A System of Evaluation for Design Based on the Principles of Product Semantics

A Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of the Ohio State University

by

Prasad Boradkar

The Ohio State University
1993

Master's Examination Committee:
Prof. Reinhart Butter
Prof. In Jae Myung

Approved by
Advisor
Department of Industrial Design
ACKNOWLEDGEMENTS

I would like to thank my advisor Prof. Reinhart Butter, and my co-advisor Prof. In Jae Myung, for their invaluable guidance and support.

I would also like to thank Prof. Klaus Krippendorff from University of Pennsylvania, Mr. Scott Belliveau from NCR Corporation, and Mr. Louis Blatt from NCR Corporation, for their help.

Thanks are due to Ms. Jessica Dulin for her help with the statistical analysis.
VITA

January 21, 1965

1986

Born - Nagpur, India

Bachelor of Mechanical Engineering,
M. S. University of Baroda, India

January 1987-July 1987

Graduate Engineer Trainee,
Godrej and Boyce Manufacturing Co.
Bombay, India

1989

Master of Design,
Indian Institute of Technology, India

1989-1990

Industrial Designer
Roplas India Ltd.,
Chinchwad, India

1990-1991

Industrial Designer
Bajaj Auto Ltd.,
Akurdi, India

February 1992-May 1992

Research Fellow
Technische Universiteit, Delft
Delft, The Netherlands

September 1992-present

Graduate Associate
Wexner Center for the Arts
Columbus, Ohio

FIELDS OF STUDY

Major Field: Industrial Design
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS** ........................................ ii

**VITA** .......................................................... iii

**LIST OF TABLES** ............................................. vii

**LIST OF FIGURES** ........................................... viii

## CHAPTERS

<table>
<thead>
<tr>
<th>I. Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>..........................</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. On Product Semantics</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Basics.</td>
<td>6</td>
</tr>
<tr>
<td>Understanding</td>
<td>6</td>
</tr>
<tr>
<td>Function</td>
<td>9</td>
</tr>
<tr>
<td>Metaphors in Design</td>
<td>10</td>
</tr>
<tr>
<td>The Role of Language in Design</td>
<td>10</td>
</tr>
<tr>
<td>Affordances</td>
<td>11</td>
</tr>
<tr>
<td>Information Theory</td>
<td>12</td>
</tr>
<tr>
<td>Application to Design</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. On Evaluation</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Role of Evaluation in Design</td>
<td>16</td>
</tr>
<tr>
<td>The Design Process</td>
<td>16</td>
</tr>
<tr>
<td>The Design Process at NCR</td>
<td>19</td>
</tr>
<tr>
<td>Evaluation in Education</td>
<td>21</td>
</tr>
</tbody>
</table>
Evaluation Strategy .......................... 28

IV. The Design Project. ......................... 31
  Design Methodology ......................... 31
  Metaphors .................................. 33
  On Modeling in Design ...................... 37
  Modelmaking Media and Technology ...... 40
  Virtual Modeling ............................ 41
  Design, Evaluation and Concurrent Engineering ... 43
  Computers in Design and Evaluation ....... 44

V. Evaluation Parameters ....................... 53
  Four Design Theories ....................... 53
  The Context of Use ......................... 54
  The Context of Language ................... 64
  The Context of Genesis ..................... 66
  The Context of Ecology ..................... 68
  Aggregation of Individual Evaluations .... 71

VI. The Testing Scenario ......................... 72
    Subjects .................................. 72
    Materials ................................. 73
    Procedure ................................ 73

VII. Methods of Data Analysis ................. 83
    Hypothesis Testing ....................... 83
    Multidimensional Scaling ................ 86
    Verbal Reports and Protocol Analysis ... 89
    Multi Attribute Utility Technology ...... 94
    Geography and Cartographic Analysis ... 95
    Analysis of Water Color Painting ......... 97
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIII. Analysis and Conclusions</td>
<td>99</td>
</tr>
<tr>
<td>Hypothesis Testing</td>
<td>100</td>
</tr>
<tr>
<td>Multidimensional Scaling</td>
<td>111</td>
</tr>
<tr>
<td>Multi Attribute Utility Technology</td>
<td>114</td>
</tr>
<tr>
<td>IX. Future Research</td>
<td>120</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>122</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>140</td>
</tr>
</tbody>
</table>
LIST OF TABLES

TABLES

1. Evaluation Matrix ........................................... 30
2. Attributes for Laptop Computers .......................... 33
3. Attributes and Medium of Presentation Matrix ........ 70
4. Average Ranks on the Six Attributes Tested .......... 125
5. Average Values on the Six Dissimilarities ............. 126
6. Target Group Compatibilities .............................. 127
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Roots of the System of Evaluation</td>
<td>2</td>
</tr>
<tr>
<td>2. Components of Traditional Semiotic Theory</td>
<td>7</td>
</tr>
<tr>
<td>3. Meanings and Contexts</td>
<td>9</td>
</tr>
<tr>
<td>4. Design Development Spiral</td>
<td>18</td>
</tr>
<tr>
<td>5. Laptop Computer Showing Metaphor of File Folder and Ring Binder</td>
<td>34</td>
</tr>
<tr>
<td>6. Computer for Young Professionals, College Students, and High School Kids (from top to bottom)</td>
<td>36</td>
</tr>
<tr>
<td>7. Modelmaking Process</td>
<td>46</td>
</tr>
<tr>
<td>8. Computer Generated Image of the Power Switch</td>
<td>47</td>
</tr>
<tr>
<td>9. Computer Generated Image of the Keyboard</td>
<td>47</td>
</tr>
<tr>
<td>10. Computer Generated Image of the Track Ball</td>
<td>48</td>
</tr>
<tr>
<td>11. Computer Generated Images of Grills for the Young Professionals, College Students, and High School Kids (from left to right)</td>
<td>48</td>
</tr>
<tr>
<td>12. Computer for Young Professionals (closed)</td>
<td>49</td>
</tr>
<tr>
<td>13. Computer for Young Professionals (open)</td>
<td>49</td>
</tr>
<tr>
<td>14. Computer for College Students (closed)</td>
<td>50</td>
</tr>
<tr>
<td>15. Computer for College Students (open)</td>
<td>50</td>
</tr>
</tbody>
</table>

viii
CHAPTER I
INTRODUCTION

Product semantic theory, initiated in 1984 by Krippendorff and Butter (Krippendorff and Butter, 1984) has since its conception dealt with various issues related to the symbolic qualities of objects in use. The body of knowledge has grown through contributions from active researchers across the world, creating for the discipline of product semantics an invaluable place in the profession of industrial design. Through its intrinsic attachment to industrial design, product semantics has bestowed upon it a uniqueness that has helped design to distinguish its empirical domain from other professions. Industrial designers now have an opportunity to develop a language of their own based on the observation that human beings perceive not pure forms but meanings.

Over and above a scrutiny of the issue of meaning, the theory of product semantics has helped formulate a design methodology to aid the creation of products that make sense to human beings and facilitate use. However, for its full development, this emergent discipline of product semantics requires a systematic survey of evaluation techniques, and the consolidation of an academic underpinning in evaluation theory. The subject of evaluation has not been sufficiently dealt with within the literature on product semantics. Research in semantics and in the theory of evaluation have progressed independently, but
little effort has been done toward combining the two. It is essential to tap the information available in evaluation research and develop a means for evaluating design based on the principles of product semantics. This thesis attempts to accomplish precisely that.

![Diagram of System of Evaluation](image)

**Fig. 1. Roots of the System of Evaluation**

Having stated the need for a means of evaluation, an outline needs to be drawn up specifying the structure of such an evaluation. A number of questions are posed in an attempt to define the structure of the evaluation:

What is the purpose of the evaluation?
What does it seek to evaluate?
How should the evaluation be conducted?
Who should conduct the evaluation?
At which stage in the product development cycle should the evaluation be conducted?
What should be the final product of the evaluation?

The ensuing discussion addresses these concerns.

The purpose is to develop a means of evaluating one or a group of artifacts for their ability to make sense to human beings. Meanings arise in interaction, and the evaluation should test the efficacy of the interaction afforded by the artifacts under scrutiny.

Evaluation of meaning will essentially involve a study and testing of those attributes of artifacts that are responsible for encouraging this human-artifact interaction. These attributes will be extracted from the existing literature on product semantics. Krippendorff has defined four essential contexts in which designed objects need to survive in the face of existing practices. These contexts, namely that of use, language, genesis and ecology form the platform on which the evaluation will be based. It is realized that meaning of an artifact cannot be evaluated by using one single method of evaluation. There are various different issues involved and what is needed is a ‘system of evaluation’ consisting of several different techniques appropriate to the attributes being investigated.

As product semantics seeks to understand users’ understanding of the artifacts in their world, any evaluation should necessarily be centered around users. The term user is interpreted in a fairly general way to include a wide range of interpretations. Typically, ‘user’ refers to the person who physically uses the artifact. However, meanings of artifacts do not depend solely upon the physical user but on other ‘participants’ as well-anyone who in any way interacts with the artifact influences it. Ideally all these participants should be involved in the
evaluation.

This evaluation will be structured and coordinated by an individual, who could be an industrial designer, a marketing expert, and finance expert—anyone who is involved in the product development process. Participation of a variety of individuals is essential for a semantic evaluation of artifacts, and Chapter V elucidates how these people are involved in the evaluation.

Product evaluation, as it is practiced today by groups such as Consumer Reports is conducted with products available on the market. This does not give designers an opportunity to modify their designs unless a new improved model is introduced. Industrial design groups need evaluation that can be conducted during the course of the design process, so that the feedback can be effectively used before the product is manufactured. It is hence essential to develop methods to evaluate designs without having a working product available. Evaluation should be possible with the help of sketches, renderings, models, prototypes, etc.

The final product of the evaluation should be a judgment of a single product or a comparative judgment of a group of products that will aid in the decision of choosing one over the others for being semantically more appropriate. The requirement of an outcome of this type of evaluation was discussed with designers in industry to get a better perspective of what exactly was desired by the professionals in the field. Scott Belliveau of NCR Corporation, Dayton, Ohio, believes that such an evaluation would help forward the cause of design if it could provide numerical results. Quantitative data is often an excellent tool for convincing ‘non-designers’ of the feasibility of a certain design solution. A product development manager would see reason if the evaluation
could, with valid justification say that from the point of view of semantics, Product A scores 7.3 while Product B scores 5.2 on a scale of 10. Quantification of abstract qualitative impressions and feelings is a non-trivial task and presents the risk of losing the richness of data in transformation from the un-measurable to the measurable. However, it brings ease into the process of decision making, makes objective what is ordinarily considered subjective, and gives designers a language that can be comprehended by management. It may not be possible to evaluate all the attributes of an artifact relevant to semantics to give a quantitative result. Certain attributes may need qualitative techniques and may have a qualitative output. Therefore the evaluation will be part quantitative, with qualitative information to back it up.

This thesis document deals with three major issues. The first, formulation of a means of evaluation for product design founded in semantics, second, the design of a product that can be used as a vehicle in an actual evaluation, and third, the testing done with the designs generated.
CHAPTER II
ON PRODUCT SEMANTICS

The Basics

"Product Semantics is the study of symbolic qualities of man-made forms in the context of their use and the application of this knowledge to industrial design" (Krippendorff & Butter, 1984). This was one of the earliest definitions of product semantics, and over the years a fair amount of research has been done by designers, psychologists, communication experts, architects, and other professionals to add to the body of knowledge which has proven to be an extremely relevant source of inspiration to industrial design.

We human beings surround ourselves with an extraordinarily large number of objects or things that play a significant role in our day to day lives. Though referred to as ‘consumer goods’, these everyday things are not consumed but held close to oneself as expressions of personal identity that make our lives richer in meaning. It is the job of the designer to understand how users interact with objects, and use this information in the practice of design. We live in the midst of products that make sense to us, about what they are, what they are to be used for, how, when, and in what context they are to be used. Objects are perceived not as pure forms, but as meanings. We are concerned with what the object is, what it is made for, what it does, how it’s parts are connected
together, how and in what situations it is to be used, what it is made of, and so on. In case of a personal object, issues like who gave it, how it blends with other personal objects, how it feels, etc. assume importance as well. It is suggested that an object is what it means, and a distinction, if any, between the two is not perceptually demonstrable (Krippendorff, 1989).

The three major components of traditional semantic theory, namely sign, referent, and thought are defined as something that is intended to represent, something that is represented, and someone that makes the connection. Product semantic theory proposes that the sign and the referent are in fact one, as the object is what it means to the user, and therefore the triangular relationship becomes a circular one between the object and its user. Objects are always seen in a context of users, situations, other objects, and so on. Over and above the physical functions, objects exist in a social, psychological, and symbolic environment.

![Diagram](image)

**Fig. 2. Components of Traditional Semiotic Theory**
Meaning is a cognitively constructed relationship. It selectively connects features of an object and features into a coherent unity. This context into which people place the object is cognitively constructed. Meaning and sense are two distinct manifestations of experiencing things. Something makes sense when we understand the role it plays in a particular context, when we have a satisfactory explanation of what it does. On the other hand, the meaning of an object is the sum total of all the contexts for which someone is capable of imagining some sense for it. A hat, for example, has numerous uses; it can protect one's ears from the cold, it can be used to give oneself a certain image, it can express one's loyalty to a certain sports group, it can be worn to express royalty, or it can be used to pull rabbits out of. All possible contexts define what a hat is to people capable of using their imagination. Something means its possible contexts of use. One can therefore say, by analogy, that sense is to meaning as actuality is to potentiality. Meaning is achieved by grouping the various senses one can make of something in different environments into a recognizable class. Krippendorff suggests that semantics, as a theoretical discourse on meaning should be embedded in human understanding (Krippendorff, 1991). Human beings should be necessarily constitutive elements in the worlds constructed in semantic theory. Understanding is a process that arises within human cognition and is personal....

**Understanding Understanding**

Form follows meaning. Not function. This dictum brings the user into the picture and suggests that designers need to discuss how artifacts make sense to anyone other than themselves. It is preposterous to assume that form (designers' meaning) and meaning (users' meaning) are the same. Product semantics
studies how these two (the designers' cognition and the users' cognition) relate.

Product semantics is concerned with human interfaces. It deals with that layer of human cognition in which we experience how we interact with our environment, and make things understandable. Product semantics seeks to understand users' understanding of their practice of interfacing with designed objects. There are many interfaces between humans and an artifact as meanings are not universal, shared, or mappable onto particular forms. Designers must be aware of the knowledge people bring to their interfaces and understand their understanding.

Fig. 3. Meanings and Contexts

Function

In the light of product semantics, the concept of function or functionalism can be comprehended better. A function is a part-whole relationship intended to account for what something does or is to do when put in its appropriate place within a system whose purpose it is to serve. Functionalism assumes the
existence of a unique, stable, coherent, and well understood system whose physical presence is taken as proof of its purposiveness. All working parts in the system are subordinate to its purpose, each with unique functions that help the maintenance of the whole. In design, the dictum *Form Follows Function* expresses that an examination of the system will naturally lead to its appropriate form.

**Metaphors in Design**

Metaphors are mental operations that transform meanings from a familiar domain to a relatively unknown or new domain. In this process they help human beings in understanding new artifacts. For metaphors to work, people must be able to see some correspondence between the source and target domains. The elements of the target domain are often reorganized in terms of structure of the source domain. Metaphors assume that meanings are thing-like entities. It is also assumed that people will be able to extract from the messages the same meaning that was put into it by the sender.

In design, the use of metaphors can help users understand and interact more effectively with artifacts they are unfamiliar with. However, metaphors have to be self-evident in order to permit this. Otherwise they can be only confounding.

**The Role of Language in Design**

Language plays an important role in the process of design as well as in the relationship between artifacts and their users. Language is not merely descriptive or instructive, but it also be constitutive of social reality. Declarations, advertisements, rumors and all other means of communication that involve the use of language influence individual perceptions and thus meanings of artifacts
cannot be shared. Language is an essential communicating tool in the process of design. The design starts with the brief, which is in linguistic form, either as a write up of clients’ needs, the contract, the terms and conditions, design specifications, etc. All ensuing discussions about the new design are carried out through the use of language. What cannot be described would be incredibly difficult to design and just as difficult to manufacture.

Language also gives designers a means of accessing users in order to understand their understanding of the environment. Talking to users and listening to them is a source of invaluable information about how people interact with their artifacts. Techniques like content analysis and protocol analysis are means to understand people’s perceptions by having them talk about their actions. Users create cognitive models on the basis of the different experiences they have in using things, and the models guide their actions. They help in developing some kind of operational logic about why things work the way they do, and how to use them. Verbal reports can give designers an insight into the user’s mind and help create better interfaces.

**Affordances**

An interesting notion referred to as ‘affordance’ was suggested by American psychologist James Gibson in the fifties. He proposed that perception was behavioral, and what one perceives are not objects but affordances, i.e. one perceives not an object ‘chair’, but something that affords sitting. The more tired one is, the more objects appear ‘chair-like’. An extremely tired individual perceives even an orange crate or the floor as a something that he can sit on. This theory of perception is also referred to as the ecological theory of perception, and suggests that behavioral meanings are perceived in the patterns
of energy that fall on our senses. Meanings are not attached to geometrical entities seen by the observer, but are picked up by the invariant patterns in the energy by our senses (Smets, 1989).

**Information Theory**

Human beings have the unique capability of developing symbols in the brain which are representative of objects and events in the outside world. These symbols are organized on the basis of a strategy developed using semantic coding. This coding is derived from sensory information which carries the primary visual component of the object or event perceived. A better understanding of how the coding process extracts semantic information from the world would be invaluable information to designers dealing with the issue of meaning. As a part of the natural process of categorization, we human beings assemble our experiences into categories at various levels of abstraction, labels these categories with lexical terms and uses these terms effectively in communication and thinking. This labeling derives cues from the object form, the specifics of which determine the category the object falls into. The mind categorizes objects on the basis of formal similarities neglecting the infinite variations when irrelevant to human response. Objects within a category are fuzzy; and variations in the product form are often subtle enough to elude categorization. A category has a central member or a typical example, which semantic representative of the core meaning of the category. A category is defined, not by its boundaries but by its center. This raises significant questions with respect to generation of new product forms. Should typicality be maintained in new product forms to suggest category belongingness or should design be innovative and disregard these issues? Athavankar suggests that
designers need to strike a balance between typicality and innovation, category belongingness and individual choices (Athavankar, 1989).

One can deduce from these discussions that meaning is not something that can be attached to an object as an external entity. Meanings arise in our interaction with the things we surround ourselves with. Industrial designers, being creators of objects, need to be sensitive to this complex nature of meaning, and should incorporate this sensitivity in their designs.

**Application to Design**

Practical application of theory is never an easy task, but researchers have outlined systematic approaches to the incorporation of semantics in the process of design. Butter has established a heuristic sequence of steps to facilitate the process of generating product forms with predetermined meanings (Butter, 1989). These steps are:

Step 1. Establishment of general objectives and constraints for the product or system.

Step 2. Identification of the product’s projected context of use covering user groups and the system’s semantic performance characteristics.

Step 3. Generation of a list of desired attributes expressing the projected semantic performance characteristics.

Step 4. Generation of a list of undesired attributes expressing those semantic characteristics to be definitely avoided.

Step 5. Analysis, grouping, and ranking of all attributes.

Step 6. Search for concrete manifestations capable of supporting desired attributes and contrasting with undesirable attributes.
Step 7. Assessing, selecting, and integrating semantically feasible manifestations into expressive wholes.

Step 8. Evaluation of compatibilities and technical feasibilities of the ideas.

This process has shown remarkable results, generated strikingly innovative forms, and created objects that users can communicate with symbolically. The success of product semantics as a discipline of industrial design is exemplified in the products designed using the methodology mentioned earlier.
CHAPTER III
ON EVALUATION

In the beginning God created the heaven and earth...
And God saw every thing that He made.
"Behold," God said, "it is very good."
And the evening and the morning were the sixth day.
And on the seventh day God rested from all His work.

His archangel came then unto Him asking:
"God, how do you know that what you have created is 'very good'?
What are your criteria?
On what data do you base your judgment?
Aren't you a little close to the situation
to make a fair and unbiased evaluation?"

God thought about these questions all that day
and his rest was greatly disturbed.
On the eighth day God said, "Lucifer go to hell."

Thus was evaluation born in a blaze of glory.
Ever since the status of the profession has been somewhat in doubt:
the road to salvation or a sure ticket to damnation?

- The Real Story of Paradise Lost (Patton, 1981)
The Role of Evaluation in Design

A study of the process of design is essential to understand the role played by evaluation in the sequence of activities. Described below is the generic process of design, followed by the corporate methodology adopted at National Cash Register (NCR), Dayton, Ohio.

The Design Process

The process of design involves, according to Zeisel, three analytically distinct elementary activities, namely imaging, presenting, and testing (Zeisel, 1988). Imaging is defined as forming a general, sometimes fuzzy, mental picture of a part of a model of the world. This process is a search for solutions, a kind of problem-solving, which in case of design is often visual and is also referred to as conceptualization. These images and ideas have to be communicated, and this is achieved through various media available to designers. In physical form, these image representations could be,

1. Sketches, renderings, drawings (Two dimensional representations)
2. Models (Three dimensional representations)
3. Computer simulated images, models
4. Prototypes

Sketches, drawings and renderings are two dimensional representations of three dimensional objects and therefore present limited information to the observer. Numerous techniques of representation are employed by designers and they offer varying degrees of information depending upon the amount of detail expressed. Sketches are rough and less detailed, whereas renderings are more true to the form, and give as close a three dimensional appearance to
objects as is possible on paper.

Designers make different kinds of physical models all through the design process as a part of the act of ‘form giving’. Models give a better understanding of the product form than renderings, as they can be viewed from various angles. In the initial phase of design, quick paper or foam models are often made to facilitate overall formal decisions. Detailed and more refined models are made as the design progresses and nears completion. At the end of the design process, appearance models are made, which mimic the final product in every detail and are usually made from harder materials like wood, ABS, polystyrene, etc.

The use of computers in design has increased tremendously, and fairly sophisticated software is available for visualization of forms. The advantage of computer simulated images over renderings is that the forms can be rotated and viewed from various perspectives. Once a computer model is generated, changes can be made fairly easily and rapidly. When the design is finalized, prototypes are usually made to study the product in use. The prototype is a true representation of the final design, both in terms of the engineering as well as form. These representations of images are as much a help to designers in ‘seeing’ what is in their minds, as they are in communicating the ideas to others. Presentations are regularly scheduled through the design process. The purpose of the presentations is to inform the concerned personnel about the development of the design. The imaging techniques mentioned above are used as tools in communicating the design to others. The mode of presentation employed depends upon the state of development or refinement of the design. The representations help the designers to scrutinize their designs, and make
modifications or improvements prior to manufacturing. Testing in design essentially involves design critiques, judgments, comparisons, reviews, etc. This is an extremely crucial activity in the design process, as it aids in evaluation of the design by offering cues as to where and how improvements can be made.

Fig. 4. Design Development Spiral

It may therefore be noted that design is not a strictly linear process, but one which involves repetition and backtracking. Information obtained from testing is fed back into the design process enabling designers to refine their ideas on the basis of the new information. The process of design can be represented visually as a spiral with each loop representing a design cycle comprised of imaging, presenting, and testing. Therefore design decisions are taken in every cycle, and decisions in one cycle may determine the context for a decision in the
next. Imaging, presenting, and testing may not be equally important in each cycle. In the earlier cycles, designers may spend more time on imaging, and as the design progresses, more time may be spent on presenting. The process of testing often goes on simultaneously through imaging and presenting. Every time a designer pauses, and steps back to look at a sketch or a model, it undergoes a process of evaluation giving cues towards further refinement.

**The Design Process at NCR**

Corporations often have well-defined design methodologies and development processes which are based on the corporate design philosophy. Listed below is the "Design Development/Approval Process" followed at National Cash Register (NCR), Dayton, Ohio. This process is divided linearly into six phases which represent six different activities typically followed in the act of designing.

I. Task Definition

This phase involves exploration of the influence of technology and the marketplace on product system configuration. The factors which contribute to design resolution are identified and documented. A design brief is prepared. As part of evaluation, reviews are conducted with appropriate plant personnel to assure that results are properly coordinated with program objectives.

II. Concept Exploration

A range of design alternatives are explored through sketches or CAD models. The general direction for design development is established through evaluation reviews conducted with appropriate plant personnel.
III. Design Development

Of the various alternatives generated in the earlier phase of concept exploration, one is selected for further development through sketches, three dimensional form studies, paper mockups, tape drawings and CAD models. The design direction is finalized through reviews conducted with the concerned plant personnel.

IV. Design Intent Approval

This phase entails presentation of the design intent with soft models made from paper or foam. A design intent review is conducted with ID representatives, plant management representatives, division management representatives, and at times representatives of the executive management. This team normally comprises of personnel from manufacturing, finance, and marketing. The results of the review are documented and distributed by the ID group.

V. Design Finalization

A design layout is developed, including final mechanical/electrical components. If required, a design detail layout is prepared for the final appearance model. The drawings are reviewed by ID and R&D.

VI. Final Design Approval

This phase is not an essential one, and the design finalization phase mentioned earlier can be the terminal phase. In this phase, a detailed model of the final design is made, and the design is photographically recorded. Design reviews are conducted with the plant manager, the division management, and the executive office. The results of the final review are documented and
distributed by ID.

**Evaluation in Education**

A major portion of the research in the field of evaluation has been done within the domain of education, and therefore is mostly focused on program evaluation. Of the large number of definitions of the term ‘evaluation’, some are substantially different from one another whereas some are fairly similar. Evaluation, according to a prominent long-standing definition, involves comparing objects and outcomes. It has also, in broader terms, been referred to as the combined study of performance and values. There is however a definition proposed by the Joint Committee on Standards for Educational Evaluation, which is, ‘Evaluation is the systematic assessment of the worth or merit of some object’. Value is the root term in this definition, and it denotes that judgment is an essential element of evaluation. Stated otherwise, if a study does not say how good or bad an object is, it is not evaluation.

Evaluators spend a considerable amount of time and effort in trying to ensure that the evaluation is free of individual biases and motives. In spite of being based on objective procedures and unbiased information, evaluation is not value free. This position has been argued against by writers who believe that value judgments are subjective interpretations, and one group’s values should not be given precedence over another’s. This point of view proves that there can be many different and often contradictory perspectives which make the evaluation indeterminant. Evaluation is indeed a complex process and demands great care on the part of the evaluator in gathering information and providing a defensible rationale for the value perspective employed. Evaluation can be
comparative or non-comparative, it can focus on a single product or compare it to available alternatives. Both methods are equally valid and the decision to use one over the other depends upon the object of the evaluation.

In social science practice, evaluation has emerged as a distinctive field in the last fifteen years. Patton, a leading researcher in evaluation theory, defines it in further detail as a practice involving the systematic collection of information about the activities, characteristics and outcomes of programs, personnel or products for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs, personnel or products are doing and affecting. Though extensive in length, the definition is fairly simple and can be broken down into four major sections - (1) The systematic collection of information about, (2) a broad range of topics, (3) for use by specific people, (4) for a variety of purposes.

William H. Gephart (Patton, 1981) believes that there can be no short, succinct definition of evaluation, and proposes six different ones each representing a definitive task.

1. The classificatory definition defines evaluation as a 'problem solving strategy' employed for establishing the relative or absolute worth of various choices.
2. The comparative definition likens evaluation to research, development, management, and other problem-solving strategies, pointing out similarities and differences with each.
3. The operational definition tells how an evaluation is conducted, from identification of the impending decision through data collection and analysis to information use.
4. The componential definition explains that evaluations include a problem, a situation involving choices, data on the worth of options, a context, a set of values, a time frame and so on.

5. The ostensive definition gives examples of what to buy.

6. The synonym definition includes such words as judgment and appraisal.

These six definitions taken together form the concept of evaluation.

Models of Evaluation

Ernest House (Patton, 1981) has identified eight different kinds of models of evaluation that prescribe evaluators ought to do, and explain how to conduct a particular type of evaluation. These are,

1. Systems analysis, which quantitatively measures program inputs and outcomes to look at effectiveness and efficiency.

2. The behavioral objectives approach, which focuses entirely on clear, measurable and specific goals.

3. Goal-free evaluation, which examines the extent to which actual client needs are being met by the program, regardless of what the staff involved in the evaluation think.

4. The art criticism approach, which makes the evaluator's own expertise-derived standards of excellence a criterion against which programs are judged.

5. The accreditation model, in which a team of external accreditors determine the extent to which the object of evaluation meets professional standards.

6. The adversary approach in which two teams battle over the summative question of whether a program or product should be continued.

7. The transaction approach, which focuses on the program process.
8. Decision-making models, in which the evaluation is structured by the decisions to be made.

The different definitions of the term 'evaluation' and the various models mentioned above, reveal alternative approaches to and conceptualizations of the practice of evaluation. Models are basically prescriptive, they suggest 'do's and don'ts' to evaluators, whereas the types of evaluations to be discussed below are descriptive. Six categories of evaluation based on the purpose of the evaluation effort and by the kinds of activities that tend to be stressed are described below.

1. Front-end analysis (also referred to as preinstallation evaluation).

   In the case of educational programs, this type of evaluation is conducted prior to installation of the program so as to facilitate the planning and implementation, and deciding whether or not the program should be implemented.

2. Evaluability assessment

   This type of evaluation is done to assess how well a certain program or product can be evaluated before a formal evaluation is undertaken. The feasibility of various evaluation methods and approaches is looked into, and issues of the scope of the evaluation, the cost, the limitations are assessed.

3. Formative evaluation (also referred to as process evaluation).

   Information about program improvement, modification, and management is achieved from formative evaluations.

4. Impact evaluation (also referred to as summative, effectiveness, or outcome evaluation)

   The impact or the results of the evaluation, are determined by this type of evaluation, and it helps in making major decisions about the continuation,
expansion, reduction, and funding of the program.

5. Program evaluation

According to the Evaluation Research Society, the kinds of activities involved in these evaluations vary widely from periodic checks of compliance with policy to relatively straightforward ‘tracking’ of services delivered, to ‘counting’ of clients.

6. Meta-evaluation (evaluation of evaluation)

Professional critiques of evaluation reports, re-analysis of data, and external reviews of internal evaluations are included in meta-evaluation.

These types of evaluations are not mutually exclusive, and all evaluations usually involve activities from more than one evaluation type, either occurring simultaneously or followed by one another. As mentioned before, this categorization is based on educational program evaluation and for the kind of product evaluation mentioned here, these might not be directly applicable.

Front-end analysis can be conducted in case of products before they are manufactured, i.e. during the design phase, to decide if the particular design should be implemented. This system of evaluation will be predominantly comparative, as the industry needs a device that will permit testing various design alternatives. However, some of the methods used will be non-comparative. Referring to the models of evaluation proposed by Ernest House, it may be said that this system of evaluation may be likened to goal-free evaluation. It may also be noted that the art criticism approach will not be suitable here as it relies on criteria defined by evaluators themselves. Studying complex issues virtually requires the employment of multiple data collection and data analysis methods. Different methods like questionnaires, personal and
group interviews, observations, etc. reveal diverse information. Also, certain information can be derived only by certain techniques. In the evaluation of meaning, the major advantage of the multimethod approach would be not just the quantity of the data collected, but also the diversity of the data collected on a single topic. This also allows for comparison of data between methods, which can, in turn help in assessing validity.

Within each category of the evaluations mentioned above, there are numerous types of specific evaluations of which some could be relevant to products. These are

1. Appropriateness evaluation: This asks questions about the services that the clients should be receiving, and whether they are appropriate to client needs.
2. Effectiveness evaluation: This evaluation is aimed at finding out to what extent the product is effective in attaining its goals. For example, is the handle designed to communicate to the user that it can be grabbed to lift the product?
3. Extensiveness evaluation: This evaluation looks to discover to what extent the product is able to deal with the total problem that has to be solved.
4. External evaluation: This evaluation is conducted by people outside the project in an effort to increase objectivity. Designers are too closely involved with their designs to be able to judge them objectively. They might also make certain assumptions about users and their perceptions which might be inaccurate. These need to be tested out before they can be implemented.
5. Needs assessment: This evaluation technique addresses the issue of client needs and how they can be met.
Patton lists over a hundred different types of evaluations in his book Creative Evaluation (Patton, 1981). There are other techniques, questionable in value, but practiced nonetheless...

Guesstimate approach: What do we think is happening to clients 'without the pain and bother of actually collecting the data'...

Quick-and-dirty evaluation: How can we do this as fast as possible at the lowest cost?

Weighty evaluation: How can we produce a thick evaluation report?

**Evaluation Strategy**

An evaluation research strategy is a plan that provides a basic direction for the evaluator, and guides in selecting particular techniques or methodological practices for the evaluation design. This system of evaluation, designed to scrutinize the aspect of meaning, will necessarily consist of data gathering through testing followed by relevant analysis. There are various levels of decision making involved in designing an evaluation study. Typically, the evaluator moves back and forth between general methods strategies and specific evaluation questions which could be a part of the study. The research strategy has to be translated into concrete data gathering techniques, instruments and operations which are specific to the area of application.

Research in cognitive engineering aimed at studying the interface between human beings and machines, has revealed that a large number of products do not respect users' cognitive models. Usability tests are designed to study whether products can be effectively handled by the people they are meant for. Some of these tests involve questionnaires, interviews, video recordings of
subjects performing predetermined tasks, followed by detailed analyses. These analyses depend upon the methods of data gathering and the specific nature of results sought. It needs to be noted that this evaluation is not aimed at testing the usability or the operational aspects of products. It is understood and appreciated that meaning emerges through use, and that the context of use constitutes an essential component of the meaning of an object. However, the system outlined here does not address this particular aspect of human-object interface, but attempts to evaluate what can be referred to as “the appeal” of products, which is primarily a function of form and surface attributes like color, texture, patterns, etc.

The importance of testing and evaluation throughout the design process has been stressed earlier. Referring to the “NCR Design Development/ Approval Process”, an evaluation or review is conducted in each one of the six phases which aids in deciding the general design direction. Decisions in design are often made on the basis of individual biases and subjective interpretations. Designers frequently make ‘educated’ guesses on issues where information from users is not easily available. This can, and should be avoided by setting up objective evaluations of design solutions or alternatives. Potential users and not designers, clients, or marketing personnel should be instrumental in deciding what a product should be. However, users cannot be expected to respond to a sketch, a rendering or a model just as they might to the actual product. These only provide limited information to the user (or observer), and testing done with sketches or models should consider this limitation. Different methods will need to be employed for testing, depending on whether sketches, renderings or models are being used. A matrix of methods and phases of the design process,
which shows which method should be employed in each of the phases is charted. Depending upon the progress of the design, the evaluation methods to be employed will be changed.

However, all cells of the matrix are not discussed in this thesis. More emphasis has been given to the “Design Intent Approval” phase of the NCR Design Development/ Approval Process”. The testing will be done using soft models made from paper and foam, which have very limited or no operational capabilities. Detailed appearance models are highly skill and labor intensive, as well as extremely expensive to make and hence it is not feasible to make appearance models of all the design alternatives for the purpose of testing. The industry needs methods that permit rapid prototyping and rapid testing, and this thesis is an attempt towards providing the solution.
Table 1. Evaluation Matrix

<table>
<thead>
<tr>
<th>The Design Process</th>
<th>Evaluation Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Task Definition</td>
</tr>
<tr>
<td>2</td>
<td>Concept Exploration</td>
</tr>
<tr>
<td>3</td>
<td>Design Development</td>
</tr>
<tr>
<td>4</td>
<td>Design Intent Approval</td>
</tr>
<tr>
<td>5</td>
<td>Design Finalization</td>
</tr>
<tr>
<td>6</td>
<td>Final Design Approval</td>
</tr>
</tbody>
</table>
CHAPTER IV
THE DESIGN PROJECT

It was decided to include in the thesis plan, in addition to a written document, a design project that would provide an object that could be used in performing an actual evaluation. The design project is viewed as a vehicle to support the evaluation by providing an object for testing.

The product selected for design was the laptop computer. Personnel at NCR Corporation provided help in design and modelmaking on account of their involvement with the design of laptop computers. Three designs were worked out for three different user groups; high school children, college students, and young professionals. Advancement in electronic technology has brought computerization to a large number of products and it is projected that it will not be long before high school children actually use portable computers at school and home.

Design Methodology

The methodology for the design was based on a heuristic sequence of steps outlined by Butter (Butter, 1989) for incorporating desired attributes in the product. This methodology is discussed in Chapter II, ‘On Product Semantics.’ The first step involved establishment of the general objectives and constraints for the laptop computer. It was decided to design three laptops with varying
expressions suitable for the respective user groups. These three laptop computers were treated as a family of products; so that they bore definite resemblances to each other but were different in their own way. The differences between them are subtle and they share common components. As these were designed primarily for the purpose of testing, all details were not worked out. More stress was laid on generating innovative forms that were semantically more appropriate to the concept of a laptop computer than are the existing laptops. The overall dimensions of the laptops and the screen were assumed on the basis of the average sizes of laptop computers and screens. Only a few of the necessary components were designed. These are:

- The Power Switch
- The Trackball (a trackball instead of a mouse is provided to prevent any additional parts that may be displaced or lost)
- Ventilation Grill
- The Keyboard
- Screen
- Brightness and Contrast Controls
- Release for the Lid

Details of the other components like SCSI ports, communication outlets, LEDs etc. were omitted. The keyboard is the same for two of the designs, the brightness and contrast controls, and the power switch are the same for all three.

The projected context of use for all three includes work as well as entertainment. A list of desired attributes was generated for all three.
Table 2. Attributes for Laptop Computers

<table>
<thead>
<tr>
<th>High School</th>
<th>College</th>
<th>Young professionals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Youthful</td>
<td>Distinctive</td>
<td>Professional</td>
</tr>
<tr>
<td>Playful</td>
<td>Trendy</td>
<td>Classy</td>
</tr>
<tr>
<td>Toyish</td>
<td>Dynamic</td>
<td>Efficient</td>
</tr>
<tr>
<td>Uninhibited</td>
<td>Mature</td>
<td>Sophisticated</td>
</tr>
</tbody>
</table>

There were some undesirable attributes, of which some were common to all three. It may be noted that certain attributes that appear desirable for one appear as undesirable for the others. The attributes were generated in a brainstorming session with senior and junior students in the product design program at The Ohio State University. The designs for all three laptops were worked out simultaneously.

**Metaphors**

Certain metaphors that relate to the concept of computers were applied to the designs of all the laptops. The electronic medium is quickly changing commonly followed practices on how information is gathered, stored, used, and relayed. Portable computers, floppy disks, modems, and other hardware have reduced the necessity of storing information on paper, in files, in books, etc. Instead of using envelopes and stamps to send reports through the postal service, we use electronic mail, phone lines and computer terminals. Instead of storing typewritten sheets of paper in folders, we save them on magnetic disks. However, users of such electronic media are not always aware of how computers work, and arguably they do not need to know. At the same time, there is
absolutely no reason for a computer to be a rectangular gray box that communicates nothing about what it is, what it can do, how it can be used, and so on... Metaphors are therefore used in design to transfer meanings from one usually unfamiliar domain of experiences to another usually less structured or novel domain. As computers are replacing paper and printed matter, the metaphor of a file folder was used in designing the overall forms of the laptops. The metaphor of a ring binder was employed in the design of the hinges for the lid of the computer. The curvature in the body coupled with the layered or stepped effect is representational of a stack of papers. The curves at the front (opening) end of the computer follow the curvature of the keyboard.

Fig. 5. Laptop Computer showing Metaphor of File Folder and Ring Binder
The curvature of the keyboard is based on anthropometric studies which states that the standard keyboard leads to undue stress on the wrist and is one of the causal factors in the carpal tunnel syndrome. A keyboard shaped in the form of a "V" is assumed to be more conducive to the angle at which our arms approach the keyboard. The basis for the curvature of the keyboard lies in this anthropometric research.

The variation in the overall forms is shown in the illustrations of the three computers. The subtle change in the forms from a rectangular, to a trapezoidal with converging sides, to a trapezoidal with converging sides and a curved top edge are manifestations of a change from a more serious, square expression, to freer, toy-like expression. It may also be noted that the lines get smoother, softer, and the radii get larger. The power switch, the keyboard, the brightness and contrast controls and the screen designs were maintained the same for all three designs. However, the release switch for the lid, the ventilator grills, and the overall forms were varied. The forms of the hinges, which are studied, subtle and conservative for the laptop for young professionals, are less so for the second one which is designed for the college students, and most playful and toy-like for the one for high school kids.
Fig. 6. Computer for Young Professionals, College Students, and High School Kids (from top to bottom)
On Modeling in Design

The role of modelmaking in industrial design has been briefly discussed in the preceding chapter. Product designers make sketches, renderings, drawings, build models, and write reports in course of the design process in order to formulate, concretize, and communicate their ideas. Three dimensional models, it is believed, come closest to depicting what the final product will look like when manufactured. This embodiment of ideas can be achieved by different techniques through the employment of a variety of materials, depending upon the progress of the design and the level of complexity desired. However, Schrage points out that models reveal much more than merely designers’ ideas, they reflect corporate culture and a company’s ability to innovate.

“The prototyping culture—the media, methods, and styles companies use to manage their models of reality—offers a wealth of critical insights and opportunities into how organizations design and build value.”

-Michael Schrage, Design Management Journal, Winter 1993

In order to understand this, it is essential to familiarize oneself with the modelmaking vocabulary employed by industrial designers. Certain terms and phrases mean different things for different individuals, companies, and cultures. ‘Prototypes’ and ‘mock-ups’ are terms that have a wide range of interpretation, and it is contested that what a prototype means to a company is a reflection of its prototype culture. In certain organizations, the word prototype refers to all the models built, including the rough study models or form studies, as well as fully detailed operational models. However, in organizations with strong engineering and technical cultures, the prototype is the one and only working model that the engineers have figured out how to manufacture. Schrage’s
interpretation of ‘prototype’ includes all the models that are created as a part of
the design process. He stresses that the quality of the final product depends on
the quantity and quality of the prototypes. While industrial designers have
always appreciated the relevance of models in designing; the business world,
probably due to the recent attention to design, is now learning to value models
as essential stepping stones in product development. Rapid prototyping has
gained considerable significance in industry with the emphasis on accelerated
product development cycles. Design-sensitive organizations generally produce
most models in the least amount of time, whereas the companies that are less
design sensitive usually want one perfect model.

Models can be classified either on the basis of the media (materials) used
in creating them, or in terms of the degree of development of the design. Here,
they are classified grossly on the basis of the development of the design into two
categories. It will be noticed that there is a fair amount of overlap between the
methods of classification, as the material selection also depends upon how
developed the design is.

Study Models

Three dimensional models give a better representation of form than
drawings. They allow viewing from different perspectives, and can be studied in
movement to give a better idea of the inter-relationship between volumes. Three
dimensional models can also be used to study the amount of space occupied by
the product in its environment. The virtues of models over representations on
paper are abundant. It is customary for designers to make their own models as a
part of creative problem solving. Usually created to add dimensionality to a
sketch or a rendering, study models are also referred to as form studies and they aid the designer in confirming or rejecting ideas generated on paper. However, models generally take considerable amount of time to create, and therefore study models are made by designers quickly, using materials that are easy to shape and reshape. Typically, they are made from paper, foam core, blue foam, clay, or a combination of the above. Various materials are invariably involved, as good and rapid modeling calls for creativity in material selection. They are therefore rough, unfinished, and not true to the final product in terms of fine detail, surface textures, or color. They are meant for feedback to the designer, and might be used for critical appraisal from fellow designers, but seldom for communicating design ideas to management or ‘non-designers.’

Appearance Models

Models made for the primary purpose of communicating the design to other groups in the product development process, are true representations of the final product in terms of the appearance. They might not be made from the true materials, but are finished so as to simulate the final product form. Appearance models are usually made from harder materials than those used to make rough models. Polystyrene, acrylic, ABS (acrylo-butadiene styrene), and PVC (polyvinyl chloride) are some of the commonly used plastics. Metals and wood are also used depending upon the requirement. Ren shape, yellow urethane foam, white urethane or acrylic foam that are denser and better at retaining imparted form, can also lend themselves as suitable materials for appearance models. If the models have to sustain excessive handling and manipulation, foam is not a suitable or wise choice.
**Modemaking Media and Technology**

The prototype culture of an enterprise is partially controlled by the type of organization. Small entrepreneurial companies are usually more prototype-driven, while larger organizations are spec-driven. Are prototypes built around predefined specs, or are specs defined and refined as the prototypes are created? David Kelly of IDEO believes that organizations need to "move from spec-driven prototypes to prototype-driven specs" to be innovative. Not everyone might agree with this line of thought, but it certainly proves that modemaking is assuredly an essential phase in the product development process.

Technology has helped discover various materials, tools, and techniques that are constantly redefining the modemaking process, as well as the quality of the model. Modeling materials and processes undoubtedly influence the design. GVO's Michael Barry argues, "As the modemaking material changes, the design process changes. Hewlett-Packard used to do its calculator modeling in cardboard, which explained why all their calculators featured angles and edges. When the company switched to foam, you saw calculators that had a softer and more rounded look... Foam lends itself to certain kinds of subtleties." In the automobile industry, the traditional modeling material has been clay, which in spite of being an excellent material for large undulating curves, is not always suitable for rapid iteration. In the words of Barry, "When a (clay) model starts to harden up, so does a lot of thinking."

Styrofoam (blue foam) and urethane foam (white and yellow foam) are excellent materials for modemaking, as they are easily workable and hence suitable for large radii and organic shapes. Foam is available in standard sizes as
slabs and blocks. Foam has a characteristic surface quality which is difficult to conceal or disguise. Ren shape, a very high density foam, comparatively new in modeling, is easy to carve or sculpt, and is not easily dented. It can be used to generate complex shapes, and it can also accept paint without revealing the porosity of the foam.

**Virtual Modeling**

With the increasing involvement of computers in design, it is now possible for designers to have virtual three dimensional models in plastic without anyone having to as much lift a file or a sanding block. Three dimensional computer modeling software available today is fairly sophisticated, and is capable of generating elaborate forms with intricate details. With the amazing potential of ‘virtual reality’, computer aided design offers advantages that palpable physical models never could. Computer models created with CAD tools can be viewed on the monitor as wire frame models or can be skinned with a variety of textures like rubber, glass, marble, plastic, chrome, etc. These digital models can be located in environments or spaces of one’s choice created with the same tools to view them in context. The biggest advantage however lies in the fact that they can be modified to create alternate solutions with the ease and rapidity of a few keystrokes at the computer terminal. Photo realistic renditions of products, with high resolution color print-outs are replacing photographs of physical models in certain organizations.

CNC milling, an automated computer controlled process accepts solid models created by CAD software as input data and machines the form out of the material loaded on the milling machine. Another computer controlled process,
called stereo lithography can create models without any machining or tool work. Stereo lithography (analogous to three dimensional printing) also uses data from the CAD system to mold special plastics called curable photopolymers using laser light. The setup consists of a bath of the liquid photopolymer containing a movable platform. Ultraviolet laser scans successive layers of the CAD model and cures the layers of plastic, creating a copy of the model, section-by-section. As the first cross-section sets, the platform is lowered fractionally and the next layer is scanned. The entire model is generated when all the layers are scanned. This is then cured in an oven and is ready for further operations, usually of surface treatment (painting, etc.).

These techniques require equipment that is rather expensive, that only large companies can afford. One of the most important properties of modeling media is cost; as it is easier to experiment with inexpensive materials like paper, cardboard, and Styrofoam. Modeling is inherently a multimedia process, and good models reflect creative material use.

In view of the emphasis on rapid prototyping and acceleration of the product development cycle, organizations are turning towards electronic media as against palpable modeling. Changing the modeling material usually affects the design process. In certain organizations, moving from renderings to three dimensional models represents a definite design direction, if not freezing the solution. To quote Schrage, "... hard models can harden thinking..." Rough foam models are therefore seen as transition media between 2D (sketches, drawings, renderings) and 3D (appearance models). It is often in the interest of the industrial designer to maintain the roughness of the foamware to emphasize the
temporariness of the design. The value of rough foam models lies in their very roughness, inaccuracy of surface, and lack or color or paint. Barry of GVO argues, “The minute you lay it in color, you finalize it... you send a cue that its finished.”

The emphasis placed by management on ‘speed to market’ as an essential requirement for success should not imply fewer models or lesser iterations to save time. Instead, better techniques for rapid modeling should be discovered to allow for necessary iterative remodeling.

**Design, Evaluation and Concurrent Engineering**

The need to evaluate often brings the design activity to a dead halt as it necessitates production of drawings that can be used in the creation of models. If designers do not make the models themselves, they are required to supervise the modeling. This is likely to cause an undesirable slowdown in the design process.

In today’s world of short product development cycles, there is tremendous emphasis on parallel or concurrent engineering. Contemporary concurrent design principles encompass all disciplines into a simultaneous product and process design paradigm including product planning, serviceability design, human factors, industrial design, safety, disassembly and reclamation, and more. Interdisciplinary teams work simultaneously on the development of new products, and there is continuous flow of information between groups. This is more time efficient than the traditional process where the product travels from one group to another in a linear sequence. Research had shown that most successful companies deploy multi-disciplinary product development teams. As
soon as the conceptual design is captured, it is followed by a quick dispersement of the design intent. This vastly reduces ‘time to market’ for the product. Similarly, if design and evaluation can run concurrently, coupled with a method for rapid prototyping and testing, there can be considerable time saving in the entire process.

Computers in Design and Evaluation

As discussed earlier in virtual modeling, the use of the computer as a tool in the design process has, to a certain extent, reduced the need to create palpable three dimensional models. However, it needs to be studied whether or not these computer generated images aid in a ‘true’ perception of form. There are certain inherent disadvantages of the computer technology, especially the output devices, that put restrictions on ‘true’ representation. The perception of a three dimensional virtual object on the monitor is hampered by the fact that the screen is a flat surface. The monitor casing, often a beige plastic frame, only adds to the sense of unreality of the object displayed within. There is also the restriction of the maximum size of the display screen. Large objects cannot be viewed full scale on the screen. The luminosity and pixellation of the screen are other factors that need to be overcome. With the advent of virtual reality and holography, these problems might have solutions. Meanwhile, the medium offers tremendous flexibility and possibilities that were unforeseen a few years ago, and should therefore be thoroughly explored.

Most design offices these days are equipped with computers. Designs can be modeled on the computer using CAD software. The involvement of the computer can be taken a step further by using it in the testing as well. An
interactive questionnaire can be created using authoring software, and administered on the computer. Presumably, this would prevent the slowdown in the design process as manpower would not be directed towards producing drawings and models.

Knowing that computer generated models do not give as true a representation of form as palpable models, an alternative solution is proposed. There exists the possibility of combining two dimensional computer images with three dimensional models. This technique proves to be time saving, and at the same time provides a three dimensional model. The overall form of the object is sculpted from the appropriate material (foam, Ren shape, plastic, etc.). What is not modeled physically is modeled on the computer. The right balance between the parts that need to be 3D and those that can be abstracted to images needs to be worked out. Minor details and complex intricacies can be sketched on the computer using drawing or painting programs. These images are then printed out and pasted on the object. The images can be either drawn from scratch using the appropriate software, or a scanner can be used to digitize photographs and objects. In spite of being two dimensional images, certain three dimensionality can be added by shadows, highlights, and shading. The amount of imagery in relation to the form would undoubtedly be dependent on the object, and would therefore vary for each object. This technique permits a calculated level of abstraction and is found rather promising.

This technique of combining 2D and 3D was adopted in the creation of the models of the laptop computers that were used in the testing.
Models were made for all three laptop computers using white polyurethane foam (also referred to as Clark foam). Along with these three models, a fourth model of a laptop computer was made, which was essentially a replica of an NCR design. This fourth design was intended to represent the typical laptop computer seen in stores today. It was decided to build the models by adopting the modelmaking process normally practiced in industry. The material was supplied by NCR, and the designs were discussed with designers and model makers to learn how these would be typically modeled in foam. On studying the design, it may be observed that the major challenge to modelmaking would be generation of the curved-stepped-layered effect on the base unit of the portable. Various suggestions were made by the NCR personnel. The method that was followed is graphically represented below.

![Diagram of modelmaking process](image)

**Fig. 7. Modelmaking Process**

This method was adopted for reasons of accuracy and better surface finish. 3/16" slices of white foam were cut on the table saw. Male and female mold parts were made from bass wood by cutting the necessary curve on the band saw. Extra curvature was incorporated in the mold to allow for spring back.
Two slices of foam were glued together at one edge to maintain their relative positions through the bending and lamination. Slow setting epoxy glue with a working time of ninety minutes was applied between the slices. They were clamped down and the epoxy was allowed to set for twenty-four hours. As the epoxy set, a hard layer was formed between the two foam layers forcing them into the curve. This process was repeated twice for the four layers. The rest of the details were then fabricated. The keyboard, power switch, release for the lid, the screen, brightness and contrast controls, and the ventilator grill were created as illustrations on the computer and printed out. These print outs were pasted on the models.

Fig. 8. Computer Generated Image of the Power Switch

Fig. 9. Computer Generated Image of the Keyboard
Fig. 10. Computer Generated Image of the Track Ball

Fig. 11. Computer Generated Images of Grills for the Young Professionals, College Students, and High School Kids (from left to right)
Fig. 12. Computer for Young Professionals (Closed)

Fig. 13. Computer for Young Professionals (Open)
Fig. 14. Computer for College Students (Closed)

Fig. 15. Computer for College Students (Open)
Fig. 16. Computer for High School Kids (Closed)

Fig. 17. Computer for High Schools Kids (Open)
Fig. 18. Model of the NCR Computer (Closed)

Fig. 19. Model of NCR Computer (Open)
CHAPTER V
EVALUATION PARAMETERS

The testing is based on the four design theories proposed by Krippendorff. The various parameters that have been isolated are discussed here along with methods for measuring or evaluating them.

An important factor to consider is the degree of roughness of the stimuli being tested. The evaluation system has to account for the roughness of the model. A certain roughness factor can be defined for the stimuli and used in the final calculation. For example, a blue foam model might be assigned a 'roughness factor' of 0.9, and a working prototype a 'roughness factor' of 0.1.

Four Design Theories

Klaus Krippendorff has been making continuous relevant contributions to the development of a product semantic theory. He has, in the form of design theories, identified four essential contexts in which artifacts need to survive. A brief explanation of these theories is given below.

It is essential to have measurable parameters which can be used to develop an evaluation strategy. A large number of product attributes have been identified on the basis of the above mentioned design theories of use, language, genesis, and ecology.
The Context of Use

Designers need to be aware of the users' cognitive models to understand how they interact with objects. People develop some sort of operational logic which can help them to comprehend how things work and how they can be made to perform according to their needs. Products need to be recognized for what they are and what they can do, they have to be self-evident. They have to be designed so that they can be handled with well understood practices - they should be self-instructive. Users' conceptions are based mostly on ideal-types. Objects are recognized and classified on the basis of their similarity or dissimilarity to the ideal-type which in Wittgenstein's words is 'the good example' (Vohra, 1986). There are many different ways in which artifacts can be used, and the meaning of an artifact equals the range of imagined contexts of use. A book is primarily meant to be read, but it can also be used to support the short leg of a table, conceal money, or knock someone down.

Interfacing with something requires that the user gets adequate and timely feedback on her/his actions. The artifact should ideally give the user information relevant to the cognitive models in use. Some users need little information to get acquainted with a product, some are able to find their way through by trial and error, and there are others who need tutors to help them through. Design should support all these user activities. Cognition is dynamic, human beings are continually learning new things through new experiences, and products have to afford the cognitive dynamics.

Evaluation Parameters for The Context of Use

1. Is the product recognizable for what it is? Is it self-evident? Can there be some way of measuring or evaluating this quality of self-evidence?
Objects can be tested for identifiability by conducting identification or recognition tasks; or for memorability through recall tests. Recognizability tests can be conducted with the help of rough models as well as working products. Recall tests can be conducted to see if any of the product features are able to create a lasting impression in the minds of the subjects. The recall test can be verbal or visual. In a testing situation, the product being tested can be taken away, and the subject can be asked to sketch the object, or describe it verbally. Though drawing and observational skills vary from person to person, the results of the test give an idea of which parts of the product are most distinctive. For a visual recall test, the drawings can be analyzed to look for product features that appear most frequently, and the protocol can be analyzed for the verbal recall test.

Self-evidence

The quality of self-evidence is expressed through the recognizability of a product; the easier it is to recognize a product, the better it is ranked on a scale of self-evidence. Recognizability can manifest itself in levels of accuracy or precision. At best, the test subjects might be able to say that the object is a portable computer meant for high school kids, and on the other hand, just as an electronic device of some sort. Recognition comes through our capability to assemble data into categories, and label them. We place objects that substantially share features and show a correlation into one category, such that each category can be perceived as a distinct concept. It is important that product form shows its belongingness to the category in order to be recognized as a legitimate member. Objects within a category define typicality. A new object can be perceived as a member of a certain category depending upon its
closeness to the typical member of the category, which can also be referred to as the degree of typicality.

A model of category system, conceived by Rosch, can be used here for studying the degree of recognizability. Human beings map things holistically as a part of a larger network of objects in search of efficient information processing techniques. The mind creates a model of the real world with concepts that are functionally linked. The model proposed by Rosch is split into levels of abstraction which are labeled the Superordinate, Basic, Subordinate levels. The fourth level is referred to as Real World Examples, and consists of objects that exist in the world.

The example used here starts with electronic devices at the superordinate level and ends at existing laptop computers like Macintosh, IBM, etc.

<table>
<thead>
<tr>
<th>Superordinate Level</th>
<th>Electronic Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Level</td>
<td>Computers</td>
</tr>
<tr>
<td>Subordinate Level</td>
<td>PCs</td>
</tr>
<tr>
<td>Real World Examples</td>
<td>IBM</td>
</tr>
</tbody>
</table>

Fig. 20. Levels of Categorization
The shared semantic features of the actual instances in the subordinate level concept 'Laptops' help define this concept. Similarly, 'PCs', 'Laptops', and other subordinate level concepts help define the meaning of the basic level concept 'Computers'. In recognition tasks, objects are first seen as members of the basic level category and further processing is required to understand them at a superordinate or subordinate level. However, this does not apply to new products with innovative forms that are not easy to categorize. In these cases, they might first be seen as members of a superordinate level which is more general and includes a variety of objects. The easier it is to recognize a product, the closer it is to the real world example.

Self-evidence can be measured at two levels. Are subjects able to recognize the model as a laptop computer at a holistic level, and are they able to recognize individual parts as the keyboard, on/off switch, brightness control, contrast control, etc.?

One of the major parameters affecting self-evidence is the degree of typicality. We judge objects in terms of typicality (degree of closeness to the ideal-type) and then try to explain how it differs from the ideal-type. New objects need to strike a balance between these similarities and dissimilarities so that the object maintains some relation to the ideal-type, at the same time introducing innovations which make for interesting deviations from the same ideal-type. The closer an object is to the typical example of the category, the easier it is to recognize it. The degree of dissimilarity should arouse interest, but should not be so high as to be confounding or bewildering. It needs to be noted that we are dealing with perceptual and not actual similarity, and that can manifest itself in three forms. One is expressive (for e.g., the subject might say,
"Both objects look rugged"), the second one is feature based (for e.g., the subject might say, "This one has more buttons on it than the other one") and the third is purely formal (for e.g., the subject might say, "This one looks too much like a cube, and this one is more rounded"). If these degrees of similarity and dissimilarity can be identified and quantified, they can provide a measure of how innovative a product form is, and whether or not it falls within company policies and philosophies.

2. Is the product configured so that it can be handled with well understood practices? Is it semantically accommodative? Can the degree of semantic accommodation be measured?

Users develop their own operational logic about how things work and how to interact with them. These are known as cognitive models and they can be constructed by having users talk about their perceptions, actions, behaviors, etc. Protocol analysis is a popular technique for studying verbal explanations obtained from users as they perform certain tasks. Users can be asked to think aloud while performing specific operations in order to study product-user interaction.

Studying products in use necessarily requires an object or product that works. A foam model cannot be tested thoroughly for handling as it does not offer true feedback. Testing for physical handling cannot be done with models unless they are made from materials that closely resemble the final materials of the product. A model cannot provide the tactile and functional feedback that is essential for studying objects in use.

On account of the techniques used in model making, models are often rather fragile and therefore unable to survive excessive handling. However, there
are a few ways of studying certain attributes of models that deal with the basic handling of a product. Test subjects cannot be allowed to manipulate soft models. For hard models with limited functionality, some physical manipulation can be permitted. For example, in case of a laptop computer, a hard model can be used to study the following tasks.

1. Pick up the laptop computer from the table. (It would be observed as to how the model is picked up and held during transportation.)

2. Take it over to the desk and set it down so that you can use it.

3. Open the lid of the computer.

4. Start the computer. (In this case, one can test to see if the subjects are able to perform what should be but is often not an easy task. The model would obviously not start, but it would reveal whether they can locate the startup switch.)

5. Adjust the brightness and contrast levels to suit yourself. (Again, this would probably not work, but would disclose whether the subjects are able to locate the controls for doing so.)

6. Adjust the screen angle to minimize glare.

7. Move the cursor on the screen and select the document called ‘test.’

8. Open this document.

And so on...

The list of tasks would depend upon the product attributes being tested. If the designer wished to test how intuitive it is to change the batteries, the list of tasks would incorporate the relevant actions. If the individual tasks are complex or take considerable amount of time to perform, a time study can be conducted with specific benchmark tasks. A benchmark task is a list of a series successive actions that a user is asked to perform in order to study specific product
functions. The amount of time taken to execute a task is a measure of the complexity of the task and of the design. Expected time for each task is calculated on the basis of the number of steps involved, the nature of the task, the user profile, etc. Subjects are then asked to perform the tasks and the amount of time taken to perform the tasks is measured. Large discrepancies between actual time taken by subjects and expected time calculated suggest that the design is not self-evident or instructive. However, none of the tasks mentioned above can be performed on the white foam models. Subjects will not be allowed to touch the models. The tasks will be defined thus.

1. How would you open the lid of the computer?
2. How would you turn the computer on?
3. How would you move the cursor on the screen?
5. How would you adjust the angle of the screen?
6. How would you adjust the brightness and contrast levels?

As subjects are not allowed to touch the models, they can be asked to indicate how they will perform these tasks instead of actually performing them; or they can be interrupted as soon as the researcher knows what action they intend to perform. They can also be asked to think aloud as they go through these operations. Here, subjects will be asked to verbalize their thoughts without having them perform the actions. As the tasks are of a fairly short duration, a time study will not be performed.

3. Is the product such that it permits the user to get competent in using it? Is it self-instructive? Can this self-instructive quality be measured?

Self-instructive quality can be studied though the scrutiny of errors. Errors should be instructive and should promote learning. They should not lead users to
dead ends where help is needed but not provided. There should always be feedback that offers the options available for further action. The number of times an error is repeated can be used as a measure of how self-instructive the product is. If the same error is repeated often, it is obvious that the product does not aid the user in its operation. In a testing situation, the subjects can be asked to perform certain tasks, and those that are performed incorrectly can be repeated after a time break to see if the same errors are repeated. If they are, there is definitely a design problem that needs to be attended to. Another parameter for testing the self-instructive quality of products is time—the amount of time taken to figure out how to perform a certain task. A time study can be conducted to evaluate the self-instructive quality of the product. Sometimes, users are required to read instructions in order to perform complex tasks, especially those involving a large number of steps. However, a product that is handled frequently should not require the reading of instructions every time it is used. Portable personal products such as laptop computers should not compel users to carry manuals with them.

4. Does the product provide timely and informationally adequate feedback about whether an action is accepted by it, what it’s internal states are, and which options are available for further action? Is this feedback in accordance with the users’ cognitive models? Can this adaptability to users’ models be measured and evaluated?

Feedback refers to the concept of sending back to the user information about what action has actually been done. Feedback should be such that it confirms user action indicating whether or not it was ‘correct’. It should also motivate further action and specify what options are available to do so.
Appearance models can be tested for feedback if they have limited moving/working parts. If the product does not provide feedback, users tend to repeat the action till there is some confirmation from the system or the product; or they try other actions. When products are being tested for usability, the sessions can be recorded on video tape to study if subjects repeat certain actions till some feedback is obtained. The product interface needs redesign if it causes confusion by not providing feedback. The number of ‘actions-without-feedback’ in a benchmark task can be counted. The purpose of the design should be to minimize if not eliminate these actions by creating an interface that provides timely and informationally adequate feedback. It is important that feedback immediately follows user action suggesting the options available.

5. Does the product support various strategies of exploration employed by users if it is a product of non-habitual use and requires expansion of their understanding to new practices? Does it support user understanding with the minimum number of trials and errors? Can this flexibility be measured?

Human beings explore unfamiliar objects differently using their own personal strategies. Some users are willing to pore over manuals and instruction books, while others require logical transparency in order to use a product. Design should support all the strategies developed by users in order to acclimatize themselves to a new product. Testing for this parameter is possible only with a product that is fully functional.

Metaphors and metonymies are processes by which users extend their understanding to new objects and practices. Test subjects can be questioned regarding the metaphors used by designers to study if the metaphors are relevant and decipherable. Human cognition is dynamic, and we are continually
learning to relate to and use new objects. We use metaphors and metonymies as the techniques to understand unfamiliar objects and practices. Designers too use metaphors in the process of design to facilitate better understanding than that afforded by black boxes, to help users relate to objects they might be unaccustomed to. The use of metaphors also helps in generating innovative forms. For example, a major influence in the design of these laptop computers was the use of forms and patterns of existing objects like the file folder and stacks of papers as metaphors, because the computer offers means of storing information that replaces paper. These forms are abstracted from their domain, and used in this domain of electronic products to aid comprehension.

The testing of whether or not users are able to perceive the metaphors employed by designers can be done using a questionnaire. The object in question can be shown to the users, and they can be asked if and they can be asked if it reminds them of other objects or things. It is known that in testing situations open ended questions which require subjects to be imaginative often produces poor or no results. The open ended question can be followed by another question which forces them to choose one from a list of alternatives. There are two advantages in using the forced choice method- one, it always produces a response as against the open ended question which, and second, these responses are easier to analyze using statistical methods. However, the open ended questions are useful, because if subjects are able to answer them correctly, it proves that the metaphor is expressive enough to be easily recognized.
6. Does the product excite motivation in the user? Does it provoke any kind of goal-achievement behavior? Can this *motivating capability* be measured?

Krippendorff has identified two kinds of motivations, extrinsic and intrinsic (Krippendorff, 1993); and ideally, artifacts should provide both at the same time. Extrinsic motivation resides in seeing the opportunity to accomplish assignments, reaching goals, and expecting to be rewarded or escaping punishment. Intrinsic motivation lies in the pleasure of involvement with the artifact. It is enjoyment for its own sake. It may be noticed that extrinsic motivation is best achieved by keeping as much control as possible in the artifact, whereas intrinsic motivation is achieved by granting users more control of the object. All the parameters discussed earlier of self-evidence, self-instruction, semantic accommodation, etc. affect motivation. Intrinsic motivation can be promoted by keeping some of the logic of processes hidden and fun to explore, allowing users to reconfigure the artifact to a certain extent and apply their own language. Freely configurable controls that enable self-pacing and learning can facilitate intrinsic motivation.

**The Context of Language**

Meanings of objects arise through language because they are talked about by users and non-users as well. Language helps to define artifacts and brings them into interpersonal communication and social practices. Advertising, rumors, and informal discussions, in a big way affect people’s perception of these products. Design is closely interwoven with language as exemplified in the design brief, the contract, specifications, constraints, presentations, etc. The structure of the spoken language plays a big role in the process of conceptualization, design, and manipulation of things. All objects that one acquires, express the person’s individual identity and help in communicating this
to others. Thus, products become symbols of social differentiation, economic status, integration, etc. Objects therefore enter human communication through the structure of language.

**Evaluation Parameters for The Context of Language**

1. Is the product *viable in human communication?* Is it talked of favorably? Does it support meanings attached to it by judges, clients, bystanders, reference groups, or other non-users? Can this viability be measured?

   The term ‘user’ typically refers to the person who physically uses the object. However, the meanings of objects do not depend solely upon the ‘physical users’ but on other ‘participants’ as well. Anyone who interacts with the objects in any way is a participant and is influential in determining what it means. One method to study if the product is viable in human communication is through focus group interviews. The models can be displayed to a group of participants (clients, designers, bystanders, judges, etc.). The group can be asked to judge the products, if necessary on the basis of pre-determined criteria. The focus group interview can be recorded on video and/or audio tape and scrutinized using protocol analysis. If the product is discussed favorably in the group, it implies that it can successfully survive human dialogue.

2. Does the product *survive in a socio-linguistic context?* Does it gain attributes through advertising, rumor, hands-on-experience, which are favorable to the product? Can this be measured?

   Testing is often done by market researchers to study the effectiveness of advertising for a product. However, this is post-production testing, and hence does not offer feedback that can be used to redesign the product. Testing can be
done to study if the product can gain favorable attributes through advertising. Models can be photographed so as to appear as final products, used in creating mock advertisements and shown to subjects. Consider the four laptop computer models. Magazine advertisements for these can be mocked using desktop publishing software. The mock ads can then be pasted within magazines (computer magazines) and shown to subjects. They can be asked to express opinions on the products featured in the magazine, including the new designs of the laptop computers. If the new designs get positive reviews, it can be assumed that the product will gain favorable attributes through advertising.

3. Does the product give it’s user an individual identity and aid in interpersonal communication? Can this aspect be evaluated?

Almost all objects one possesses become expressive of one’s individual identity. All products are suggestive of certain user profiles on account of their physical attributes like form, colors, textures, etc. If three products are designed for three different user groups, population samples of people who represent those groups can be called in as subjects for the testing. They can be asked to look at the products and predict what kind of users would be attracted to each of those products. This would reveal if products are expressive and meaningful to certain types of users.

**The Context of Genesis**

Products have to go through a complex process of conceptualization, realization, manufacturing, advertisement, use, etc. This process connects a large number of people like designers, financiers, engineers, consumers, market researchers, recyclers, etc. To be successful in this genesis, the product has to
pave it's way through this network of production and consumption, and must be meaningful to all those involved in every stage. This is a continuous transformation of material culture in a circle without beginning or end.

**Evaluation Parameters for The Context of Genesis**

1. Does the product pave it's own way through the network of production and consumption? Is it meaningful to everyone involved at every stage of the genetic process? Does it communicate it's potential to engineers, financiers, managers, users, market researchers, recyclers, etc.? Does it express where it comes from, what can be done with it, and how and why and where it might go next? Can this *communicability* be measured?

To study the survival of a product in the context of genesis, it has to be studied through its entire developmental cycle, from the conceptual stage all the way up to its recycling. The information relating to a product in its developmental cycle is spread over different media like engineering drawings, cost sheets, production plans, models, instruction booklets, user manuals, etc. A product's ability to communicate to the personnel involved in its genetic cycle depends upon all of these associated information packets or gestalts. A model plays its role in the production and consumption cycle by communicating its manufacturability to the engineer. One of the parameters mentioned earlier that deals with the artifact's viability in human communication can be evaluated by doing focus groups. The same technique can be applied to study the artifact's survival in the context of genesis. A focus group consisting of engineers, financiers, designers, marketers, managers, recyclers, can be organized. The discussion can be centered around the different media like drawings, plans, models, etc. The proceedings can be recorded on audio/video tape and analyzed
by protocol analysis or content analysis.

The Context of Ecology

Ecology in biology deals with the interaction between species of animals and plants within a certain environment. Similarly, artifacts and species of artifacts interact among themselves resulting in cooperation, competition, parasitism, symbiosis, etc. Often, the introduction of one product in the market triggers the birth of many new products or pushes some others out of the market. Products develop family resemblances through such ecological interactions. Designers contribute to this by creating new forms, transferring technology from other domains, supporting similarities and dissimilarities, and so on. The meanings of entire species of artifacts are relative to all other artifacts in the environment that they interact with. Therefore, meanings are the key to understanding relationships within an environment. Cultural identities are maintained and communicated through generations by the artifacts that are carried over from one generation to the next. Mythology is an essential component which keeps cultures alive. The way products are used often reflects a deep-rooted mythological heritage.

Evaluation Parameters for The Context of Ecology

1. Does the product help in preserving cultural identities and communicating them to subsequent generations? Can this ability to preserve cultural identity be measured?

The survival of an artifact within the context of ecology is largely dependent upon its relationship with the environment it belongs to, as well as human beings and objects within that environment. The relationship between a
product and the local cultural practices is based on a two way interaction between them. Both influence and are influenced by each other. People often discover innovative ways of using products that were never expected while designing.

Products insensitive to culture can cause a disruption of local practices. It is however difficult to predict precisely how products are accepted and become a part of the culture. Testing for these parameters is again possible only with a product that is a part of the lifestyle in an environment. The actual product has to be observed and studied in use in people's homes to comprehend its influence on culture.

2. Does the product survive in a variety of relationships or interactions like cooperation, competition, parasitism and symbiosis? Can this survival ability be measured?

One can speculate about a product's influence on the market, but actual testing can be done only by introducing the product in the market. Test marketing is often done for products in sample populations in certain key cities. This attempts to study users' reaction to the new product, and its survival against existing products. Test marketing can be a suitable vehicle to investigate parasitism, cooperation, competition, and symbiosis amongst artifacts. It can be studied whether the product being tested pushes others out of their niches, or helps others, or depend upon other products in order to survive.

As is obvious from the discussion above, some attributes can be tested with models and rough media, while form some attributes, it is essential to have a working product that has been used and is a part of people's lifestyles.
The following matrix shows this relation between the medium of presentation and the technique of evaluation.

**Table 3. Attribute vs. Medium of Presentation Matrix**

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Medium of Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sketch/Rendering</td>
</tr>
<tr>
<td>Self-evidence</td>
<td>✓</td>
</tr>
<tr>
<td>Semantic Accommodation</td>
<td>✓</td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
</tr>
<tr>
<td>User Understanding</td>
<td>✓</td>
</tr>
<tr>
<td>Viability in Human Communication</td>
<td>✓</td>
</tr>
<tr>
<td>Viability in Socio-Linguistic Context</td>
<td>✓</td>
</tr>
<tr>
<td>User Compatibility</td>
<td>✓</td>
</tr>
<tr>
<td>Viability in Genesis</td>
<td>✓</td>
</tr>
<tr>
<td>Cultural Identity</td>
<td></td>
</tr>
<tr>
<td>Ecology</td>
<td></td>
</tr>
<tr>
<td>Self-instructive Quality</td>
<td>✓</td>
</tr>
</tbody>
</table>
**Aggregation of Individual Evaluations**

A method to aggregate the information of all the parameters is Multi Attribute Utility Technology which is discussed in further detail in the following chapter. In this method, individuals interacting with the object are identified as stakeholders. Product attributes or values are elicited from the stakeholders and organized in a hierarchical structure called the value tree. The relative importance of each of these attributes is determined through judgments. It is then ascertained how well each object of evaluation serves each value in the value tree. The measures of importance are then aggregated to give rise to a numerical result for the product being tested.
CHAPTER VI
THE TESTING SCENARIO

The models that were created were tested on the basis of the evaluation system. It needs to be noted that the testing done as a part of the thesis is a subset of the total evaluation. Practical limitations of availability of subject population and time for the study prevented execution of a complete evaluation. This evaluation is an example to demonstrate the system of evaluation, and the results have not been checked for validity.

Subjects

In an ideal situation, the testing should be conducted with a large population of the target user group if all products being tested are designed for the same user group. If the products being tested are designed for different user groups, all groups should be equally represented. In this case, the products designed were meant for three different user groups, but the testing was done only with college students for ease of availability of test subjects. In all, twenty five subjects were tested. These subjects were students from courses ID160 and ID199 as they have an interest in design and hence would be motivated to get involved in the experiment. At the same time, they are not granted admission into the industrial design program yet, and are therefore uninitiated to design methodology.
Test results are usually more reliable if the number of subjects is high. Evaluators have to exercise caution and prudence in deciding the sample size for the experiment. This depends upon the problem being researched, the nature of the testing, as well as the methods of analysis. For example, when hypothesis testing is done, there are means of determining the minimum sample size required to make the test worthwhile. In case of multidimensional scaling, it is sometimes essential to have a large number of subjects, and at times it is necessary to have a sizeable number of stimuli. This depends upon the type of matrix being used in the analysis. For protocol analysis, it is possible to get valid data by talking to as few as 10 subjects. Therefore, evaluator discretion is fundamental in deciding the sample size for the experiment. Errors due to sample size should be avoided.

Materials

The four models of the laptop computers made from white foam were used in the experiment. The objects being tested will be referred to as 'evaluands.' A list of instructions was provided to the subjects before the experiment commenced. The questionnaire that was used is discussed in this chapter. Note that the question as posed to the subject is in italics, rest is discussion. Some questions were posed verbally, some required the subjects to write in their responses. At the end of the test, subject information was requested, concerning the amount of familiarity with computers.

Procedure

The testing consists of two parts. The first portion (Questions 1 through 20 in this case) deals with the issue of self-evidence, user understanding of
metaphors, and compatibility with target user group; and is conducted individually with each one of the evaluands. The second part (Questions 21 through 27 in this case) deals with the other parameters discussed in the preceding chapter, and here all objects are treated as a group. Therefore, the experiment design is based on a 'within subjects design.' The procedure was administered individually and took an average of about 35 minutes to complete.

Objects exist within contexts. The same object viewed in different contexts might not be necessarily perceived similarly. Indeed, the environment a product is viewed within, plays an important role in the recognition of the product. This is true of the testing situation too. Environments can be designed and manipulated to assist or impede recognition of objects within it. There could be different types of contexts that the product could be tested in.
1. The Real World Situation: The model of the laptop computer could be tested on a desk in a real office.
2. A Simulated World Situation: An office environment, with the necessary furniture and the related paraphernalia like papers, files, computers etc., could be created for testing.
3. Neutral context: A blank, expressionless environment is referred to as a neutral context here. Testing the laptop computer on a table devoid of any other objects in a seemingly austere setting is an attempt to extract it from the influence of immediate physical surroundings. This, it is expected, would help test whether the product is expressive enough to be recognized despite a neutral background. Researching the influence of the environment on the recognizability of a product would assuredly reveal relevant information, but is not pursued here. Instead, a neutral context is assumed for the experiment. It needs to be noted here that
contexts could also refer to the psychological, cultural, ethnic backgrounds too. However, context here refers to the immediate physical surrounding.

The subject is asked to enter a room, in which one of the models is placed on an otherwise empty table in the closed position. The first question is asked when the subject is at least ten feet away from the table.

**Question 1. Can you say what the object on the table is?**

It is expected that this first question will not only reveal whether the design of the laptop computer is self-evident enough to be recognized at a certain distance, but will also reveal whether a model as rough as this one is recognized as a product at all, or whether it is seen just as a object made out of foam. The word 'object' was used to prevent using the word 'product' as it might give rise to pre-conceived notions about what is being shown in the test. The word 'artifact' was not used for the same reason.

If it is recognized as a laptop computer, further questions should be aimed at eliciting information about specifics. The subject is then asked to approach closer to the model and is asked to view it carefully from all sides.

**Question 2. Now that you see it better, can you say what the object on the table is?**

Two types are problems are anticipated when this question is posed to the subjects. On one hand, they might feel a compulsion to stick to their first response even if they think it is something else on closer inspection. On the other hand, there might be a bias to change simply because the question is asked. It is assumed that these effects will not skew the responses significantly.

If the model is not recognized as a laptop computer, it might be recognized as an electronic/electric device of some sort. One of the intentions of the questions is to discover whether subjects are able to pinpoint the model as a
real world example of a laptop computer, or whether they cannot even tell the superordinate level. If the subject is unable to say what the product is, she/he can be presented alternatives to choose from. Explanations and examples are given along with the category labels, as all the subjects might not be able to understand the differences in the meanings of the terms.

**Question 3. Do you think it is**

a. A mechanical object

   (An object that works due to relative mechanical movement between parts, and is not powered by electricity, like a mechanical typewriter, a manual camera, a mechanical clock, etc.)

b. An electromechanical object

   (An object that is electrically powered, like an electric typewriter, a toaster, etc.)

c. An electronic object

   (An object that is electronically controlled, like a computer, an automatic camera, a camcorder, etc.)

**Question 4. Do you think it is some kind of**

a. Audio equipment

b. Video equipment

c. Computer equipment

d. Kitchen equipment

e. Household appliance

f. Communication equipment

g. Chemical equipment

h. Maintenance equipment

i. Any other ___________
The responses to these questions will show how many subjects recognize the model at the superordinate level as an electronic product, how many recognize it at the basic level as some computer equipment, how many recognize it at the subordinate level as a laptop computer on close inspection, and how many can recognize it from a distance of approximately ten feet. The larger the number of positive responses at the subordinate level, the more self-evident is the stimulus.

**Question 5.** If you were asked to use this object, how would you start? What do you think you would do first?

The subject is not asked directly to open the laptop computer. It needs to be discovered whether or not it is perceived that the object has a lid that needs to be opened.

**Question 6.** What do you think will happen if you press this switch?

The researcher points out to the release catch for the lid

**Question 7.** Now that you see it opened, can you tell me what the object on the table is?

If the object is not recognized by then, the lid is opened and the same question is asked. If it is still not recognized after asking Question 7, the subject is told that it is a laptop computer.

**Question 8.** Do you know what this is?

The researcher points out to the trackball.

**Question 9.** Do you know what this is?

The researcher points out to the ventilator grill.

The testing will attempt to unravel whether or not users perceive the metaphors employed by designers. One of the models will be shown to the subject and they will be asked.
Question 10. Do you know what this is?

The researcher points out to the power switch.

Question 11. How would you adjust the brightness and contrast of the screen?

Question 12. How would you move the cursor on the screen?

Question 13. Does this object remind you of anything else that you have used or seen? It could be humans, products, animals, plants, virtually anything... Consider the object as a whole, and also consider the parts. You could say that the object reminds you of a raccoon, or you could say that this bottom-left corner looks like a raccoon, or you could say both.

If the subject is unable to respond to the question, some clues or hints could be provided to suggest the metaphors employed by the designer.

Question 14. Consider this part of the object (and the researcher points out to the lid {metaphor of file folder} in question). Can you think of some products or things that you use in your daily life, possibly at work, that it resembles?

If this question does not generate any reply, the subject could be offered alternatives to choose from, in the following question.

Question 15. Would you think that this part of the object bears a resemblance to any of these things?

1. File folder
2. Suitcase lid
3. Shopping bag
4. Ziploc bag
5. __________ other

This series of Questions 13 through 15, designed to study the appropriateness and strength of the metaphors, are repeated for the different parts of the model. These responses will give an indication of whether the
designer's intentions are realized through the actions performed by the user. Designers need to know whether users perceive the metaphors employed by them, and if so to what extent; or whether they invariably envision their own metaphors.

**Question 16.** Consider this part of the object (and the researcher points out to the hinges {metaphor of ring binder} in question). Can you think of some products or things that you use in your daily life, possibly at work, that it resembles?

**Question 17.** Would you think that this part of the object bears a resemblance to any of these things?

1. Ring binder
2. Door hinges
3. __________ other

**Question 18.** Consider this part of the object (and the researcher points out to the body of the laptop computer {metaphor of paper stack} in question). Can you think of some products or things that you use in your daily life, possibly at work, that it resembles?

**Question 19.** Would you think that this part of the object bears a resemblance to any of these things?

1. Steps or stairway
2. Stacks of papers
3. __________ other

**Question 20.** As you can see, these laptop computers vary from each other by less or more degrees. Rate the following pairs of laptop computers for the level of difference between them, so that

1 = The two laptop computers are very similar to each other
9 = The two laptop computers are very different from each other

Question 20 is designed to elicit information about the perceived users. The form should help define the user profile. It will also help determine whether products would help express the users' individual identity. Subjects are often wary of mentioning their opinions. An indirect question (Who do you think would like this object?) is expected to elicit a truer response than a direct question (Would you like this object?).

When testing for typicality, the four models can be shown at a time to the subject with the question

**Question 21.** Think of the laptop computers that you have seen or used before, and try to visualize what you think would be a ‘typical laptop computer’ in terms of the shape, color, textures, etc. Can you rank these objects in order of preference from the one that gives the impression of being most like a the typical example, to the one that gives the impression of being least like a typical laptop computer? You can assign numbers to the objects so that

1 = Least like a typical laptop computer
4 = Most like a typical laptop computer

The data from the responses to this question can be analyzed by multidimensional scaling for the quality of 'typicality'. However, it needs to be stressed that this quality of 'typicality' is not necessarily essential. In fact, if all products attempted to stick to their 'typical', there would be no innovative product forms. The product has to strike a balance between being close to the typical and being innovative in form. Question 22 tests for typicality. Question 23 attempts to test for degree of innovation.

**Question 22.** Fill in the blanks so that each of the four groups are matched with the portable computers that the individuals of that group might be most attracted to.
Each age group should be matched with one computer only.

Age group 14-18 yrs. might be attracted to Computer

Age group 19-23 yrs. might be attracted to Computer

Age group 24-28 yrs. might be attracted to Computer

Age group 28 and more yrs. might be attracted to Computer

**Question 23.** Can you rank these objects in order of preference from the one that gives the impression of being the most innovative, to the one that gives the impression of being the least innovative? You can assign numbers to the objects so that

1 = Least innovative in appearance

4 = Most innovative in appearance

This data can be analyzed by multidimensional scaling for the quality of 'Innovativeness.' Over and above 'typicality' and 'innovativeness', recognizability also depends on the roughness of the model. Representing this graphically,

![Diagram](image.png)

Fig. 21. Typicality vs. Innovativeness vs. Model Roughness
As the models being tested are foam models with the same amount of detail, the
degree of roughness is constant, and the innovativeness and typicality are the
major influencing factors.

**Question 24.** Can you rank these objects in order of preference from the one that
gives the impression of being the most playful, to the one that the one that gives the
impression of being the least playful? You can assign numbers to the objects so that
1 = Least playful in appearance
4 = Most playful in appearance

**Question 25.** Can you rank these objects in order of preference from the one that
gives the impression of being the most trendy, to the one that gives the impression of
being the least trendy? You can assign numbers to the objects so that
1 = Least trendy in appearance
4 = Most trendy in appearance

**Question 26.** Can you rank these objects in order of preference from the one that
gives the impression of being the most classy, to the one that the one that gives the
impression of being the least classy? You can assign numbers to the objects so that
1 = Least classy in appearance
4 = Most classy in appearance

**Question 27.** Can you rank these objects in order of preference from the one that
gives the impression of being the most ordinary, to the one that gives the impression
of being the least ordinary? You can assign numbers to the objects so that
1 = Least ordinary in appearance
4 = Most ordinary in appearance

All rank order data will be analyzed by multidimensional scaling.
CHAPTER VII
METHODS OF DATA ANALYSIS

Possible Methods of Evaluation

Presented below, are some techniques drawn from different disciplines which can be used in the evaluation of products.

Hypothesis Testing

A hypothesis is a formal statement or expectation about the outcome of a study. Scientific hypotheses are usually stated in terms of dependent and independent variables, and the relationships between them. Hypothesis testing is a set of inferential procedures, which uses data from samples to establish the credibility of a hypothesis about a population. In the behavioral sciences, formulating theories, stating hypotheses, gathering data to test the hypotheses, revising theories, making new hypotheses, and conducting more research are common practices towards accumulation of knowledge.

Statistical hypotheses are numerical statements regarding the potential outcomes of an experiment. Hypotheses in an experiment are stated in a pair, the null hypothesis denoted by $H_0$, and the alternative hypothesis denoted by $H_1$. The null and alternative hypotheses are mutually exclusive in that only one of them can possibly be true. Once the data gathered from an experiment is analyzed, a decision has to be made whether to believe in the null hypothesis or
the alternative hypothesis.

There are certain tests which are used to analyze data and decide whether or not to reject the null hypothesis. The z test, the t test, the F test, and the Chi square test are examples of such tests. In experimental situations where the responses are categorical, and the data are in the form of frequency counts, the Chi square test can be used in testing hypotheses. The Chi square test is an example of a non-parametric test, because it does not make assumptions about the shape of the population distribution. Therefore such tests are also called distribution-free tests. Parametric tests like the F tests assume that populations are normally distributed and have equal variances. If these assumptions are grossly violated, interpretations of test results can be misleading. When the Chi square statistic is used, a null hypothesis is first stated that indicates what the population data would look like if there is no effect. The frequencies of distribution specified by the null hypothesis are called expected frequencies, and the frequencies obtained from the sample are called observed frequencies.

The frequencies specified in the null hypothesis are determined either rationally or empirically. If there is no preference among subjects for any of the categories, the probabilities are considered equal. If there are two categories, the probabilities for each of them would be 0.5. Therefore the null hypothesis would be,

\[ H_0 : P_1 = P_2 = 0.5 \quad \text{Eq. 1} \]

\[ H_1 : P_1 \neq P_2 \quad \text{Eq. 2} \]

where, \( H_0 = \) null hypothesis

\( P_1 = \) probability of selection of category 1

\( P_2 = \) probability of selection of category 2
Similarly, if there are three categories, the null hypothesis would be,

\[ H_0 : P_1 = P_2 = P_3 = 0.33 \]  

Eq. 3

where, \( H_0 \) = null hypothesis

\[ P_1 = \text{probability of selection of category 1} \]
\[ P_2 = \text{probability of selection of category 2} \]
\[ P_3 = \text{probability of selection of category 3} \]

These are examples of the rational approach. Another example of the rational approach would be when the distribution of the expected frequencies is predicted by a theory. In the empirical approach, the expected frequencies are specified by a theory. In the empirical approach, the expected frequencies are specified on the basis of existing data or past research. The purpose of the goodness-of-fit test is to determine if the observed frequencies obtained from a sample of subjects differs from expected frequencies. The Chi square allows the researcher to determine the probability that the difference between the observed and the expected frequencies is due to chance. The computation of the Chi square is explained in Chapter IX titled ‘Analysis and Conclusions.’

\[ \chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e} \]  

Eq. 4

As the difference between the observed and expected frequencies increases, the value of Chi square increases. A small value of \( \chi^2 \) suggests a good fit between the observed and expected frequencies. This calculated value of \( \chi^2 \) is compared to the critical value of \( \chi^2 (\chi^2_{\text{crit}}) \). If \( \chi^2_{\text{obs}} \) is greater than \( \chi^2_{\text{crit}} \), the null hypothesis is rejected. The value of \( \chi^2 \) depends on the Chi square distribution and the value of the degree of freedom. The degree of freedom is the number of categories less one. The Chi square distribution is a theoretical distribution formed by taking an infinite number of samples in which the Chi
square statistic is computed for each sample. The relative frequency of each value of Chi square is plotted to show the Chi square distribution. The mathematical underpinnings of the Chi square distribution are not discussed here.

There are certain problems in which it is necessary to decide whether a difference between two sample proportions is due to chance, or whether it is indicative of the fact that corresponding proportions are indeed unequal. The test used in such cases is referred to as the z test. If \( x_1 \) and \( x_2 \) are the numbers of the successes obtained in \( n_1 \) trials and \( n_2 \) of another, the trials are all independent, and the corresponding probabilities of a success are, respectively \( p_1 \) and \( p_2 \), then the sampling distribution of \( x_1/n_1 - x_2/n_2 \) has the mean \( p_1 - p_2 \). Since the value of \( p \) is not known, it is estimated by pooling the data.

\[
p = \frac{x_1 + x_2}{n_1 + n_2}
\]

Eq. 5

The statistic \( z \) can be calculated by the formula

\[
z = \frac{x_1/n_1 - x_2/n_2}{\sqrt{p(1-p)(1/n_1 + 1/n_2)}}
\]

Eq. 6

The null hypothesis is rejected if the value of \( z \) is less than or equal to \(-z_a\), or if the value of \( z \) is more than or equal to \( z_a \).

**Multidimensional Scaling (MDS)**

Multidimensional scaling is a set of mathematical procedures that enables a researcher to uncover the 'hidden structure' of data bases. The basic concept of this scaling method can be best explained by an example. Consider a geographical map of an area showing locations of various cities. If one was asked to construct a table of distances between cities, it could be easily done by
measuring the distances off the map and converting them to the actual distances by using the scale. If this problem was posed in the reverse order, where a table of distances was given and the task was to construct a map, it could be solved by multidimensional scaling. Geometric procedures could also be used but would require much more effort and time. The input to multidimensional scaling is a set of 'proximities' among any kinds of objects. A proximity is the measure of similarity or dissimilarity between the objects. The output is a spatial representation, consisting of a geometric configuration of points, such that each point corresponds to one object or stimulus. This configuration makes the data easier to comprehend because it is a graphic representation. The larger the dissimilarity between objects, the further apart they are represented in the spatial map.

This spatial representation could be in two, three, four, or more dimensions depending on the complexity of the data. The analysis of the configuration is usually done by examining the arrangement of points which represent the objects. There are other analysis and interpretation methods which are often used to supplement direct examination.

**Methods of Obtaining Data**

A common procedure for obtaining proximities data is to ask people to judge the 'psychological distance' or 'closeness' of the stimulus objects. 'Similarity or 'dissimilarity' are most frequently used to elicit proximity judgments, however other words like association, complementarity, relatedness, etc. are often used. If the number of stimuli is low, each of the stimulus objects can be paired with every other in the group, and the subjects can be asked to
judge the pairs. The data obtained from this technique are referred to as pair comparisons data. For a larger number of stimuli, the method of rank ordering can be used, wherein all the stimulus objects are ranked in order of preference. For large stimulus sets, subjects can be asked to sort or cluster the stimuli according to perceived similarity. The stimuli are placed in categories so that there is more similarity amongst stimuli in one category than there is between stimuli in different categories. A matrix of proximities can be generated by counting the number of times each pair of stimuli is put in the same category.

**Analysis and Interpretation**

The proximities data result in a two way matrix, which forms the basic input for the multidimensional scaling.

As the method of scaling is mathematically extremely complex, it is almost always executed by computer programs. Mathematically it is possible to carry out multidimensional scaling not only in two- and three-dimensional space but also in R-dimensional space, where R = 1, 2, 3, 4, ... In an ordinary sense, four-dimensional space impossible to visualize, but there are numerous ways to express it visually (though imperfectly), and there is no difficulty in dealing with it mathematically.

As the purpose of MDS is to discover the hidden structure in data sets, the subjects are not given specific instructions about characteristics on which the similarity judgments are to be based. This is the information which has to be discovered, and not imposed. However, they can be asked to base their decisions on specific kinds of similarities, like cultural similarity, political similarity, etc.
The choice of axes is usually based on direct observation. However, this can be handled more objectively by the use of multiple regression analysis. In this technique, along with the similarity judgments, the subjects are also asked to rate each of the stimuli on bipolar scales. The mean values of these ratings are then used in identifying the dimensions in the spatial configuration.

The data obtained from the rank ordering is usually analyzed by multidimensional scaling with the help of a computer program called MD_PREF (MultiDimensional Preference Scaling) which provides an internal analysis of preference data. From the data, the program positions the stimuli as points in Euclidean space and each subject is represented by a vector to the region where the subject's highest preference lies. The stimuli are represented as points in the same space, so that the projections of the stimuli on a given subject's vector maximally reproduce her/his preferences.

**Verbal Reports and Protocol Analysis**

The human being is at the center of things because it is for us that the products are designed, it is we who will be using the products, and we will evaluate them. To understand how human beings relate to products around them, how and what they read into them are questions which need to be addressed. Watching and talking to people gives valuable information and therefore the question of verbal information is extremely relevant here. The method outlined below is used specifically in analyzing verbal information.

Protocol analysis is a method used for analyzing data gathered from what are commonly known as interviews. The raw data consists of verbal information
usually on a tape recorder, which is transcribed, segmented, encoded and
analyzed. These methods are described below. Verbal protocols are a source of
enormous information, and can be classified into three different forms such as,

- Thinking-aloud (TA) protocols
- Retrospective responses to specific probes
- Introspective reports of trained observers

Verbal reports were initially treated as "soft" data as against "hard" data,
the difference between the two being the amount of subjectivity and
correspondence to observed behavior. Since the introduction of tape recorders
into data gathering, it has become a standard procedure to make transcripts of
recorded tapes, and thus preserving it as "hard" data. In producing the
transcript, some selection is required, some information is eliminated and this
process is referred to as preprocessing. These preprocessed data are then edited
and encoded so that they can be used to test theories or make predictions. In
this process, coding categories are determined, coding assessments are made
and the preprocessed statements are encoded into the terminology of the
theoretical model. Verbal protocols can be analyzed in two different ways. One
method does not require the analysis of meanings, because the subject and the
experimenter have pre-decided specific symbols for communication. These
symbols could be speech signals or button presses (a subject could say 'cef'
instead of 'yes'). To analyze these verbalizations, the experimenter has to only
categorize each signal into the pre-decided category. In the other method of
analysis, the observed verbalizations are analyzed in terms of their meanings.
Again, the protocols are encoded and thus limited to certain selected aspects
and features rather than the full meaning. Each stimulus or instance can be represented as a unique combination of features. These features can be combined to represent a concept. Although there may be a large number of concepts in a protocol, it would be much more limited than the amount of variability existing in natural language. Therefore, an incorporation of a model or theory aids in limiting the range of possible interpretations allowing the analysis to be selective. Many analyses do not fit into a scheme and in such cases, the search for interpretations proceeds parallel to the search for a processing model.

The Processing Model

The processing model helps in interpreting verbal data and this assumes that human cognition is information processing. A cognitive process is seen as a sequence of internal states successively transformed by a series of information processes. Information is stored in different types of memories— a short term memory (STM) with limited capacity and intermediate duration and a long term memory (LTM) with a large capacity and relatively permanent storage. Information recently acquired is stored in the STM and is available for producing a verbal report, but information from LTM has to be retrieved and transferred to STM before it can be reported. There are other aspects to the processing model, namely cognition, control of attention, fixation and automation which will not be discussed here.

Thinking-aloud Protocols

In this kind of verbal reports, the cognitive process described as successive states of heeded information are verbalized directly. Subjects are asked to express their thoughts directly as they emerge. The advantage here is that naive and
inexperienced subjects can be used in experiments.

**Retrospective Reports:**

Retrospective reports can be based on either short term or long term memory. When a task is being accomplished, a durable track of the information being attended to is laid down. After the task is done, this track can be accessed from the STM and verbalized. Parts of it can be retrieved from LTM but that would involve an additional process of retrieval and might cause loss of information. Ideally, a retrospective report should be asked of a subject immediately after completion of the task. If the investigator is interested in a particular aspect of the subject's behavior, then specific probes can be constructed to induce the subject to generate the desired information.

**Introspective Reports:**

Introspection, of the direct observation of the mind in operation was the primary method of investigation in the early years of psychology. However, it was often accused of being non-scientific and therefore it changed form over the years becoming more systematic and rigorous. Subjects are asked to report on their thought processes in relation to a certain experience of task. There is a lot of subjectivity involved in these reports but they can also be very informative.

**Segmentation and Encoding Processes:**

Once the verbal report is transcribed, it is taken up for segmentation. The protocol is broken down in such a way that each segment represents a statement. If the protocols are completely grammatically correct, each segment would be a clause or a sentence, but in normal speech single words of
incomplete phrases are often used in place of entire sentences. Each segment is then encoded. The coding system depends on the type of problem but usually involves these four types of statements:

Intentions: Verbs like ‘shall’, ‘will’, ‘have to’ represent goals and future states.
Cognitions: Information based on attention to selected aspects of the current situations are coded as cognitions.
Planning: Intermediate constructions which mentally explore possibilities, like ‘If X then Y, if Z...’ are planning sentences.
Evaluations: Keywords like ‘no’, ‘yes’, ‘dammit’, fine’ are evaluations.

There could be specific coding systems for specific tasks like problem-solving in chess, logic problems, math and thermodynamic problems etc. The statement-by-statement encoding retains most of the semantic content of the protocol. It also retains all of the idiosyncrasies of individual behavior. This is then analyzed by the experimenter. The protocol can be compared with a detailed computer simulation of the behavior and a matching can be observed. It can also be used to test behavioral theories, or to study commonalities of behavior in a group. This method of analysis has a wide range of applications and can be extremely versatile. Human behavior in various situations and environments can be studied. The process of decision making and problem solving can also be scrutinized in detail. Computers are now being used in protocol analysis and a program called Mini Protocol Analysis System is now available which aids in encoding segments. It facilitates presenting the protocol statements to the experimenter and also makes the encodings directly available for further analysis on the computer.
Multi Attribute Utility Technology (MAUT)

This is a technique which is used for social and educational program evaluation. It requires representatives of the program to identify the most relevant values or attributes that constitute the program. The basic steps in MAUT are:

1. Identification of the objects of evaluation and the function or functions that the evaluation is intended to perform. Evaluations should help the process of decision making and the nature of the decisions to be made controls the objects of evaluation.
2. Identification of stakeholders, individuals or groups who have a stake or an intent in the program and who are important enough that their opinions should be considered.
3. Elicitation of the relevant attributes or value structures from the stakeholders and organization of these into a value tree. The stakeholders are asked to identify topics that are relevant to the decision. Once these are known, they are arrayed in an organizational hierarchy called the value tree.
4. Assessment of the relative importance of the values identified by each of the stakeholders/groups. This is known as weighting of values and is an essential process, as not all the attributes carry equal importance and this difference needs to be brought out in the evaluation.
5. Assessment of how well each object of evaluation serves each value at the lowest end of the value tree. These are called location measures or utilities and could be measurements or expert opinions.
6. Aggregation of location measures with measures of importance. The values derived in step 4 and step 5 are aggregated into one composite using certain
rules and mathematical formulae. Therefore the two sets of numbers are now merged into one.

7. Sensitivity analysis is a method which consists of changing some of the numbers that were input into the MAUT analysis and doing it again to see the change in results.

The result of MAUT is a numerical value for each of the objects of evaluation and thus facilitates the decision making process. In this system of evaluation, MAUT has been modified and used to aggregate data obtained from the questionnaire into a numerical value for each of the evaluands.

**Geography and Cartographic Analysis**

Geographic inquiry has two approaches- humanistic and behavioral, but a number of concepts recur throughout geographic work regardless of the approach. Some of these are distance, direction, shape, size, location, scale, etc. For example, distance can be treated as actual distance on a geographical map, and also in terms of individual perceptions (far or close), and it can refer to space as well as human relations. Cartography is a technique used by geographers for both, data analysis as well as communication. Any type of data- surveys, interviews, observations, tests can be mapped if they can be quantified or put into qualitative categories.
Fig. 22. Mental Maps Drawn by Dave, Ernest, and Ralph (from top to bottom).
Mental maps can be used to understand and analyze people’s perceptions and mental images of objects, artifacts, environments, events, etc. The figure shows three mental maps drawn by Black children of their neighborhood, which has White and Black population on the two sides of Parker Street. Ernest’s map emphasizes the street which is a kind of a territorial barrier between the two areas. Dave’s map shows less of the street, has a lot of detail of the area he lives in, but no detail of the White dominated Mission Hill area. It can thus be seen how these children perceive their environment.

A drawing recall test in which subjects are asked to make a sketch of an object they have just seen or used, can give similar information of how people perceive products.

**Analysis of Water Color Painting**

Composition of the image can be regarded as an important element in a painting. Described below is a technique of evaluation based on the analysis of composition in a painting. Eight elements which make up a composition are identified. These are unity, dominance, conflict, repetition, alteration, balance, harmony and gradation. Also defined are seven techniques which can be employed by the artist to achieve these compositional elements. These techniques are line, color, value, shape, texture, direction and size. The eight elements and seven techniques give rise to a 8X7 matrix in which the elements of design are the variables. Every composition has varying degrees of unity and/or balance and/or harmony and so on. Construction of such a matrix creates a heuristic tool as shown in Figure 23.
The fifty six cells in the matrix represent directions of thinking, it gives an idea of all the possible design solutions to the problem of composition. This matrix has been used to serve as a heuristic device in the composition of an evaluation in the form of a three dimensional figure with multiple layers, tiers, and columns. The author considers evaluation as applying a set of value depicting variables to a group of options and therefore the vertical dimension is labeled, options, the horizontal as value depicting variables, and the third as decision maker levels. The options are those entities which are to be evaluated, the variables are the parameters against which the options are evaluated, and the decision maker levels can be seen in terms of ‘participant’, ‘institution’ and ‘society’. In an evaluation, the options are defined first, the variables next and the decision maker levels last.

![Design Principles]

**Fig. 23. 8 X 7 Matrix for Evaluating Water Color Paintings**
CHAPTER VIII
ANALYSIS AND CONCLUSIONS

The test was run with twenty eight subjects, but the data from the first three subjects were used to modify and remove major bugs in the questionnaire that might have been overlooked. For example, it was found necessary to add a question at the stage when the computer is opened, as to what they thought it was. Certain phrases and words were changed to prevent undue stress on the subject. For example, "Do you think that..." was changed to "Does it give an impression of..."

It was mentioned earlier that the evaluation would include both qualitative and quantitative results. As all parameters are not predisposed to numerical analysis, qualitative methods will be required to maintain the richness of data. Therefore, the quantitative results will include information that is quantifiable, including results from MAUT, MDS and hypothesis testing; and the qualitative results will have information from focus groups, protocol analysis, content analysis, etc.

**Qualitative Results**

Certain conclusions can be drawn about the this methodology of testing in general, and some conclusions can be drawn about the actual evaluation.

Testing with models does not permit an exhaustive evaluation of design.
The more finished a model, the larger are the number of semantic attributes that can be tested with it. In the testing of the laptop computers, some of the subjects when asked what the object was, said it was a "...foam book...", or a "...notebook made from Styrofoam..." The material of the model overshadowed the fact that it was a representation of some object. The metaphor of the book with a ring binder that was used in the design was recognized by all subjects. There were however minor variations within it which were interesting. Some of these are listed below.

1. Open book
2. Book cut into half
3. A bound telephone book
4. An old big notebook

Only one subject was able to recognize the model as a laptop computer from up close (Question 2). 6 subjects could tell that it was computer equipment of some sort, and 7 subjects could tell that it was an electronic product. All subjects were able to recognize the model when it was opened. This was probably due to the fact that the keyboard is a strongly identifiable element that helps in the recognition of the computer.

**Quantitative Methods**

**Hypothesis Testing**

Various hypotheses can be made depending upon which attributes of the evaluands need investigation. The hypotheses that were made here are listed below.
Recognition of the object from a distance and up close

It is assumed that when the evaluand is first displayed to the subject and they are asked what the object might be, they might perceive it as a notebook or an open book, which are the metaphors employed; or they might be more specific and see the metaphor of the file folder; or they might perceive it as some kind of a product. The same three responses are possible when the object is shown up close.

<table>
<thead>
<tr>
<th>Question 1</th>
<th>Question 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Metaphor</td>
<td>General Metaphor</td>
</tr>
<tr>
<td>Specific Metaphor</td>
<td>Specific Metaphor</td>
</tr>
<tr>
<td>Product</td>
<td>Product</td>
</tr>
</tbody>
</table>

Fig. 24. Possible Responses to Question 1 and 2

The arrows indicate that there are nine possible responses from the subjects to questions 1 and 2 (Refer to Question 1 and 2).

There are three possibilities,

\[ P_1 = \text{The response to Question 2 as subjects come closer to the evaluand, changes to more specific from general metaphor, or from general metaphor to product.} \]
Fig. 25. Change of Response from General to Specific

\[ P_2 = \text{There is no change in response as subjects come closer to the evaluand.} \]

Fig. 26. No Change in Response
\( P_3 = \) The response to Question 2 as subjects come closer to the evaluand changes from more specific metaphor or product to general metaphor.

![Diagram](image)

**Fig. 27. Change in Response from Specific to General**

Without additional knowledge about subjects' decision making process, we hypothesize that the probabilities of \( P_1 \), \( P_2 \), and \( P_3 \) are equal. Therefore, the null hypothesis can be represented as,

\[
H_0 : P_1 = P_2 = P_3 = 0.33 \quad \text{Eq. 7}
\]

The alternative hypothesis \( H_1 \) can be represented as

\( H_1 : \) At least one probability is different from the others

This hypothesis will be tested using the Chi-Square test for goodness of fit.

Here, \( \alpha = 0.05 \).

**Calculations**

Number of subjects, \( N = 25 \)

Number of categories, \( C = 3 \)

The expected frequency will be the same for all categories
\[ f_e = \frac{N}{C} = \frac{25}{3} = 8.33 \quad \text{Eq. 8} \]

The observed frequencies for the three categories were

\[ f_0 = 15 \text{ for } P_1, \quad 10 \text{ for } P_2, \quad \text{and } 0 \text{ for } P_3 \]

\[ \chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e} \quad \text{Eq. 9} \]

Therefore, \( \chi^2 = 14 \)

The degree of freedom is \( df = C-1 = 3-1 = 2 \) \( \text{Eq. 10} \)

The value of a being 0.05, the critical value of \( \chi^2 \) is 5.99. As the observed value of \( \chi^2 \) is greater than 5.99, the null hypothesis is rejected. This implies that the probabilities cannot be assumed to be equal. It can therefore be interpreted that subjects will tend to be more specific in their responses when they view the evaluand from up close.

**Recognition of Components**

*There is equal probability of recognition of the trackball, the ventilator grill, and the power switch.* (Refer to Questions 8, 9, and 10)

The reason for assuming equal probabilities is that on analysis if it shows that one of the components has a fairly low probability, there is a definite recognition problem that needs to be addressed for that component. Therefore, the null hypothesis can be represented as

\[ H_0 : P_1 = P_2, \quad H_0 : P_2 = P_3, \quad H_0 : P_1 = P_3 \quad \text{Eq. 11} \]

\[ H_1 : P_1 \neq P_2, \quad H_1 : P_2 \neq P_3, \quad H_1 : P_1 \neq P_3 \quad \text{Eq. 12} \]

where \( P_1 = \text{Probability that the track ball is recognized by subjects} \)

\[ P_2 = \text{Probability that the ventilator grill is recognized by subjects} \]

\[ P_3 = \text{Probability that the power switch is recognized by subjects} \]

This hypothesis will be tested using the \( z \) test.
Here $\alpha_{\text{test}} = 0.01$ and $\alpha_{\text{expt}} = 0.03$.

**Calculations**

Number of subjects, $N = 25$

Number of categories, $C = 3$

The observed frequencies for the three categories were

$$f_0 = 24 \text{ for } P_1, \; 1 \text{ for } P_2, \; \text{and } 6 \text{ for } P_3$$

The $z$ test is used to compare two probabilities. $P_1$ will be compared to $P_2$, $P_2$ will be compared to $P_3$ and $P_3$ will be compared to $P_1$.

**Comparing $P_1$ to $P_2$,**

$$H_0 : P_1 = P_2 \quad \text{Eq. 13}$$

$$x_1 = 24, \; x_2 = 1, \; n_1 = 25, \; n_2 = 25$$

$$p = \frac{x_1 + x_2}{n_1 + n_2} \quad \text{Eq. 14}$$

$$= 0.5$$

$$z = \frac{x_1/n_1 - x_2/n_2}{\sqrt{p(1-p)(1/n_1 + 1/n_2)}} \quad \text{Eq. 15}$$

$$= 6.5$$

As $z$ is more than 2.17, the null hypothesis is rejected. Therefore it can be concluded that $P_1$ and $P_2$ may not be equal.

**Comparing $P_2$ to $P_3$,**

$$H_0 : P_2 = P_3 \quad \text{Eq. 16}$$

$$x_2 = 1, \; x_3 = 6, \; n_2 = 25, \; n_3 = 25$$

$$p = \frac{x_1 + x_2}{n_1 + n_2} \quad \text{Eq. 17}$$

$$= 0.14$$

$$z = \frac{x_1/n_1 - x_2/n_2}{\sqrt{p(1-p)(1/n_1 + 1/n_2)}} \quad \text{Eq. 18}$$
\[ z = -2.03 \]

As \( z \) is more than -2.17, the null hypothesis is retained. Therefore it can be concluded that \( P_2 \) and \( P_3 \) may be equal.

Comparing \( P_1 \) to \( P_3 \),

\[ H_0 : P_1 = P_3 \quad \text{Eq. 19} \]

\[ x_1 = 24, x_3 = 6, n_1 = 25, n_3 = 25 \]

\[ p = \frac{x_1 + x_2}{n_1 + n_2} \quad \text{Eq. 20} \]

\[ = 0.6 \]

\[ z = \frac{x_1/n_1 - x_2/n_2}{\sqrt{p(1-p)(1/n_1 + 1/n_2)}} \quad \text{Eq. 21} \]

\[ = 5.19 \]

As \( z \) is more than 2.17, the null hypothesis is rejected. Therefore it can be concluded that \( P_1 \) and \( P_3 \) may not be equal.

On studying the results of the \( z \) test, it is obvious that \( P_1 \), \( P_2 \) and \( P_3 \) do not have equal probabilities. However, the pair \( P_1 \) and \( P_2 \) are comparable. It can be concluded from the equal probabilities of \( P_1 \) and \( P_2 \), and the frequency count for the ventilator grill and power switch that these two elements are not recognized and that need redesign.

**Perception of Individual Metaphors**

a. *There is a 0.2 probability that the metaphor of the file folder will be recognized for the profile of the lid.* (Refer to Question 14 and 15)

Question 14 is used as a reference to study the expressiveness of the metaphor, but the analysis is performed with the response of Question 15, as it has definite categories and is hence easy to quantify in terms of probabilities. As
there are 5 categories, the probability of each response is 0.2.

Out of 25 subjects, 4 subjects could answer the open ended Question 14 correctly, and were hence not asked Question 15. Of the remaining,
Number of subjects who selected Category 1 = 3
Number of subjects who selected Category 2 = 17
Number of subjects who selected Category 3 = 0
Number of subjects who selected Category 4 = 1
Number of subjects who selected Category 5 = 0

\[ H_0 : P_1 = P_2 = P_3 = P_4 = P_5 = 0.2 \]  \hspace{1cm} \text{Eq. 22}

where \( P_1 \) = Probability of recognition of the metaphor of the suitcase lid
where \( P_2 \) = Probability of recognition of the metaphor of the file folder
where \( P_3 \) = Probability of recognition of the metaphor of the shopping bag

where \( P_4 \) = Probability of recognition of the metaphor of the ziploc bag
where \( P_5 \) = Probability of recognition of the metaphor of other category

\( H_1 : \) at least on eprobability is different than the others

This hypothesis will be tested using the Chi-Square test for goodness of fit.

Here, \( \alpha = 0.05 \).

Calculations

Number of subjects, \( N = 21 \)
Number of categories, \( C = 5 \)

The expected frequency will be the same for all categories

\[ f_e = N/C = 21/5 = 4.2 \]  \hspace{1cm} \text{Eq. 23}

The observed frequencies for the three categories were

\[ f_o = 3 \text{ for } P_1, 17 \text{ for } P_2, 0 \text{ for } P_3, 1 \text{ for } P_4, 0 \text{ for } P_5 \]
\[ \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \]  
Eq. 24

= 50.95  
Eq. 25

The degree of freedom is \( df = C-1 = 5-1 = 4 \)

The value of \( \chi^2 \) being 0.05, the critical value of \( \chi^2 \) is 9.49. As the observed value of \( \chi^2 \) is greater than 9.49, the null hypothesis is rejected. This implies that there is a higher probability that the metaphor of the file folder is recognized. Therefore, the alternative hypothesis is accepted. It needs to be noted here that though this methodology can be successfully applied, the value of \( f_e \) should be equal to or more than 5. The value of \( f_e \) is 4.2 here, but this problem can be overcome by testing a larger number of subjects.

b. There is a 0.33 probability that the metaphor of the ring binder will be recognized for the hinges of the lid. (Refer to Questions 16 and 17)

As there are 3 categories, the probability of each response is 0.33.

Out of 25 subjects, 22 subjects could answer the open ended question 16 correctly, and were hence not asked question 17. Of the remaining,

Number of subjects who selected Category 1 = 1

Number of subjects who selected Category 2 = 2

Number of subjects who selected Category 3 = 0

\( H_0 : P_1 = P_2 = P_3 = 0.33 \)  
Eq. 26

\( H_1 : \) at least one probability is different from the others

where \( P_1 = \) Probability that the metaphor of the hinges will be perceived by subjects

where \( P_2 = \) Probability that the metaphor of the ring binder will be recognized by subjects

where \( P_3 = \) Probability that the subjects will perceive a metaphor other
than the ring binder or door hinges.

This hypothesis will be tested using the Chi-Square test for goodness of fit. Here, $\alpha = 0.05$.

Calculations

Number of subjects, $N = 3$

Number of categories, $C = 3$

The expected frequency will be the same for all categories

\[ f_e = \frac{N}{C} = \frac{3}{3} = 1 \quad \text{Eq. 26} \]

The observed frequencies for the three categories were

\[ f_o = 1 \text{ for } P_1, \ 2 \text{ for } P_2, \ 0 \text{ for } P_3 \]

\[ \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \quad \text{Eq. 27} \]

\[ = 2 \]

The degree of freedom is

\[ df = C-1 = 3-1 = 2 \quad \text{Eq. 28} \]

The value of a being 0.05, the critical value of $\chi^2$ is 5.99. As the observed value of $\chi^2$ is less than 5.99, the null hypothesis is retained. This leads to the interpretation that there are equal probabilities of recognition the hinges as the metaphor of ring binders, door hinges, or something other than these two. Here too, as in Case a above, the value of $f_e$ is less than 5, and more subjects should be tested in order to get valid results.

c. There is a 0.2 probability that the metaphor of the ring binder will be recognized for the hinges of the lid. (Refer to Question 18 and 19)

As there are 3 categories, the probability of each response is 0.33.

Out of 25 subjects, 9 subjects could answer the open ended Question 18 correctly, and were hence not asked Question 19. Of the remaining,

Number of subjects who selected Category 1 = 10
Number of subjects who selected Category 2 = 6
Number of subjects who selected Category 3 = 0

\[ H_0 : P_1 = P_2 = P_3 = 0.33 \quad \text{Eq. 29} \]

where \( P_1 \) = Probability of recognition of the metaphor of the steps
where \( P_2 \) = Probability of recognition of the metaphor of the paper stack
where \( P_3 \) = Probability of recognition of the metaphor of other category

\[ H_1 : \text{at least one probability will be different than the others} \]

This hypothesis will be tested using the Chi-Square test for goodness of fit.

Here, \( \alpha = 0.05 \).

Calculations

Number of subjects, \( N = 16 \)
Number of categories, \( C = 3 \)

The expected frequency will be the same for all categories

\[ f_e = \frac{N}{C} = \frac{16}{3} = 5.33 \quad \text{Eq. 30} \]

The observed frequencies for the three categories were

\[ f_o = 10 \text{ for } P_1, 6 \text{ for } P_2, 0 \text{ for } P_3 \]

\[ \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \quad \text{Eq. 31} \]

\[ = 9.505 \]

The degree of freedom is \( df = C-1 = 3-1 = 2 \quad \text{Eq. 32} \)

The value of a being 0.05, the critical value of \( \chi^2 \) is 5.99. As the observed value of \( \chi^2 \) is more than 5.99, the null hypothesis is rejected. Therefore the probabilities of recognition of the metaphors of the steps, stack of papers, or something other than these two, are not equal.
Many other hypotheses can be drawn up depending upon the information desired. This list of hypotheses can be specifically designed for each evaluation.

**MultiDimensional Scaling**

The rank order data from the testing is analyzed using scaling software. The rank order data is converted into a proximities matrix by a PC version of MDPREF (Multidimensional Preference Scaling) called PCPREF. This software is an adaptation of the main frame statistician program MDPREF and is written for the PC by Prof. Tom Nygren at The Ohio State University. This analysis gives a graphic representation of the evaluands in Euclidean space. The analysis was done by creating an input file consisting of a matrix of the stimuli against the attributes. The rows of the matrix represented the computers A, B, C, and D (normally represented by subjects), and the columns represented the attributes (normally represented by stimuli). When the program was run, an output file was generated which contained the results of the scaling. This file is included in the appendix. It generates a population matrix of the subjects which give the subject coordinates for the dimensions specified in the input file. Similarly, the stimulus matrix gives the stimulus coordinates for the dimensions specified. These values were used in creating two and three dimensional graphs representing the computers and attributes in Euclidean space. The scaling was done in two and three dimensions, and the graphs were generated using a statistical program called SYSTAT for the Macintosh. It may be noticed that the three computers which are similar to each other, are found close together in space. However, Computer D was perceived as very different from the rest and is hence seen distant from the other three computers.
A = Laptop computer for young professionals
B = Laptop computer for college students
C = Laptop computer for high school kids
D = Replica of NCR Laptop computer
1 = Attribute ‘typicality’
2 = Attribute ‘innovativeness’
3 = Attribute ‘playfulness’
4 = Attribute ‘trendy’
5 = Attribute ‘classiness’
6 = Attribute ‘ordinariness’

Fig. 28. Scaling in Two Dimensions
Fig. 29. Scaling in Two Dimensions with Attribute Vectors

Fig. 30. Scaling in Three Dimensions
The graphs show how differently Computer D is perceived compared to Computers A, B, and C. The attributes as well as the computers are represented in Euclidean space. If vectors are drawn from the attribute points to the origin, it gives an idea of how the stimuli are rated on these attributes. For example, it can be seen that Computer D rates highly on Typicality and Ordinariness, but Computers A, B, and C rate highly on the Innovativeness and Trendiness scales. Three dimensional plots are more complex to visualize as compared to two dimensional plots, and therefore also more complex to analyze.

**Numerical Output for the Evaluands Using MAUT**

The technique referred to as Multi Attribute Utility Technology (MAUT) is described in detail in Chapter 7. The value tree shown below should exemplify the process of derivation of the various weights and location measures. The first step involves identification of the attributes that need to be evaluated. In the case of the laptop computers, these are self-evidence, comprehension of metaphors, compatibility of metaphors, and expressiveness of the evaluand. This list will vary for each evaluation, as the evaluator can be selective in deciding which attributes need to be considered. This will also be governed by the medium of presentation of the evaluand. For example, if the testing is being done with a blue foam model, the attribute of ‘Informationally Adequate and Timely Feedback’ cannot be considered, as evaluation of that parameter requires a working product or prototype.

In MAUT, there is a strong emphasis on the stakeholders in the evaluation. Many of the decisions are made by the stakeholders. The attributes and weights are decided by the stakeholders and modified by the evaluator. In this case the stakeholders will be the industrial designer, the product development engineer,
the marketing specialist, the financier, the recycler, the physical user. All those individuals or groups actively involved in the product development cycle should be part of the list of stakeholders. The concept of stakeholders is similar to that of participants explained earlier.

The active attributes are defined by the evaluator on the basis of those defined in the system. All stakeholders are asked to give weights to the attributes. The weights are scrutinized by the evaluator. It is the decision of the evaluator as to how these weights are to be integrated. An average of all weights specified by the stakeholders can be assumed as the final weight. In the case of the laptop computer, these figures are assumed as no actual stakeholders were involved. The principal attributes are then broken down into its components to simplify and facilitate measurement. For example, consider the second attribute of 'Comprehension of Metaphors.' This principal attribute is divided into four second-level attributes, namely comprehension of the overall metaphor, comprehension of the metaphor of a file folder, comprehension of the metaphor of a ring binder, and comprehension of the metaphor of a stack of papers. These again have been assigned weights. This weighting should also be done by stakeholders and modified by the evaluator. It may be noted that the weights within one ‘twig’ of the value tree add up to one.

The next step is to decide the location measures which are also referred to as utilities. This is where the data from the testing is used. These are typically measurements or expert judgments. All location measures occur on a common scale, which is usually 0 to 100. Consider the second-level attribute of comprehension of the overall metaphor. On a scale of 0 to 100, 0 would mean that the evaluan d affords absolutely no comprehension of the metaphor, whereas a location measure of 100 signifies that the metaphor is comprehended
by the entire sample population. A frequency count is done for all the attributes evaluated. The individual location measures are calculated by simple arithmetic. If 12 out of 25 subjects are able to comprehend the overall metaphor, the location measure on extrapolation to 100 would be 48. These location measures are calculated for all the active attributes.

The final summation is a product of the location measure and the importance weight of the attribute across all the levels of the value tree. If

\[ L_{ij} = \text{Location measure for attribute } i_j \]

\[ W_{ij} = \text{Importance weight for attribute } i_j \]

\[ S = \text{Final sum of evaluation} \]

\[ i = \text{The number of levels in the value tree} \]

the final sum would be mathematically represented as,

\[ S = \sum_{i=1}^{n} (L_{ij} \times W_{ij}) \quad \text{Eq. 33} \]

The importance weight is the product of all the weights for that attribute multiplied from the principal to the lowest level in the value tree.

In this case, for the attribute, ‘Comprehension of Metaphors’, the final importance weight for the attribute comprehension of overall metaphor will be

Importance weight for comprehension of overall metaphor = 0.20 \times 0.40 = 0.08

The final weight for this attribute is the product of the final weight and the location measure = 0.08 \times 100 = 8

Similarly,

Importance weight for comprehension of metaphor of file folder = 0.20 \times 0.20 = 0.04

Final weight for this attribute
= 0.04\times 84 = 3.36

Importance weight for comprehension of metaphor of ring binder
= 0.20\times 0.20 = 0.04
Final weight for this attribute
= 0.04\times 96 = 3.84

Importance weight for comprehension of metaphor of stack of papers
= 0.20\times 0.20 = 0.04
Final weight for this attribute
= 0.04\times 60 = 2.4

The utility for the attribute 'Comprehension of Metaphors' is
= 8 + 3.36 + 3.84 + 2.4
= 17.6

The value tree shown on the following page gives a pictorial representation of all the attributes that were used in computing the utility of Computer A. The other utilities were calculated similar to the 'Comprehension of Metaphors.'
Self-evidence = 11.68
Semantic Accommodation = 7.6
Compatibility With Target User Group = 4
Expressiveness = 5.99
The final utility for Computer A is the sum of all the utilities of all the attributes.
Final Utility for Computer A = 46.87
The location measures for the attributes of Self-evidence and Semantic Accomodation are calculated using the frequency counts. For example, if 10 out of a total of 25 subjects are able to recognize the component or function, the location measure will be \(10 \times 100/25 = 40\). The value is extrapolated to 100 from 25. For the attribute of Compatibility With Target User Group, the location measures are calculated using the frequency count of the number of subjects who were correctly able to match the computer with the target user group. In case of the attribute of Expressiveness, the location measures are calculated using the output of the multidimensional scaling. PCPREF generates first and second score matrices that give the stimulus(attributes) projections on the subjects(computers) vectors. For example, for Computer A the second score matrix gives -0.590 for typicality, 0.336 for innovativeness, 0.365 for playfulness, 0.124 for trendiness, -0.544 for ordinariness and so on. Similarly, Computer D scores 0.572 on typicality, but -0.336 on innovativeness, -0.333 on playfulness, -0.311 on trendiness, and 0.571 on ordinariness. These are extrapolated to a scale of 0 to 100 for the location measures. Therefore, 0.336 would be 33.6, -0.333 would be 33.3, and so on.

The final utilities are calculated for all the evaluands. The most favored evaluand is the one that ends up with the highest score. Figure 31 shows the value tree for Computer A. The attributes that are greyed out are not considered as they are not tested.
Fig. 31. MAUT Value Tree for Computer A
CHAPTER IX
FUTURE RESEARCH

While testing, there are a large number of variables that can be changed to study different effects. For example, the state of the evaluand—whether it is a rough study model or a working prototype affects people's perception. The environment or context the evaluand is placed within influences its recognition. The laptop computer had it been placed within an office environment, a larger number of people who would have probably recognized it.

These two aspects form two major areas of future research. One can study the effects of the state of the evaluand on user perception. Experiments can be conducted with sketches, renderings, mock-ups (painted and unpainted) using the same set of questions. This would be a between-subjects design, with the different evaluands being tested in a fixed environment with a fixed set of questions. The results of this can be analyzed by hypothesis testing.

The effects of the physical context can also be studied similarly. A single model, or identical copies of a model can be placed in different environments and tested with the same questionnaire. This too would be a between-subjects design. This research, as explained involved a certain amount of modelmaking along with computer simulated images for creation of the evaluand. The testing was done by the researcher with one subject at a time in a controlled environment. Future research can focus on the involvement of the computer
throughout the entire process. CAD software and 3D modeling might or might not eliminate the necessity of physical modelmaking. Various questions need to be addressed. Can users perceive objects on the monitor screen and respond as they would to a physical model? Would this help in bringing the user into the design process?

The other issue is whether or not testing can be done on the computer. Interactive media can help generate questionnaires, CAD software can help create models and this eliminates the need for modelmakers and test administrators. However, the feasibility of such an approach needs to be evaluated.

Various methods of testing and analyzing data have been discussed in this thesis. However not all of them were actually performed with the models of the laptop computers. However, it has been outlined as to how these methods can be used in testing. Further research can focus on these techniques like protocol analysis, cartographic analysis, focus group study, content analysis, etc. It is possible to extract qualitative and quantitative information from these techniques. Additional research needs to be done to discover means of doing so.

There is a dire need in the industry for techniques of rapid modeling and rapid testing. Now that a system of evaluation has been outlined, it needs to be tested with a large number of alternative designs and a considerable number of subjects. This can be done with the involvement of the industry, possibly with live projects. This would add validity to the system and would also help tremendously in refining, modifying, and improving it.
BIBLIOGRAPHY


Krippendorff, K. Product Semantics, Course Notes for ID420, 1993


Smets, G. J. F., Overbeeke, C. J. and Maurer, C. Semantics of Street Furniture Third International Conference on Semantics, Helsinki, 1991


APPENDIX

The appendix contains information that was collected from the testing. These data are not required for understanding the thesis document, but are provided here as reference.

Rank Order Data

Table 4 shows the average ranks of the computers A, B, C, and D. These values were calculated from the subjects' responses and utilized in the multidimensional scaling.

Table 4. Average Ranks on the Six Attributes Tested.

<table>
<thead>
<tr>
<th>Computer</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Typical</td>
</tr>
<tr>
<td>A</td>
<td>1.72</td>
</tr>
<tr>
<td>B</td>
<td>1.84</td>
</tr>
<tr>
<td>C</td>
<td>2.44</td>
</tr>
<tr>
<td>D</td>
<td>4.00</td>
</tr>
</tbody>
</table>

125
Dissimilarities Data

Table 5 shows the average dissimilarities between the four computers, calculated on the basis of pair comparisons. Subjects were asked to rate all possible pairs for the level of dissimilarity on a scale of 1 to 10. These data were used in generating the dissimilarity matrix used in scaling.

Table 5. Average Dissimilarities on the Six Pairs Tested.

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Dissimilarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A &amp; B</td>
<td>2.36</td>
</tr>
<tr>
<td>B &amp; C</td>
<td>2.36</td>
</tr>
<tr>
<td>A &amp; D</td>
<td>8.00</td>
</tr>
<tr>
<td>B &amp; D</td>
<td>8.00</td>
</tr>
<tr>
<td>C &amp; D</td>
<td>7.84</td>
</tr>
<tr>
<td>A &amp; C</td>
<td>3.56</td>
</tr>
</tbody>
</table>

Compatibility with User Group

The frequency count for compatibility with target group, which shows the number of times the computers were matched up with their respective target groups, is displayed in Table 6. These values are a reflection of people's perception of which target group will be attracted to each of the four laptop computers.
### Table 6. Target Group Compatibility

<table>
<thead>
<tr>
<th>Ages</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-18 yrs.</td>
<td>10</td>
<td>9</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>18-23 yrs.</td>
<td>3</td>
<td>9</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>23-28 yrs.</td>
<td>10</td>
<td>5</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>28 yrs. +</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>

**Analysis Done by PCPREF**

The input file that was created to run PCPREF, and the output file that was generated by the program are shown below. The output file also gives a graphic representation of where the stimuli and the attributes lie in space. These representations though graphic are not easily comprehensible, and thus are not included here. The scaling was initially done with two dimensions, and followed by a three dimensional scaling.

**Two dimensions**


PCPREF Analysis of Computers

6 stimuli (attributes)
4 subjects (computers)
2 dimensions
Data are rank order preferences

NP NS NF IREAD MDATA NS1 NORP NCARD LABNP LABNS ISN
4 6 2 2 0 0 0 7 1 1 0

****IDENTIFICATION KEY FOR PLOTS WITH IDENTIFIED POINTS****

PT #  1 2 3 4 5 6 7 8 9 10
CHAR  A B C D E F G H I J

In the joint space plot, the first 6 points are stimuli and the next 4 points are subjects.

Subject labels have been provided. They are:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
</tr>
</tbody>
</table>

Stimulus labels have been provided. They are:

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>typicali</td>
</tr>
<tr>
<td>2</td>
<td>innovati</td>
</tr>
<tr>
<td>3</td>
<td>playful</td>
</tr>
<tr>
<td>4</td>
<td>trendy</td>
</tr>
<tr>
<td>5</td>
<td>classy</td>
</tr>
<tr>
<td>6</td>
<td>ordinary</td>
</tr>
</tbody>
</table>

MEAN OF THE RAW SCORES (BY SUBJECT)
SD OF THE RAW SCORES (BY SUBJECT)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>2.580</td>
<td>2.607</td>
<td>2.673</td>
<td>2.120</td>
</tr>
<tr>
<td>MEAN</td>
<td>.546</td>
<td>.546</td>
<td>.292</td>
<td>1.342</td>
</tr>
</tbody>
</table>
FIRST SCORE MATRIX

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>STIMULUS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>typicali innovati playful  trendy  classy  ordinary</td>
</tr>
<tr>
<td>A</td>
<td>-1.575  .769  .843  .769  .403  -1.209</td>
</tr>
<tr>
<td>B</td>
<td>-1.404  .867  1.086  .720  -.012  -1.257</td>
</tr>
<tr>
<td>C</td>
<td>-.800  .709  .023  .709  1.120  -1.760</td>
</tr>
<tr>
<td>D</td>
<td>1.401  -.805  -.835  -.745  -.417  1.401</td>
</tr>
</tbody>
</table>

CORRELATION MATRIX OF SUBJECTS

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.000</td>
<td>.977</td>
<td>.825</td>
<td>-.994</td>
</tr>
<tr>
<td>B</td>
<td>.977</td>
<td>1.000</td>
<td>.745</td>
<td>-.977</td>
</tr>
<tr>
<td>C</td>
<td>.825</td>
<td>.745</td>
<td>1.000</td>
<td>-.862</td>
</tr>
<tr>
<td>D</td>
<td>-.994</td>
<td>-.977</td>
<td>-.862</td>
<td>1.000</td>
</tr>
</tbody>
</table>

ROOTS OF THE FIRST SCORE MATRIX

22.186  1.728  .084  .002

PROPORTION OF VARIANCE ACCOUNTED FOR BY EACH FACTOR

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.924</td>
<td>.072</td>
<td>.004</td>
<td>.000</td>
</tr>
</tbody>
</table>

CUMULATIVE PROPORTION OF VARIANCE ACCOUNTED FOR

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.924</td>
<td>.996</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

SECOND SCORE MATRIX — STIMULUS PROJECTIONS ON THE SUBJECTS VECTORS

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>STIMULUS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>typicali innovati playful  trendy  classy  ordinary</td>
</tr>
<tr>
<td>A</td>
<td>-.590  .336  .365  .308  .124  -.544</td>
</tr>
<tr>
<td>B</td>
<td>-.621  .332  .434  .296  .034  -.475</td>
</tr>
<tr>
<td>C</td>
<td>-.338  .283  .008  .289  .468  -.709</td>
</tr>
<tr>
<td>D</td>
<td>.572  -.336  -.333  -.311  -.164  .571</td>
</tr>
</tbody>
</table>
### POPULATION MATRIX

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>DIMENSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>.993</td>
</tr>
<tr>
<td>B</td>
<td>.968</td>
</tr>
<tr>
<td>C</td>
<td>.889</td>
</tr>
<tr>
<td>D</td>
<td>-.999</td>
</tr>
</tbody>
</table>

### STIMULUS MATRIX

<table>
<thead>
<tr>
<th>STIMULUS</th>
<th>DIMENSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>typicali</td>
<td>-.555</td>
</tr>
<tr>
<td>innovati</td>
<td>.335</td>
</tr>
<tr>
<td>playful</td>
<td>.302</td>
</tr>
<tr>
<td>trendy</td>
<td>.312</td>
</tr>
<tr>
<td>classy</td>
<td>.199</td>
</tr>
<tr>
<td>ordinary</td>
<td>-.593</td>
</tr>
</tbody>
</table>

Correlation between subjects original preferences and their projections on the fitted vectors.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Coefficient of Correlation</th>
<th>Spearman Rank Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.9960</td>
<td>.9856 T( 4)= .000</td>
</tr>
<tr>
<td>B</td>
<td>.9971</td>
<td>1.0000 T( 4)= .000</td>
</tr>
<tr>
<td>C</td>
<td>.9998</td>
<td>.9856 T( 4)= .000</td>
</tr>
<tr>
<td>D</td>
<td>.9999</td>
<td>.9276 T( 4)= .000</td>
</tr>
</tbody>
</table>

Three dimensions

---

PCPREF: MULTIDIMENSIONAL PREFERENCE SCALING.
VERSION 1.2, JULY, 1989.

pcpref analysis of computers

6 stimuli (attributes)
4 subjects (computers)
3 dimensions
Data are rank order preferences

NP NS NF IREAD MDATA NS1 NORT NCARD LABNP LABNS ISN
4 6 3 2 0 0 0 7 1 1 0

****IDENTIFICATION KEY FOR PLOTS WITH IDENTIFIED POINTS*****

PT # 1 2 3 4 5 6 7 8 9 10
CHAR A B C D E F G H I J

In the joint space plot, the first 6 points are stimuli and the next 4 points are subjects.

Subject labels have been provided. They are:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
</tr>
</tbody>
</table>

Stimulus labels have been provided. They are:

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>typicali</td>
</tr>
<tr>
<td>2</td>
<td>innovati</td>
</tr>
<tr>
<td>3</td>
<td>playful</td>
</tr>
<tr>
<td>4</td>
<td>trendy</td>
</tr>
<tr>
<td>5</td>
<td>classy</td>
</tr>
<tr>
<td>6</td>
<td>ordinari</td>
</tr>
</tbody>
</table>

MEAN OF THE RAW SCORES (BY SUBJECT)
SD OF THE RAW SCORES (BY SUBJECT)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>2.580</td>
<td>2.607</td>
<td>2.673</td>
<td>2.120</td>
</tr>
<tr>
<td>MEAN</td>
<td>.546</td>
<td>.546</td>
<td>.292</td>
<td>1.342</td>
</tr>
</tbody>
</table>
### FIRST SCORE MATRIX

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>typically</th>
<th>innovati</th>
<th>playful</th>
<th>trendy</th>
<th>classy</th>
<th>ordinary</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1.575</td>
<td>.769</td>
<td>.843</td>
<td>.769</td>
<td>.403</td>
<td>-1.209</td>
</tr>
<tr>
<td>B</td>
<td>-1.404</td>
<td>.867</td>
<td>1.086</td>
<td>.720</td>
<td>-.012</td>
<td>-1.257</td>
</tr>
<tr>
<td>C</td>
<td>-.800</td>
<td>.709</td>
<td>.023</td>
<td>.709</td>
<td>1.120</td>
<td>-1.760</td>
</tr>
<tr>
<td>D</td>
<td>1.401</td>
<td>-.805</td>
<td>-.835</td>
<td>-.745</td>
<td>-.417</td>
<td>1.401</td>
</tr>
</tbody>
</table>

### CORRELATION MATRIX OF SUBJECTS

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.000</td>
<td>.977</td>
<td>.825</td>
<td>-.994</td>
</tr>
<tr>
<td>B</td>
<td>.977</td>
<td>1.000</td>
<td>.745</td>
<td>-.977</td>
</tr>
<tr>
<td>C</td>
<td>.825</td>
<td>.745</td>
<td>1.000</td>
<td>-.862</td>
</tr>
<tr>
<td>D</td>
<td>-.994</td>
<td>-.977</td>
<td>-.862</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### ROOTS OF THE FIRST SCORE MATRIX

22.186  1.728  .084  .002

### PROPORTION OF VARIANCE ACCOUNTED FOR BY EACH FACTOR

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>.924</td>
<td>.072</td>
<td>.004</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

### CUMULATIVE PROPORTION OF VARIANCE ACCOUNTED FOR

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>.924</td>
<td>.996</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

### SECOND SCORE MATRIX — STIMULUS PROJECTIONS ON THE SUBJECTS VECTORS

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>typically</th>
<th>innovati</th>
<th>playful</th>
<th>trendy</th>
<th>classy</th>
<th>ordinary</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-.642</td>
<td>.311</td>
<td>.348</td>
<td>.311</td>
<td>.167</td>
<td>-.495</td>
</tr>
<tr>
<td>B</td>
<td>-.572</td>
<td>.351</td>
<td>.447</td>
<td>.291</td>
<td>-.003</td>
<td>-.514</td>
</tr>
<tr>
<td>C</td>
<td>-.326</td>
<td>.288</td>
<td>.011</td>
<td>.288</td>
<td>.458</td>
<td>-.719</td>
</tr>
<tr>
<td>D</td>
<td>.574</td>
<td>-.335</td>
<td>-.332</td>
<td>-.311</td>
<td>-.165</td>
<td>.570</td>
</tr>
</tbody>
</table>
## Population Matrix

<table>
<thead>
<tr>
<th>Subject</th>
<th>Dimension 1</th>
<th>Dimension 2</th>
<th>Dimension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.989</td>
<td>-0.114</td>
<td>-0.089</td>
</tr>
<tr>
<td>B</td>
<td>0.966</td>
<td>-0.249</td>
<td>0.076</td>
</tr>
<tr>
<td>C</td>
<td>0.889</td>
<td>0.458</td>
<td>0.019</td>
</tr>
<tr>
<td>D</td>
<td>-0.999</td>
<td>0.054</td>
<td>0.002</td>
</tr>
</tbody>
</table>

## Stimulus Matrix

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Dimension 1</th>
<th>Dimension 2</th>
<th>Dimension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>typical</td>
<td>-0.555</td>
<td>0.338</td>
<td>0.619</td>
</tr>
<tr>
<td>innovati</td>
<td>0.335</td>
<td>-0.031</td>
<td>0.264</td>
</tr>
<tr>
<td>playful</td>
<td>0.302</td>
<td>-0.569</td>
<td>0.182</td>
</tr>
<tr>
<td>trendy</td>
<td>0.312</td>
<td>0.025</td>
<td>-0.056</td>
</tr>
<tr>
<td>classy</td>
<td>0.199</td>
<td>0.635</td>
<td>-0.481</td>
</tr>
<tr>
<td>ordinary</td>
<td>-0.593</td>
<td>-0.397</td>
<td>-0.528</td>
</tr>
</tbody>
</table>

Correlation between subjects original preferences and their projections on the fitted vectors.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Coefficient of Correlation</th>
<th>Spearman Rank Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.0000</td>
<td>0.9856</td>
</tr>
<tr>
<td>B</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>C</td>
<td>1.0000</td>
<td>0.9856</td>
</tr>
<tr>
<td>D</td>
<td>0.9999</td>
<td>0.9276</td>
</tr>
</tbody>
</table>