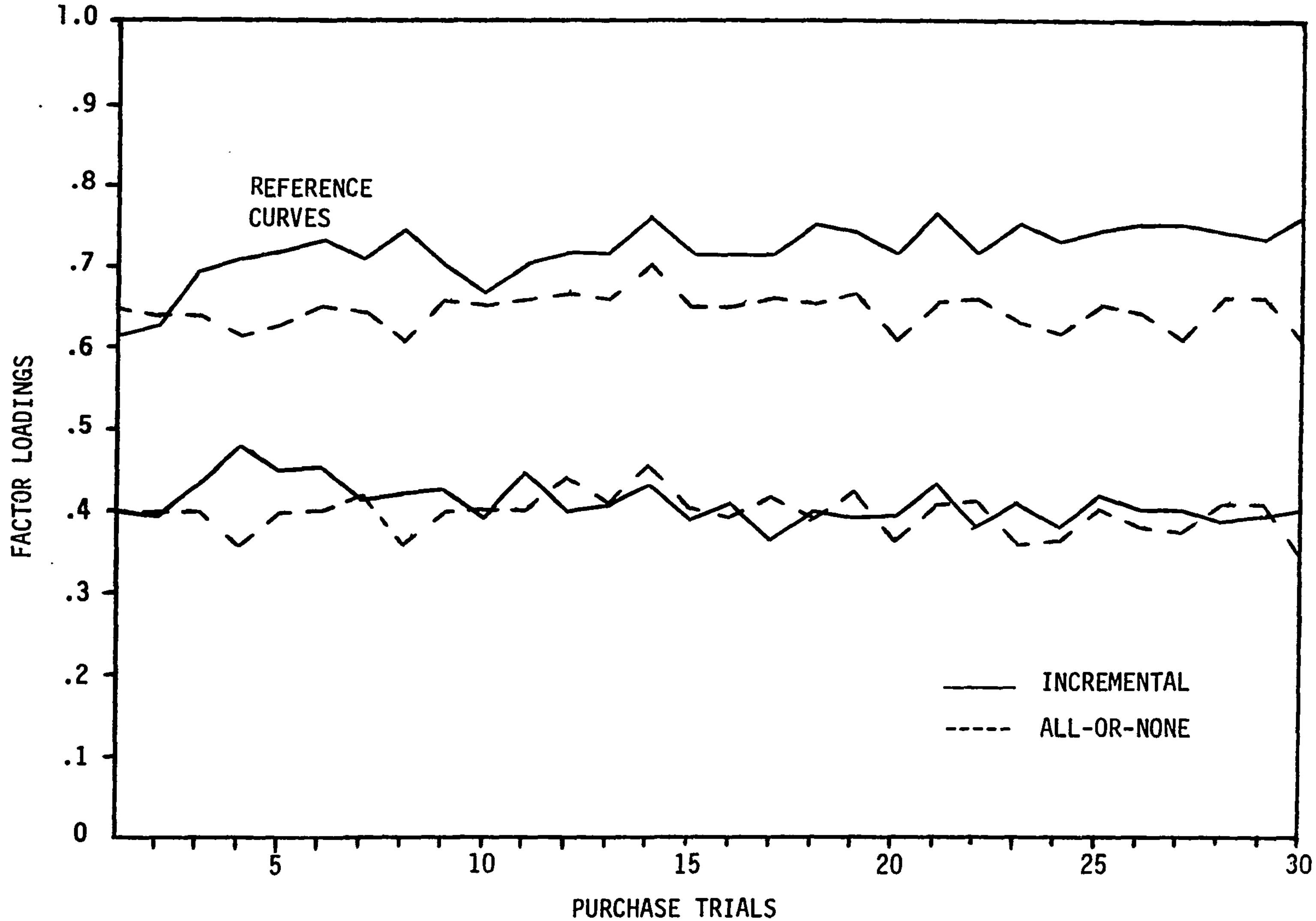

Fig. 10.--Reference curves with corresponding mean learning curves for set number 4.

TABLE 12

REFERENCE CURVES (FACTOR LOADINGS) FOR INCREMENTAL LEARNING DATA FROM SET 5
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.72018	.37424	.15121
2	.70213	.41949	.18093
3	.72863	.31737	.24089
4	.71769	.37838	.22989
5	.73537	.34909	.15987
6	.73211	.27221	.03164
7	.75662	.24030	.00681
8	.75624	.24098	.07050
9	.71919	.26992	-.07213
10	.76893	.13743	-.06052
11	.75921	.14944	-.13609
12	.76531	.01415	-.08365
13	.72691	.23232	-.31860
14	.78486	.00862	-.30947
15	.76711	-.04587	-.12128
16	.76394	-.01533	-.04847
17	.75658	-.04512	-.10649
18	.77844	-.06573	-.08974
19	.77292	-.04480	-.14798
20	.77727	-.08230	-.30639
21	.75063	-.15618	-.13759
22	.76764	-.28687	-.00156
23	.73346	-.20931	.12029
24	.74912	-.29110	.19466
25	.73955	-.25658	.01677
26	.77889	-.35539	.09929
27	.74533	-.34147	-.02804
28	.73864	-.38690	.17332
29	.73919	-.38690	.15017
30	.68476	-.37092	.23390
% Of Variation Explained	55.89	6.80	2.57
Eigenvalue	16.7682	2.0385	0.7715

TABLE 13

REFERENCE CURVES (FACTOR LOADINGS) FOR ALL-OR-NONE
LEARNING DATA FROM SET 5
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.59006	.41914	-.29278
2	.59042	.43632	-.25417
3	.63431	.38518	-.33287
4	.57987	.47803	-.24722
5	.68171	.32084	-.27348
6	.67422	.36285	-.07887
7	.70138	.29185	-.05832
8	.72258	.26159	.00949
9	.73011	.25693	.01539
10	.71542	.15643	.10132
11	.68028	.28024	.05763
12	.66631	.23958	.23420
13	.72319	.19607	.24480
14	.71761	.07602	.21701
15	.64977	.04051	.28381
16	.68186	.09736	.37290
17	.72763	-.10221	.27462
18	.73961	-.08867	.21570
19	.68351	-.03949	.31628
20	.73436	-.23289	.18914
21	.74880	-.23665	.11921
22	.67585	-.28095	-.03616
23	.66079	-.35038	-.06106
24	.64812	-.39434	-.11993
25	.63919	-.34773	-.18484
26	.71028	-.42851	-.23870
27	.66437	-.38981	-.17352
28	.67782	-.39267	-.17941
29	.68231	-.47606	-.18528
30	.64512	-.39312	-.16990
% Of Variation Explained	46.38	9.64	4.37
Eigenvalue	13.8986	2.8921	1.3101

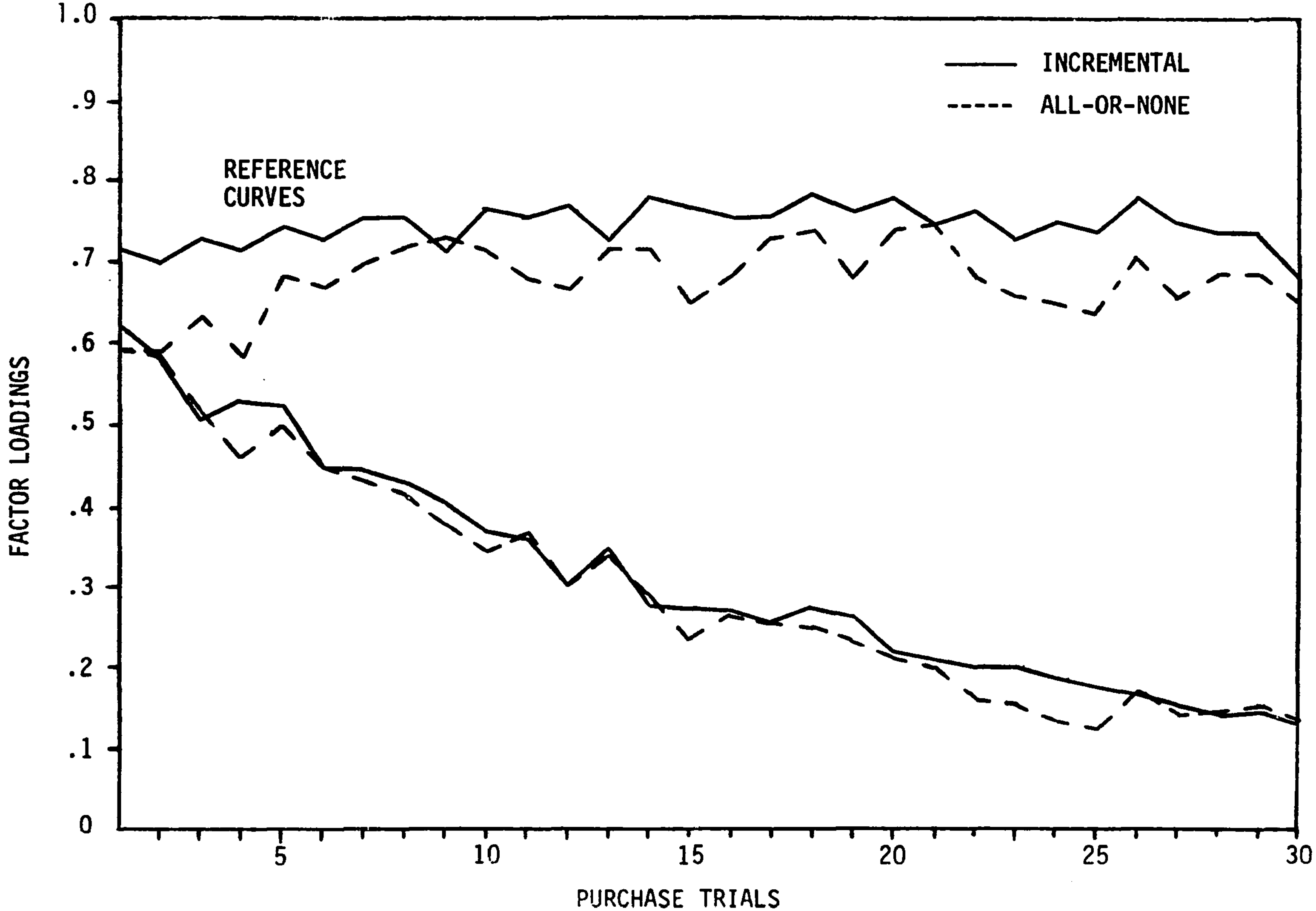

Fig. 11.--Reference curves with corresponding mean learning curves for set number 5.

CHAPTER V

DISCUSSION AND CONCLUSIONS

The discussion and conclusions are divided into two sections. First, the pertinent aspects of the reference curves are elaborated and conclusions tendered. Secondly, the brand loyalty scores generated are interpreted and warranted conclusions are provided.

Analysis and Evaluation of Reference Curves

Sheth[1] describes the factor loadings of the reference curve as aggregate parameters for each of the purchase trials. It is given that the buyer's behavior results from both environmental influences and internal motives. The factor loadings, then, characterize the environmental influences on the buyer's manifested behavior; in other words, the effects of aggregate purchase behavior. Because the purchase trials aggregate the environmental influences, the corresponding factor loadings remain constant across consumers. The individual parameters (factor scores or brand loyalty scores) measure the nature of individual buyer influence. Thus, the factor analytic model separates the aggregate and individual parameters during the market period studied.

---

[1]Jagdish N. Sheth, " A Factor Analytical Model of Brand Loyalty," Journal of Marketing Research, (November, 1968), p. 397.

Since the results for all 10 learning situations are remarkably similar on the relevant criteria, the outcome is discussed in general.

Table 4 through Table 13 give the first three reference curves for each of the 10 different learning situations. The first reference curve for the 10 situations explains from a high of about 90% of the variation for the incremental data of set 1 to a low of 41.6% for the all-or-none data of set 4. The second reference curve explains from a high of 9.6% of the variation for the all-or-none data of set 5 to a low of about 3.2% for both the incremental data of set 4 and the all-or-none data of set 1. Lastly the third reference curve explains from a high of 4.4% of the variation for the all-or-none data of set 5 to a low of 2% for the incremental data of set 1. Actually, 13 reference curves (factors) for each case were printed out. Because the principal components analysis program generates the eigenvalues in sequence from highest to lowest, however, it is known that the fourth and remaining eigenvalues are negligible, and therefore, are not presented.

In all cases the first reference curve seems to be the only significant curve. Clearly, the eigenvalues drop off rapidly after the first. The only case that may be of some contention is the all-or-none data of set 5 where 9.6% of the variation is explained by the second reference curve. Howard and Sheth[2] when explaining an identical outcome (9.6% of variation explained on the second reference curve of their rice sample having 6 trials and 17 panel members) offer that

---

[2]John A. Howard and Jagdish N. Sheth, <u>The Theory of Buyer Behavior</u>, (New York: John Wiley and Sons, 1969), p. 258.

the second and other reference curves act as correction terms for the first reference curve. They further report that the magnitude of the variation accounted for is insufficient for acceptance as another dimension of the underlying function. In the present data, 7.5% of variation explained for incremental set 2 is the next highest amount captured by a second reference curve.

For these reasons it is evident that only one dominant reference curve underlies each of the 10 different learning conditions. Based upon the present data analysis, it makes no difference whether consumers are learning to buy brand A gradually or in an all-or-none manner. In all cases there is only one significant reference curve. Accordingly, it is concluded that the factor analytic model is not responsive when number of reference curves is the criterion for differentiating between incremental and all-or-none learning.

To facilitate interpretation, the first reference curves (factor loadings) for both incremental and all-or-none learning by set number, are plotted in Fig. 7 through Fig. 11. Also, for ease of visual comparison, the respective mean curves (proportion of brand A purchased at each trial) already observed in Fig. 2 through Fig. 6 are provided again in Fig. 7 through Fig. 11.

Remember that when only one dominant factor is found, the shape of the reference curve should be nearly identical to the mean learning curve computed on the original data. The first reference curves and the mean curves, by set, do have the same shape but are not identical because the reference curves were computed from the standardized matrix

$D^{-\frac{1}{2}}Y$ and not Y the raw data matrix.[3] Additionally, Howard and Sheth[4] indicate that the reference curves are somewhat different from the mean curves because the reference curve is only a good approximation to the mean curve, and that the two curves will be the same only if the shape and form of the individual learning curves are the same.

Unquestionably, the first reference curves, within a set, are more similar to each other than they are to the mean curve. By inspection, just as with a given set of mean learning curves, there is no way to discriminate between the incremental reference curve and the all-or-none reference curve within a given set. Therefore, on this aspect of data analysis, it is concluded that the factor analytic model does not allow the identification of type of learning: the known incremental and all-or-none learning data built into the research design remain disguised.

It must be observed, however, that in general the reference curves do reflect the mean learning curve. The preceding statement supports and is consistent with Sheth's[5] suggestions that the factor analytical model (basically a statistical technique) is empirical in that it provides brand loyalty measures from a sample of consumer purchase data. That is, it is not necessary that any _a priori_ hypothesis be formulated about the nature of the brand loyalty curve

---

[3]Jagdish M. Sheth "A Behavioral and Quantitative Investigation of Brand Loyalty" (unpublished doctoral dissertation, University of Pittsburgh, 1966), p. 147.

[4]Howard and Sheth, _loc. cit._

[5]Sheth, _op. cit._ 1968, pp. 395, 401-402.

across purchase trials. Or alternatively, if a theoretical hypothesis is projected, use of the factor analytic outcome allows the acceptance or rejection of the theoretical hypothesis in that particular situation.

The crucial point to recognize is that Sheth is referring to reference curves as indicative of the form or shape of the aggregate brand loyalty curve over time. Seemingly, if the interest in learning curves is restricted to mean performance it may be more economical and accurate (as opposed to an approximation procedure) to merely calculate, from the raw data, and plot the proportion of preferred brands at each trial. Obviously, whether reference curves or mean curves are analyzed the possibility of identifying the indivduals' underlying learning process is eliminated.

Because of the consistent findings of one dominant reference curve for both incremental and all-or-none data regardless of fast learning or slow learning, in the aggregate respectively, it was decided to run more data combinations.[6] The existing learning data were mixed in various combinations of "slow" and "fast" learners, and incremental and all-or-none learners from sets 1 and 3. In total, four new runs were made.

First, in Table 14 is represented the factor analytic outcome of one-half (150) slow all-or-none learners from set 3 and the other half fast incremental learners from set 1. These four combinations are not

[6]Recall that all-or-none learners are either totally learned, or learned nothing; however, the mean learning curve (with its averaging effect) is gradual. For convenience, all-or-none learners will be referred to as slow or fast, depending upon the shape of the mean curve.

TABLE 14

REFERENCE CURVES (FACTOR LOADINGS) FOR ALL-OR-NONE LEARNING DATA FROM SET 3 AND INCREMENTAL LEARNING DATA FROM SET 1
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.38564	-.58653	.60921
2	.57826	-.56577	.14909
3	.70889	-.46531	-.10192
4	.79104	-.34809	-.18539
5	.83075	-.28207	-.25094
6	.84450	-.23110	-.21588
7	.86948	-.18647	-.22745
8	.88715	-.11154	-.18679
9	.90613	-.02344	-.17818
10	.91233	-.01909	-.12869
11	.92031	.02360	-.09093
12	.92989	.03683	-.04280
13	.92864	.04753	.03476
14	.93712	.07370	-.02835
15	.93318	.05798	-.04437
16	.93054	.04990	-.04868
17	.92888	.09014	.04594
18	.94493	.10880	.06625
19	.93601	.08801	.08062
20	.94868	.12833	.03970
21	.93559	.13422	.07401
22	.94236	.14119	.08955
23	.95072	.15223	.04187
24	.93909	.13935	.05760
25	.94046	.11646	.12662
26	.94232	.12625	.09535
27	.93369	.12389	.12336
28	.93757	.12051	.11628
29	.94099	.12932	.12776
30	.93487	.12230	.12417
% Of Variation Explained	79.21	4.71	2.73
Eigenvalues	23.7644	1.4141	0.8178

TABLE 15

REFERENCE CURVES (FACTOR LOADINGS) FOR INCREMENTAL LEARNING DATA FROM SET 3 AND ALL-OR-NONE LEARNING DATA FROM SET 1 (n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.44892	-.73799	-.36331
2	.63018	.12702	-.52514
3	.70440	.08949	-.44895
4	.77925	.35031	-.17780
5	.82267	.26919	-.13061
6	.85149	.17568	-.12511
7	.86209	.21053	-.03334
8	.88212	.10176	.02965
9	.88726	.11838	.01967
10	.91269	.03254	.02826
11	.90470	.01678	.08627
12	.92358	-.02176	.05108
13	.91625	.03759	.10965
14	.92820	-.00931	.06668
15	.92543	.00763	.11422
16	.92185	-.08318	.06557
17	.92952	-.03233	.04433
18	.92748	-.05874	.05826
19	.93221	-.00753	.05563
20	.92822	-.08166	.03918
21	.92640	-.08559	.01071
22	.93397	-.03855	.06746
23	.92068	-.02939	.04002
24	.92171	-.06880	.04444
25	.92386	-.03773	.07603
26	.92565	-.10196	.06101
27	.92967	-.07398	.03868
28	.93041	-.04340	.10319
29	.92170	-.12857	.00186
30	.90920	-.09134	.08117
% Of Variation Explained	77.75	3.15	2.56
Eigenvalues	23.3238	0.9435	0.7672

TABLE 16

REFERENCE CURVES (FACTOR LOADINGS) FOR INCREMENTAL LEARNING DATA FROM SET 3 AND ALL-OR-NONE LEARNING DATA FROM SET 3
(n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.50542	-.14715	.23277
2	.50618	-.08179	-.58989
3	.56688	-.21889	.22971
4	.54658	-.34378	.03809
5	.54680	-.52189	.29187
6	.58533	-.32475	-.29924
7	.56644	-.37448	-.16863
8	.62955	-.20026	-.00162
9	.60598	-.19060	-.32271
10	.65149	-.12685	-.17829
11	.65276	-.04039	.10420
12	.70096	-.09823	.31529
13	.70137	.06502	-.01618
14	.72910	.02317	.02015
15	.71043	-.04496	.04312
16	.70373	-.07023	.07273
17	.73939	.04731	.01173
18	.74290	.11446	.14404
19	.74426	-.01959	.04837
20	.77493	.12920	.06121
21	.76745	.08751	-.03726
22	.77421	.16211	-.01858
23	.75264	.16387	-.05526
24	.76032	.15325	-.04302
25	.74740	.20644	-.06087
26	.78164	.18882	.07870
27	.77256	.13903	-.00307
28	.76850	.20380	-.00811
29	.78391	.21966	-.01751
30	.74309	.23881	.00538
% Of Variation Explained	47.77	3.96	3.19
Eigenvalues	14.332	1.1892	0.9575

TABLE 17

REFERENCE CURVES (FACTOR LOADINGS) FOR INCREMENTAL LEARNING DATA FROM SET 3 AND INCREMENTAL LEARNING DATA FROM SET 1 (n=30, N=300)

Trials	Reference Curve I	Reference Curve II	Reference Curve III
1	.52600	-.46270	.62935
2	.65137	-.50862	.14018
3	.76195	-.43116	-.04530
4	.78977	-.40243	-.22272
5	.82947	-.32052	-.24793
6	.86012	-.24791	-.16954
7	.87472	-.12894	-.21742
8	.89452	-.06469	-.12423
9	.89991	-.04402	-.16144
10	.91528	-.02137	-.06140
11	.90296	.04085	-.06653
12	.92019	.03216	-.03929
13	.92193	.09620	-.05458
14	.92962	.08100	-.00339
15	.92537	.11660	-.04159
16	.92049	.12546	.07186
17	.92832	.10918	.04586
18	.92613	.11674	.05308
19	.93021	.10301	-.00726
20	.92545	.11107	.09090
21	.92445	.07997	.08839
22	.93092	.11579	.02450
23	.91766	.11032	.02999
24	.91870	.10985	.06332
25	.91965	.12786	.02752
26	.92217	.12837	.09760
27	.92556	.10344	.06589
28	.92611	.14706	.02430
29	.91834	.09769	.14299
30	.90575	.13967	.08275
% Of Variation Explained	78.50	4.14	2.45
Eigenvalues	23.5492	1.2426	0.7351

necessarily presented in a particular order, however, the Table 14 combination seemed intuitively the most rigorous test for the model. The reason is that the fast incremental very nearly approaches all-or-none type learning; if the person buys brand A twice in a row she is totally learned.

Secondly, in Table 15 the combination is reversed. One-half slow incremental learners from set 3 were run with one-half fast all-or-none learners from set 1. Here, for the same type reasoning as above, it was considered that the chances were best for generating two or more dominant reference curves. In general, the all-or-none learners learn completely on the first few trials, while the incremental learners learn quite gradually over all trials.

In the third mixed combination in Table 16 the mean curve characteristics were held constant by mixing half and half incremental and all-or-none learners both from set 3. The last combination shown in Table 17 mixes one-half slow incremental learners from set 3 and one-half fast incremental learners from set 1. Again, intuitively there seemed some basis for anticipating at least two underlying dimensions; notice the difference in appearance between the mean curves of set 1 and set 3.

Investigation of the four Tables shows these results to be consistent with the previous analyses in that, regardless of either differential levels of learning or of type of learning process within a learning data set, the factor analytic model produces only one reference curve. Such research findings only can strengthen the conclusion that Sheth's factor analytic model cannot reliably be used as an instrument for verifying the incremental learning aspect of the Howard-Sheth theory of buyer behavior.

In specific reference to the hypotheses in this study, all three must be rejected. Hypothesis 1, that Sheth's proposed factor analytic measure of brand loyalty does differentiate between incremental learning data and all-or-none learning data, is rejected based upon the following research findings.

The first finding is that in all data sets examined in this study, for both incremental and all-or-none learning data, there is only one dominant reference curve. The second finding is that the first reference curves within a given set, are more similar to each other than they are to their corresponding mean curves. Therefore, the conclusion is made that the factor analytic model is inappropriate when either the number of reference curves or the characteristics of shape of reference curves is the criterion for differentiating between incremental and all-or-none learning.

Hypothesis 2 for factor analysis applied to incremental learning demands that there be only one dominant reference curve reflecting the mean curve and accordingly each buyer will have one BLS. Hypothesis 3, applied to all-or-none learning, demands that multiple reference curves account for the underlying learning data and that each buyer will have as many BLS's as there are significant reference curves. For the same reason, based upon the present research findings, neither of these two demands can be honored. The reason simply is that with real world data, where there is no _a priori_ structuring of the learning data as in this experiment, the researcher cannot know the individual underlying psychological behavior process that is embodied in a sample of purchase data. Therefore, even all those cases where incremental learning data did demonstrate one dimension, they must be rejected because these data might just as well have been all-or-none learners.

There is a possible explanation for the apparent total insensitivity of the factor analytic model to the many different configurations and learning patterns offered in this study. The use of binary data (1 or 0) as input may blunt the effectiveness of the model. Conventionally, the input data are comprised of continuous variables. Further investigation on this point seems warranted.

Analysis and Evaluation of Brand Loyality Scores

Recall that the BLS's (factor scores) are weights to the reference curves. Additionally each buyer has as many BLS's as there are reference curves. In the present study since there is only one reference curve in all cases, each buyer has only one BLS. The BLS's are unique for each possible combination of purchase protocols of n trials. In other words a given BLS accounts for both frequency and pattern of purchase.

The number of different mathematically possible BLS's is quite surprising. With two alternatives, and in this study 30 trials, an initially homogeneous group of buyers could be differentiated into over one billion different BLS's! It is evident that after a few trials a group of initially homogeneous buyers, in the sense of all having the same probability distribution over the response alternatives, could become too finely discriminated. That is, this tremendous richness of the BLS's rapidly exceeds practical experimental limits.

Selected purchase protocols, associated BLS's and number of brand A purchases are presented in Table 18 for inspection. Note that the first two buyers have the same frequency and pattern of purchase and so have identical BLS's. Also, it is readily observable that those buyers with different frequencies of purchase (and of course different patterns) have different BLS's. Lastly, all the buyers with the same frequency of purchase but with different patterns have different BLS's.

Actually, the number of different BLS's generally is fewer than 300. The number of BLS's is especially restricted for a group characterized by a fast rate of learning i.e., they quickly learn to buy brand A. The consequence is that their buying patterns are quite homogeneous. For example, the incremental learners (300) of set 1 represent only 27 unique BLS's with 21 different frequencies of brand A purchases. By contrast, it was laboriously determined that the all-or-none buyers of set 3 produced 283 different BLS's: the patterns obviously are much more heterogeneous. Relatedly, Sheth's [7] and Howard and Sheth's[8] data from their two samples of brand loyalty for rice demonstrate considerable homogeneity i.e., for the sample of 17 people there were only 7 different BLS's (7 protocols the same) and for the sample of 14 people there were only 6 different BLS's (8 protocols the same).

With the general nomenclature of BLS's elaborated, now, it is demonstrated that isolating individual BLS's is essentially trivial

---

[7]Sheth, op. cit., 1966; Sheth, op. cit., 1968.

[8]Howard and Sheth, op. cit., 1969.

TABLE 18

ILLUSTRATION OF BRAND LOYALTY SCORES SELECTED FROM ALL-OR-NONE LEARNING DATA OF SET 1

Row Number	Buyer Number	1	2	3	4	5	6	7	8	9	10	. . . ,	30	Brand A Purchases	Brand Loyalty Scores
1	273	1	1	1	1	1	1	1	1	1	1		1	30	.06544
2	266	1	1	1	1	1	1	1	1	1	1		1	30	.06544
3	296	0	1	1	1	1	1	1	1	1	1		1	29	.06329
4	40	1	0	1	1	1	1	1	1	1	1		1	29	.06316
5	285	0	0	1	1	1	1	1	1	1	1		1	28	.06101
6	127	1	1	1	0	0	1	1	1	1	1		1	28	.06092
7	151	1	1	0	0	1	1	1	1	1	1		1	28	.06088
8	9	0	0	1	0	1	1	1	1	1	1		1	27	.05874
9	102	0	1	0	0	1	1	1	1	1	1		1	27	.05873
10	293	0	0	0	1	1	1	1	1	1	1		1	27	.05872
11	213	1	0	0	0	1	1	1	1	1	1		1	27	.05860
12	55	0	0	1	0	1	0	1	1	1	1		1	26	.05650
13	64	0	0	1	0	0	1	1	1	1	1		1	26	.05649
14	62	0	0	0	1	0	1	1	1	1	1		1	26	.05648
15	149	0	0	0	0	1	1	1	1	1	1		1	26	.05645
16	11	1	1	0	0	0	0	1	1	1	1		1	26	.05639
17	132	0	0	0	0	0	1	1	1	1	1		1	25	.05420
18	297	0	0	0	0	0	0	1	1	1	1		1	24	.05196
19	119	0	0	0	0	0	0	0	1	1	1		1	23	.04972
20	44	0	0	0	0	0	0	0	0	1	1		1	22	.04751

in regard to conceivable practical applications. In Table 19, the correlations between total purchases of brand A for each person by type of learning and the corresponding brand loyalty scores are provided. The correlations were computed only on the first 50 purchases; which in effect is the same as random sampling, since each of the 300 protocols is statistically independent of the others.

Immediately, the almost total redundancy of the BLS's is apparent. The sampling of 1 in 6 BLS's and some severe rounding may have reduced some of the correlations. Especially the correlations in set 1 may have been affected by restricted range. Correlations also were run on the Howard-Sheth (referenced above) data. In this instance, 100% of the panel members were included. For the sample of N = 17, the correlation between total purchases and BLS's is r = .999, and for the sample of N = 14 the correlation is also r = .999.

These combined discoveries of the almost perfect substitutability between total purchases and BLS's can only reinforce the conclusion that BLS's as generated from the factor analytic model have negligible value. According to Sheth,[9] the individual BLS's are the most valuable feature of the factor analytic model because the distribution of the indivdiual BLS's allows analysis and interpretation for market segmentation decisions. Howard and Sheth,[10] when highlighting the implications of the factor analytic model of brand loyalty state in effect that the proportion of purchases of brand A in the total purchases of a buyer is a much

[9]Sheth, op. cit., 1968, p. 395.

[10]Howard and Sheth, op. cit., p. 271.

TABLE 19

CORRELATIONS BETWEEN PURCHASE FREQUENCY BY TYPE OF LEARNING AND CORRESPONDING BRAND LOYALTY SCORES

Set Number	Purchase Frequency Correlated With BLS	r
1	Incremental	.935
	All-or-None	.920
2	Incremental	.989
	All-or-None	.988
3	Incremental	.974
	All-or-None	.996
4	Incremental	.994
	All-or-None	.971
5	Incremental	.996
	All-or-None	.981

N=300, n=50

less desirable measure than are brand loyalty scores. The last two statements are not in concert with the research findings of this study or with the treatment of their own findings.

The following is offered as a detailed and rather technical explanation of why the correlation of first factor scores with total purchases (row sums of Y) is so high.[11] Factor scores (BLS's) were computed from the equation,

$$F_{(N,r)} = YD_{Y'Y}^{-\frac{1}{2}}AD^{-1},$$

which implies that the vector of scores on the first reference learning curve, $\underline{f}_{\cdot 1}$, is equal to,

$$\underline{f}_{\cdot 1(N,1)} = Y_{(N,n)}D^{-\frac{1}{2}}_{Y'Y(n,n)}\underline{a}_{\cdot 1(n,1)}\frac{1}{\lambda_1},$$

where $\underline{a}_{\cdot 1}$ is the first column of A, i.e., the first reference learning curve. A question that needs to be examined is the correlation between $\underline{f}_{\cdot 1}$ and the row sums of Y. Since correlation is invariant under linear transformations of the data, it is simpler to use

$$\underline{f}_{\cdot 1} = Y\,\underline{a}_{\cdot 1}.$$

This set of first component factor scores is simply a linear transformation of those given above, and will have an identical correlation with the row sums of Y. If people having identical row sums (total purchases) have identical brand loyalty scores on the first component, then obviously the correlations between total purchases and the brand loyalty scores is one. Let

---

[11] The author is indebted to W. Alan Nicewander, Assistant Professor, Department of Psychology, University of Oklahoma, for suggesting this explanation for the high correlations.

$$f_{i1} = \underline{y}_{i\cdot}\, \underline{a}_{\cdot 1} \qquad (1)$$

be the brand loyalty score of the ith person on the 1st component, and

$$f_{j1} = \underline{y}_{j\cdot}\, \underline{a}_{\cdot 1} \qquad (2)$$

be the brand loyalty score for the jth person. Using these equations one can answer the questions about the correlations between total purchases and BLS's raised implicitly in Table 19.

1. Persons i and j will have identical brand loyalty scores when $\underline{y}_{j\cdot} = \underline{y}_{i\cdot}$, i.e., when they have exactly the same pattern of purchases across the n trials.

2. Persons i and j will have identical brand loyalty scores when they have the same total number of purchases across the n trials (but not necessarily the same pattern) <u>and</u> the reference curve $\underline{a}_{\cdot 1}$ is a vector of constants.

3. Persons i and j will have nearly identical brand loyalty scores when they have the same total number of purchases across trials <u>and</u> the reference curve $\underline{a}_{\cdot 1}$ is <u>nearly</u> constant.

These three statements are based on equations (1) and (2) and are fairly obvious if one remembers that $\underline{y}_{i\cdot}$ and $\underline{y}_{j\cdot}$ are vectors of ones and zeros. The vector multiplication $\underline{y}_{i\cdot}\, \underline{a}_{\cdot 1}$ simply sums the elements in $\underline{a}_{\cdot 1}$ corresponding to non-zero elements in $\underline{y}_{i\cdot}$ (or $\underline{y}_{j\cdot}$).

Case 3 discussed above explains why the total purchases of persons correlated so highly with their first brand loyalty score. The elements of the first reference curve, $\underline{a}_{\cdot 1}$, were nearly constant after the first few trials in all sets of data analyzed. Therefore, people with the same total number of purchases had nearly identical brand loyalty socres, and the correlation approached one in most sets of data.

The last data finding to be presented in this study concerns Sheth's statement:

> More important, the brand loyalty scores support the a priori, theoretically based expectation of learning brand preferences. The statistical learning theory with the traditional negatively accelerating curve states that each additional consecutive purchase of a brand will add a fraction of learning still incomplete. The less the prior learning, the greater the increment.[12]

He demonstrates the last statement by showing 3 protocols taken from one of his samples:

1	1	1	1	1	1	1
0	1	1	1	1	1	1
0	0	1	1	1	1	1

By taking the first differences of the BLS's of protocols 1, 2, and 3, he points out that the last first difference is larger than the first because the amount of learning attained was less. In other words, as the BLS's get smaller, the first differences get larger. In a similar demonstration, Howard and Sheth[13] show, with one exception in six protocols, that the magnitude of increments in BLS's are inversely related to the magnitude of prior learning. Stated in another way, as learning increases, the differences between successive BLS's decrease.

---

[12]Sheth, op. cit., p. 400

[13]Howard and Sheth, op. cit., p. 262.

If the citations above are intuitively appealing in terms of showing how BLS's can be used to reflect exponential curves (incremental learning) for individuals, then the findings of this study must be counter-intuitive. Two brief examples will illustrate why.

In the first example only the first 5 trials of 5 protocols selected from all-or-none learning data of set 3 are shown. The remaining trials are all ones.

					BLS	First Differences
1	1	1	1	1	.14476	
0	1	1	1	1	.14186	.00290
0	0	1	1	1	.13801	.00385
0	0	0	1	1	.13405	.00396
0	0	0	0	1	.12945	.00460

The same relationships are expressed here as those in Sheth's demonstration; however, it is known that the data are from all-or-none learners.

The second example, set up the same way as the first, employs incremental data from set 1.

					BLS	First Differences
1	1	1	1	1	.06559	
0	1	1	1	1	.06316	.00243
0	0	1	1	1	.06077	.00239
0	0	0	1	1	.05843	.00234
0	0	0	0	1	.05611	.00232

The outcome is the opposite of Sheth's demonstration, even though incremental learning data are used.

It is concluded, then, that BLS's are not reliable indicators of incremental learning, and therefore, that Sheth's factor analytic model for measuring brand loyalty does not identify type of learning by either reference curve or BLS analysis.

APPENDIX A

IBM 360/50 PROGRAM FOR GENERATION OF LEARNING DATA FROM THE TWO-OPERATOR LINEAR LEARNING MODEL

```
      DIMENSION PB(300,31),SUMD(30),SUMA(300),PERD(30),PERA(300)
      INTEGER TR(300,30)
      IR = 5
      IW=6
      READ(IR, 112) NSET
 112  FORMAT (I2)
      DO 200 IF =1, NSET
      READ (IR, 102) PR, A, B, B1, IX
 102  FORMAT(4F10.0,I10)
 2    DO 20 J=1, 300
      P=PR
      DO 20 I = 1, 30
      CALL RANDU(IX,IY,R)
      IX = IY
      IF(R-P) 6,6, 7
 6    TR (J, I) = 1
      P = (1.0-B)*P+A*B
      PB(J,I) = P
      GO TO 20
 7    TR(J,I) = 0
      P=(1.0-B1)*P
      PB(J,I) = P
 20   CONTINUE
      WRITE(7,121)((TR(J,I),I=1,30),J=1,300)
 121  FORMAT(30I1)
      WRITE (IW, 113) PR, A, B, IX
 113  FORMAT (1H1, 3F10.2, I15, ///)
      WRITE (IW, 103) ((TR(J,I), I=1,30), J=1,300)
 103  FORMAT (1X, 30 I4)
      WRITE (IW,104)
 104  FORMAT (1H1)
      WRITE (IW, 109) ((PB(J,I),I = 1,30), J = 1, 300)
 109  FORMAT (1X, 30F4.2)
      DO 70 I = 1, 30
 70   SUMD(I) = 0.0
      DO 71 J = 1, 300
 71   SUMA (J) = 0.0
      DO 80 I = 1, 30
      DO 80 J = 1, 300
 80   SUMD (I) = SUMD (I)+TR(J,I)
      DO 82 I = 1, 30
 82   PERD (I) = SUMD (I)/300.0
      DO 81 J = 1, 300
      DO 81 I = 1, 30
 81   SUMA (J) = SUMA (J)+TR(J,I)
      DO 83 I = 1,300
 83   PERA (I) = SUMA (I)/30.0
      WRITE (IW, 104)
      WRITE (IW,107)
 107  FORMAT (1H1, 'SUM OF ONES OF 30 COLUMNS WITH PERCENTAGES', //)
      WRITE (IW, 108) (SUMD(I), PERD (I), I = 1, 30)
 108  FORMAT (1X, 2F15.4)
      WRITE (IW, 110)
```

```
110  FORMAT (1H1, 'SUM OF ONES OF 300 ROWS WITH PERCENTAGES', //)
     WRITE (IW, 108) (SUMA (J), PERA (J), J = 1, 300)
200  CONTINUE
     CALL EXIT
     END
```

APPENDIX B

IBM 360/50 PROGRAM FOR GENERATION OF LEARNING DATA FROM THE ONE-ELEMENT LEARNING MODEL

```
      DIMENSION SUMO(30), SUMA(300),PERD (30), PERA (300)
       INTEGER TR(300,30), U
       IR = 5
       IW = 6
       IX = 427
      READ(IR,120) BU,CU,BL,CL
 120  FORMAT(4F10.0)
      DO 10 J = 1, 300
      WRITE(IW,106) J
 106  FORMAT(1H1, 'PERSON NUMBER ', I5, //)
      U=0
      DO 20 I=1, 30
C     CHECK IF U OR L
       IF (U) 3, 4, 3
C     UNLEARNED BRANCH
 4    CALL RANDU (IX, IY, R)
       IX=IY
C     CHECK IF BUY
       IF (R-BU) 6, 6, 7
C     UNLEARNED BUY
 6     TR (J, I) = 1
       GO TO 15
C     UNLEARNED BUY
 7     TR(J,I) = 0
  15  WRITE(IW,104)J,U,TR(J,I)
 104  FORMAT(1X,3I10)
C     CHECK IF CHANGE OF STATES
      CALL RANDU (IX, IY, R)
      IX=IY
      IF (R-CU) 8,8, 20
C     CHANGE TO L STATE
 8    U = 1
      GO TO 20
C     LEARNED BRANCH
 3    CALL RANDU (IX, IY, R)
      IX = IY
C     CHECK IF BUY
      IF (R-BL) 22, 22, 23
C     LEARNED BUY
 22   TR(J,I) = 1
      GO TO 24
C     LEARNED NOT BUY
 23   TR(J, I) = 0
  24  WRITE(IW,104) J,U,TR(J,I)
C     CHECK IF CHANGE STATES
      CALL RANDU (IX, IY, R)
      IX = IY
      IF (R-CL) 27, 27, 20
C     CHANGE TO U
 27   U = 0
 20   CONTINUE
      WRITE (IW, 103) (TR (J, I),I=1, 30)
 103  FORMAT(1X, 30I1)
```

```
10    CONTINUE
      WRITE(7,121)((TR(J,I),I=1,30),J=1,300)
121   FORMAT(30I1)
      DO 70 I = 1, 30
 70   SUMD (I) = 0.0
      DO 71 J = 1, 300
71    SUMA(J) = 0.0
      DO 80 I =  1, 30
      DO 80 J = 1, 300
 80   SUMD (I) = SUMD (I) + TR (J, I)
      DO 82 I = 1, 30
 82   PERD(I) = SUMD(I)/300.0
      DO 81 J = 1, 300
      DO 81 I = 1, 30
 81   SUMA (J) = SUMA (J) + TR (J, I)
      DO 83 I = 1, 300
 83   PERA(I) = SUMA(I)/30.0
      WRITE (IW, 107)
107   FORMAT (1H1, 'SUM OF ONES OF 30 COLUMNS WITH PERCENTAGES', //)
      WRITE (IW, 108) (SUMD(I), PERD (I), I = 1, 30)
108   FORMAT (1X, 2F15.4)
      WRITE (IW, 110)
110   FORMAT (1H1, 'SUM OF ONES OF 300 ROWS WITH PERCENTAGES', //)
      WRITE (IW, 108) (SUMA (J), PERA (J), J = 1, 300)
      CALL EXIT
      END
```

APPENDIX C

IBM 360/50 PROGRAM FOR COMPUTATION OF THE NORMALIZED CROSS-PRODUCTS MATRIX, A STANDARD PRINCIPAL COMPONENTS SUBROUTINE, AND FACTOR SCORES

```
      DIMENSION Y(300,30),D(30),C(30,30),A(30,30),E(30)
      DIMENSION F(300,30),V(30,30),EV(30),IQ(30),IC(30,30)
      L = 30
      NR = 200
      NC = 30
      NNV=0
      DO 1 I=1,NR
    1 READ(5,100)(Y(I,J),J=1,NC)
C     READS BASIC MATRIX OF ONES AND ZEROS
  100 FORMAT(30F1.0)
      DO 20 J=1,NC
      SUM=0.0
      DO 2 I=1,NR
    2 SUM=SUM+Y(I,J)
   20 D(J)=SUM
C     SUMS COLUMNS OF Y
      DO 3 I=1,NR
      DO 3 J=1,NC
    3 Y(I,J)=Y(I,J)*(1.0/SQRT(D(J)))
C     FORMS Y D-1/2 - STORED IN Y
      WRITE(6,101)
  101 FORMAT(1H1, 60X,13HCOLUMN TOTALS)
      WRITE(6,102) (D(J),J=1,NC)
  102 FORMAT(1H , 15F8.3)
      DO 5 I=1,NC
      DO 5 J=1,NC
      C(I,J)=0.0
      DO 5 K=1,NR
    5 C(I,J)=C(I,J)+Y(K,I)*Y(K,J)
C     MATRIX C IS D(-1/2) Y' Y D(-1/2)
      WRITE(6,103)
  103 FORMAT(1H1,50X,25HNORMALIZED CROSS PRODUCTS)
      DO 75 I=1,NC
      DO 75 J=1,NC
   75 IC(I,J)=1000.*C(I,J)+0.5
      DO 4 I=1,NC
    4 WRITE(6,108) (IC(I,J),J=1,NC)
  108 FORMAT(1H ,30I4)
      CALL JAC (NC,C,V,EV,IC)
C     V CONTAINS EIGENVECTORS,EV IS THE VECTOR OF EIGENVALUES
      DO 11 I=1,NC
      IF(EV(I))11,11,40
   40 NNV=NNV+1
   11 CONTINUE
      WRITE(6,107)
  107 FORMAT(1H1, 60X,11HEIGENVALUES)
      WRITE(6,200)(EV(I),I=1,NNV)
  200 FORMAT(1H , 15F8.4)
C     NNV = NO. OF POSITIVE EIGENVALUES
      DO 6 J = 1,NNV
      DO 6 I = 1,NC
    6 V(I,J)=V(I,J)*SQRT(EV(J))
C     V NOW CONTAINS FACTOR LOADINGS
```

```
      WRITE (6, 104)
  104 FORMAT(1H1,50X, 15HFACTOR LOADINGS)
      DO 7 I=1,NC
    7 WRITE(6,105)I,(V(I,J),J=1,13)
  105 FORMAT(1H0,I3,4X,13F8.5)
      DO 8 I=1,NR
      DO 8 J=1,NNV
      F(I,J) = 0.0
      DO 8 K = 1,NC
    8 F(I,J)=F(I,J)+Y(I,K)*V(K,J)
C     COMPUTES       F=YD-1/2*V
      DO 9 J=1,NNV
      DO 9 I=1,NR
    9 F(I,J)=F(I,J)*1.0/EV(J)
C     F CONTAINS FACTOR SCORES NORMALIZED TO UNIT LENGTH.
      WRITE(6,106)
  106 FORMAT(1H1,50X, 13HFACTOR SCORES)
      DO 10 I=1,NR
   10 WRITE(6,105)I,(F(I,J),J=1,13)
      CALL EXIT
      END
```

```
      SUBROUTINE JAC (N,H,U,X,IQ)
      DIMENSION H(30,30),U(30,30),X(30),IQ(30)
10    DO  14 I=1,N
      DO  14 J=1,N
      IF(I-J)12,11,12
11    U(I,J)=1.0
      GO TO 14
12    U(I,J)=0.0
14    CONTINUE
15    NR=0
      IF(N-1) 1000,1000,17
17    NMI1=N-1
      DO 30 I=1,NMI1
      X(I)=0.0
      IPL1=I+1
      DO  30 J=IPL1,N
      IF(X(I)-ABS (H(I,J))) 20,20,30
20    X(I)=ABS (H(I,J))
      IQ(I)=J
30    CONTINUE
      RAP=7.450580596E-9
      HDTEST=1.0E38
40    DO 70 I=1,NMI1
      IF(I-1) 60,60,45
45    IF(XMAX-X(I)) 60,70,70
60    XMAX=X(I)
      IPIV=I
      JPIV=IQ(I)
70    CONTINUE
      IF(XMAX) 1000,1000,80
80    IF(HDTEST) 90,90,85
85    IF(XMAX-HDTEST) 90,90,148
90    HDIMIN=ABS (H(1,1))
      DO 110 I=2,N
      IF(HDIMIN-ABS (H(I,I))) 110,110,100
100   HDIMIN=ABS (H(I,I))
110   CONTINUE
      HDTEST=HDIMIN*RAP
      IF(HDTEST-XMAX) 148,1000,1000
148   NR=NR+1
150   TANG=SIGN (2.0,(H(IPIV,IPIV)-H(JPIV,JPIV)))*H(IPIV,JPIV)/(ABS (H(
     1IPIV,IPIV)-H(JPIV,JPIV))+SQRT ((H(IPIV,IPIV)-H(JPIV,JPIV))**2+4.0*H
     2(IPIV,JPIV)**2))
      COSINE=1.0/SQRT (1.0+TANG**2)
      SINE=TANG*COSINE
      HII=H(IPIV,IPIV)
      H(IPIV,IPIV)=COSINE**2*(HII+TANG*(2.*H(IPIV,JPIV)+TANG*H(JPIV,JPIV
     1)))
      H(JPIV,JPIV)=COSINE**2*(H(JPIV,JPIV)-TANG*(2.*H(IPIV,JPIV)-TANG*H
     1II))
      H(IPIV,JPIV)=0.
      IF(H(IPIV,IPIV)-H(JPIV,JPIV)) 152,153,153
152   HTEMP=H(IPIV,IPIV)
```

```
      H(IPIV,IPIV)=H(JPIV,JPIV)
      H(JPIV,JPIV)=HTEMP
      HTEMP=SIGN (1.0,-SINE)*COSINE
      COSINE=ABS (SINE)
      SINE=HTEMP
153   CONTINUE
      DO 350 I=1,NMI1
      IF(I-IPIV) 210,350,200
200   IF(I-JPIV) 210,350,210
210   IF(IQ(I)-IPIV) 230,240,230
230   IF(IQ(I)-JPIV) 350,240,350
240   K=IQ(I)
250   HTEMP=H(I,K)
      H(I,K)=0.0
      IPL1=I+1
      X(I)=0.0
      DO 320 J=IPL1,N
      IF(X(I)-ABS (H(I,J))) 300,300,320
300   X(I)=ABS (H(I,J))
      IQ(I)=J
320   CONTINUE
      H(I,K)=HTEMP
350   CONTINUE
      X(IPIV)=0.0
      X(JPIV)=0.0
      DO 530 I=1,N
      IF(I-IPIV) 370,530,420
370   HTEMP=H(I,IPIV)
      H(I,IPIV)=COSINE*HTEMP+SINE*H(I,JPIV)
      IF(X(I)-ABS (H(I,IPIV))) 380,390,390
380   X(I)=ABS (H(I,IPIV))
      IQ(I)=IPIV
390   H(I,JPIV)=-SINE*HTEMP+COSINE*H(I,JPIV)
      IF(X(I)-ABS (H(I,JPIV))) 400,530,530
400   X(I)=ABS (H(I,JPIV))
      IQ(I)=JPIV
      GO TO 530
420   IF(I-JPIV) 430,530,480
430   HTEMP=H(IPIV,I)
      H(IPIV,I)=COSINE*HTEMP+SINE*H(I,JPIV)
      IF(X(IPIV)-ABS (H(IPIV,I))) 440,450,450
440   X(IPIV)=ABS (H(IPIV,I))
      IQ(IPIV)=I
450   H(I,JPIV)=-SINE*HTEMP+COSINE*H(I,JPIV)
      IF(X(I)-ABS (H(I,JPIV))) 400,530,530
480   HTEMP= H(IPIV,I)
      H(IPIV,I)=COSINE*HTEMP+SINE*H(JPIV,I)
      IF(X(IPIV)-ABS (H(IPIV,I))) 490,500,500
490   X(IPIV)=ABS (H(IPIV,I))
      IQ(IPIV)=I
500   H(JPIV,I)=-SINE*HTEMP+COSINE*H(JPIV,I)
      IF(X(JPIV)-ABS (H(JPIV,I)))510,530,530
510   X(JPIV)=ABS (H(JPIV,I))
```

```
      IQ(JPIV)=I
530   CONTINUE
540   DO 550 I=1,N
      HTEMP=U(I,IPIV)
      U(I,IPIV)=COSINE*HTEMP+SINE*U(I,JPIV)
550   U(I,JPIV)=-SINE*HTEMP+COSINE*U(I,JPIV)
      GO TO 40
1000  DO 600 I=1,N
      X(I)=H(I,I)
600   CONTINUE
      RETURN
      END
```

BIBLIOGRAPHY

Books

Atkinson, Richard C,; Bower, Gordon H.; and Crothers, Edward J. An Introduction to Mathematical Learning Theory. New York: John Wiley and Sons, Inc., 1955.

Bugelski, B. R. Psychology of Learning, New York: Henry Holt and Company, Inc., 1956.

Bush, Robert R., and Mosteller, Frederick. Stochastic Models for Learning. New York: John Wiley and Sons, Inc., 1955.

Cooley, W. W., and Lohnes, P. R. Multivariate Procedures for Behavioral Sciences. New York: John Wiley and Sons, Inc., 1962.

Coombs, Clyde H.; Dawes, Robyn M.; and Tversky, Amos. Mathematical Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Dukta, Solomon, and Frankel, Lester. Markov Chain Analysis: A New Tool of Marketers. New York: Audits and Surveys Company, Inc., 1962.

Engel, James F.; Kollat, David T.; and Blackwell, Roger D. Consumer Behavior. New York: Holt, Rinehart and Winston, Inc., 1968.

Fourt, Louis A. Applying Markov Chain Analysis to NCP Brand-Switching Data. Chicago: Market Research Corporation of America.

Fruchter, Benjamin. Introduction to Factor Analysis. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1954.

Gilbert, Eugene. Advertising and Marketing Young People. Pleasantville, New York: Printer's Ink Books, 1957.

Green, Paul E., and Tull, Donald S. Research for Marketing Decisions. 2d ed. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1970.

Harmon, H. H. Modern Factor Analysis. 2d ed. Chicago: University of Chicago Press, 1960.

Hilgard, Ernest R., and Bower, Gordon H. Theories of Learning. 3rd ed. New York: Appleton-Century-Crofts, 1966.

Horst, Paul. Factor Analysis of Data Matrices. New York: Holt, Rinehart and Winston, Inc., 1965.

Howard, John A. Marketing: Executive and Buyer Behavior. New York: Columbia University Press, 1963.

Howard, John A. Marketing Management: Analysis and Planning. Homewood, Illinois: Richard D. Irwin, Inc., 1963.

Howard, John A., and Sheth, Jagdish N. The Theory of Buyer Behavior New York: John Wiley and Sons, 1969.

Hull, Clark L. A Behavior System. New Haven: Yale University Press, 1952.

Hull, Clark L. Principles of Behavior: An Introduction to Behavior Theory. New York: Appleton-Century-Crofts, 1943.

Kemeny, J. G., and Snell, J. L. Finite Markov Chains. Princeton, New Jersey: D. Van Nostrand Company, 1957.

Kendall, M. G. A Course in Multivariate Analysis, 2d ed. New York: Hartner Publishing Company, 1965.

Lashley, K. S. Brain Mechanisms and Behavior. Chicago: University of Chicago Press, 1928.

Newell, A., and Simon, H. A. "Computers in Psychology," in Luce, R. D.; Bush, R. R.; and Galanter, E. eds. Handbook of Mathematical Psychology. New York: Wiley, 1963.

Restle, Frank, and Greeno, James G. Introduction to Mathematical Psychology. Reading, Massachusetts: Addison-Wesley Publishing Company, 1970.

Thurstone, L. L. Multiple Factor Analysis. Chicago: University of Chicago Press, 1947.

Ledyard R. Tucker. "Learning Theory and Multivariate Experiment: Illustration by Determination of Generalized Learning Curves," in Cattell, R. B. ed. Handbook of Multivariate Experimental Psychology. Rand McNally, 1966.

Wells, William D. "Children as Consumers," in Joseph Newman ed. On Knowing the Consumer. New York: John Wiley & Sons, 1966.

Journals

Bush, Robert R., and Mosteller, Frederick. "A Mathematical Model for Simple Learning." Psychological Review, Vol. 58 (1951), pp. 313-323.

Draper, Jean E., and Nolin, Lassy H. "A Markov Chain Analysis of Brand References," Journal of Advertising Research, Vol. 4 (September, 1964), pp. 33-39.

Estes, W. K., and Burke, C. J. "A Theory of Stimulus Variability in Learning." Psychological Review, Vol, 60 (1953), pp. 276-286.

Estes, W. K. "Toward a Statistical Theory of Learning." Psychological Review, Vol. 57 (1950), pp. 94-107.

Farley, John U., and Ring, Winston L. "An Empirical Test of the Howard-Sheth Model of Buyer Behavior." Journal of Marketing Research, (November, 1970), pp. 427-438.

Guest, Lester. "Genesis of Brand Awareness." Journal of Applied Psychology, Vol. 26 (1942), pp. 800-808.

Guthrie, Edwin R. "Psychological Facts and Psychological Theory." Psychological Bulletin, Vol. 43 (1946), pp. 1-20.

Harary, Frank, and Lipstein, Benjamin. "The Dynamics of Brand Loyalty: A Markovian Approach." Operations Research, Vol. 10(January-February, 1962), pp. 19-40.

Hayes, C. L. "The Backward Curve: A Method for the Study of Learning." Psychological Review, Vol. 60 (1953), pp. 269-275.

Herniter, Jerome D., and Magee, John F. "Customer Behavior as a Markov Process." Operations Research, Vol. 9 (January-February, 1961), pp. 105-122.

Kemeny, J. G. and Snell, J. L. "Markov Processes in Learning Theory." Psychometrika, Vol. 22, pp. 221-230.

Krugman, Herbert E. "Processes Underlying Exposure to Advertising." American Psychologist, Vol. 23 (1968), pp. 245-253.

Krugman, Herbert E. "The Impact of T.V. Advertising: Learning Without Involvement." Public Opinion Quarterly, Vol. 29 (Fall, 1965), pp. 349-356.

Laughery, K. R., and Gregg, L. W. "Simulation of Human Problem-Solving Behavior." Psychometrika, Vol. 27 (1962), pp. 265-282.

Maffei, Richard B. "Brand Preference and Simple Markov Processes." Operations Research, Vol. 8 (March-April, 1960), pp. 210-218.

Merrill, M. "The Relationship of Individual Growth to Average Growth." Human Biology, Vol. 3 (1931), pp. 37-40.

Ross, John. "Mean Performance and the Factor Analysis of Learning Data." Psychometrika, Vol. 29 (1964), pp. 67-73.

Sheth, Jagdish N. "A Factor Analytical Model of Brand Loyalty." Journal of Marketing Research, (November, 1968), pp. 395-404.

Sheth, Jagdish N. "Learning of Brand Preference at the Adult Age." Journal of Advertising Research, (September, 1968), pp. 25-36.

Sheth, Jagdish N. "The Multivariate Revolution in Marketing Research." Journal of Marketing, (January, 1971), pp. 13-19.

Styan, George, and Smith, Harry, Jr. "Markov Chains Applied to Marketing." Journal of Marketing Research, Vol. 1 (February, 1964), pp. 50-55.

Tucker, Ledyard R. "Determination of Parameters of a Functional Relation by Factor Analysis." Psychometrika, Vol. 23 (1958), pp. 19-23.

Weedlock, J. W. "A Clue to Purchase Patterns - Markov's Mathemagic." Sales Management, Vol. 93 (September 18, 1964), pp. 71-72.

Weitzman, R. A. "A Factor Analytic Method for Investigating Differences Between Groups of Individual Learning Curves." Psychometrika, Vol. 28 (1963), pp. 69-80.

Other

"A New Theory on Why People Buy." Business Week, (January, 1970), p. 70.

"Fall Conference, American Marketing Association." The Marketing News, First-of-July, 1971, pp. 4-10.

Day, George S. "Buyer Attitudes and Brand Choice Behavior." Unpublished doctoral dissertation, Columbia University, 1967.

Perry, Michael. "Computer Simulation of Consumer Brand Choice." Unpublished doctoral dissertation, Columbia University, 1968.

Professional Consultation with W. Alan Nicewander, Assistant Professor, Department of Psychology, University of Oklahoma, August, 1971.

Schlomo, Lamport. "Word-of-Mouth Activity During the Introduction of an New Product." Unpublished doctoral dissertation, Columbia University, 1969.

O'Brian, Terrence U. "Information Sensitivity and the Sequence of Psychological States in Brand Choice Process." Unpublished doctoral dissertation, Columbia University, 1969.

Sheth, Jagdish N. "A Behavioral and Quantitative Investigation of Brand Loyalty." Unpublished doctoral dissertation, University of Pittsburgh, 1966.