Consumer evaluations of early product-concepts

C.J.P.M. de Bont
15-12-1992
1. Om inzicht te krijgen in de evaluatie van (vroege) produktconcepten, dient men zich behalve van de hoeveelheid informatie over en de aanbiedingsvorm van die concepten ook rekenschap te geven van de bij consumenten aanwezige kennis van het betreffende produktveld (Dit proefschrift).

2. De eigenschappen van produktconcepten zijn slechts door consumenten met voldoende kennis van het betreffende produktveld op hun waarde te schatten (Dit proefschrift).

3. Indien relevante informatie over een belangrijk attribuut op niet-eenduidige wijze gecommuniceerd wordt in een produktconcept, wordt het attribuut prijs door respondenten veelal als indicator voor de aantrekkelijkheid van dat concept gebruikt¹. Dit resulteert bij een conjuncte analyse in een geringe betrouwbaarheid en een beperkte validiteit van de deelutiliteiten van het attribuut prijs (Dit Proefschrift).

4. Door de beperkte theoretische onderbouwing bevindt het zogenoemde "produkt-concept testen" zich ondanks de talloze praktische toepassingen nog altijd in de concept-fase.

5. De bevinding dat veel gepubliceerde studies naar de betrouwbaarheid en de validiteit van resultaten van conjuncte analyses zich kenmerken door methodologische tekortkomingen, roept twijfels op ten aanzien van de kwaliteit van beoordelingsprocedures van wetenschappelijke tijdschriften.

6. Ontwerpen van produkten op basis van kennis van preferenties van consumenten betekent voor de ontwerper een inperking van zijn vrijheidsgraden, waardoor hij zijn creatieve vermogens efficiënter kan aanwenden.

7. De grafische mogelijkheden van C.A.D.-systemen worden pas ten volle benut bij het testen van de marktacceptatie van produktconcepten als het ontwerpproces zodanig ver is voortgeschreden dat de vormdetaillering van de concepten bekend is.

8. Gezien het belang van de presentatie van ideeën voor de acceptatie daarvan, verdient de wijze waarop de Faculteit van het Industrieel Ontwerpen studenten opleidt in de presentatie van de eigen produkten, navolging bij andere opleidingsinstituten.

9. Gelet op de hoge kosten die gepaard gaan met het ontwikkelen en introduceren van nieuwe produkten, op het toegenomen belang van de marketing discipline in Nederlandse bedrijven en op de geavanceerdheid van marktonderzoekstechnieken zoals conjuncte analyse, zijn de bestedingen aan marktonderzoek relatief beperkt te noemen.

10. Door de relatief grote aandacht voor de Inca-cultuur wordt onevenredig weinig aandacht geschonken aan de minstens zo belangrijke culturen van voor de Inca-tijd.

11. De veronderstelling dat ten gevolge van het ongunstige imago van de krijgsmacht als werkgever het moeilijk zal zijn een beroepsleger op de been te brengen (cf. Commissie Meijer: Naar dienstplicht nieuwe stijl, 1992), moet niet tot de instandhouding van de militaire dienstplicht, maar tot de verbetering van dit imago leiden.
Consumer Evaluations of Early Product-Concepts

C.J.P.M. DE BONT
Consumer Evaluations of Early Product-Concepts

Proefschrift
ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft
op gezag van Rector Magnificus, prof. drs. P.A. Schenck,
in het openbaar te verdedigen ten overstaan van
een commissie aangewezen door het College van Dekanen
op dinsdag 15 december 1992
om 14.00 uur door

Cornelis Joachim Petrus Maria de Bont
geboren te Raamsdonk op 12 april 1964
doctorandus in de Sociale Wetenschappen
Delft University Press / 1992
Promotiecommissie

Promotoren:

Prof. dr. W.M. Oppedijk van Veen
Dr. J.P.L. Schoormans (toegevoegd promotor)

Overige leden:

Prof. dr. ir. J.A. Buijs
Prof. ing. A.H. Marinissen
Prof. dr. G.J.F. Smets
Prof. dr. G.M. van Veldhoven (Katholieke Universiteit Brabant)
Prof. dr. Th.M.M. Verhallen (Katholieke Universiteit Brabant)

ISBN 90-6275-833-9 / CIP
Copyright 1992 by C.J.P.M. de Bont
All rights reserved

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission from the publisher: Delft University Press, Stevinweg 1, 2628 CN Delft, The Netherlands.

Printed in the Netherlands.
Acknowledgements

This thesis could not have been made without the efforts of a large number of persons. To express my gratitude, I am glad to mention them here.

To begin with, I would like to thank those persons who assisted me in formulating the research problem and who suggested pathways for finding solutions. In both respects, my promotor Walie Oppedijk van Veen and my co-promotor Jan Schoormans have contributed to a large extent. This thesis has much benefited from the theoretical and practical capacities of both my promoters.

Furthermore, I would like to thank other researchers for discussing elements of this thesis with me. In particular, Dick Wittink, Mia Stokmans, Roland Ortt, Dirk Snelders and Gerard Loosschilder provided me with useful comments.

In addition, the remarks made and questions posed, with regard to presentations of parts of my thesis, have helped me to improve this thesis. For these remarks and questions, I would like to thank the AIO-group of the Network Bedrijfseconomie, and attendants of IAREP and EMAC conferences.

Finally, I wish to thank those persons who have supported me in all kinds of practical areas. From the Department of New-Product Development, I mention Agnes Tan, Jan Poot, Anneke Oosterom and Sandra Snoek. Their enthusiasm, respectively in inviting respondents, taking care of computer facilities, improving lay-out, and administrating, has facilitated my work in a very pleasant way. I received further practical support from Marjan Wessel, Liedewij de Graaf, Mees de Lind van Wijngaarden, Marja van der Pasch, Maarten ten Houten and Dorien de Vries, all students who participated in different stages of the research project.

In the final stage of the research project, Gary Calderwood was kind enough to revise my secondary-school English. In this final stage, I was very pleased with the attitude of staff of the Department of Social Sciences of the Royal Military Academy (KMA). They gave me the opportunity to finish my thesis while fulfilling my military obligations.

Even more than any of the persons already mentioned, I would like to thank Judith for her help. She fully supported my work on the thesis, and only she entirely realizes the (behavioral) costs it has taken.

Breda, July 1992
CONTENTS

1. Introduction 1
   1.1 The aim of the study 1
   1.2 Theoretical perspective 2
   1.3 Plan of the thesis 3

2. New-Product Development and Consumer Research 5
   2.1 Introduction 5
   2.2 Stages of the new-product development process and consumer-research activities 7
   2.3 Practical reasons for focusing on concept testing 14
   2.4 Summary and conclusions 14

3. The Problem of Testing the Consumer Acceptance of Early Product-Concepts 17
   3.1 The nature of early new-product concepts 17
   3.2 The problem of evaluating early product-concepts 19

4. The Evaluation of Early Product-Concepts 21
   4.1 Models of product evaluation 21
   4.2 Integration of models of product evaluation into a Model of Product-Concept Evaluation 25
   4.3 The evaluation of early product-concepts: the formulation of hypotheses 30
   4.4 Summary and conclusions 36

   5.1 Product-concept information: choice of test product 39
   5.2 Processing: level of processing and type of processing 46
   5.3 Evaluative responses: choice of conjoint analysis 49
   5.4 Consumer characteristics: product-category knowledge 53
   5.5 Summary 55

   6.1 Experiment 1: The effect of “amount of information” on the processing of early product-concepts 58
   6.2 Experiment 2: the effects of “amount of information” and “product-category knowledge” on the evaluative responses 61
   6.3 Experiment 3: the effects of “presentation format” on the processing of early product-concepts and on its evaluative responses 68
   6.4 Experiment 4: The effects of missing information on a relevant attribute on the evaluation of the attribute “price” 72
   6.5 Summary and conclusions 75
## Contents

7. The Reliability and the Validity of Conjoint-Analysis Results: A Review of Literature 81
   7.1 Reliability and validity in general 82
   7.2 Studies on the reliability and validity of conjoint-analysis results 81
      7.2.1 The reliability of conjoint-analysis results 84
      7.2.2 The concurrent validity of conjoint-analysis results 89
      7.2.3 The predictive validity of conjoint-analysis results 92
   7.3 Summary and conclusions 96
   7.4 Hypotheses to be tested 98

8. Operationalizations for Testing Hypotheses about the Reliability and the Validity of Conjoint-Analysis Results 101
   8.1 Reliability 101
   8.2 Concurrent validity 102
   8.3 Predictive validity 102
   8.4 Data-collection method 104

9. Testing Hypotheses about the Reliability and the Validity of Conjoint-Analysis Results 105
   9.1 An empirical study on the reliability of conjoint-analysis results 105
      9.1.1 The effects of amount of information 106
      9.1.2 The effects of presentation format 107
      9.1.3 The effects of product category knowledge 110
      9.1.4 The effects of data-collection method 111
   9.2 An empirical test of the concurrent validity of conjoint-analysis results 112
      9.2.1 The effects of presentation format 113
      9.2.2 The effects of product category knowledge 113
   9.3 An empirical test of the predictive validity of conjoint-analysis results 114
      9.3.1 The effects of amount of information 114
      9.3.2 The effects of presentation format 116
      9.3.3 The effects of product category knowledge 117
      9.3.4 The effects of data-collection method 119
   9.4 Summary and conclusions 120

10. General Discussion 125
    10.1 Introduction 125
    10.2 Theoretical implications 125
        10.2.1 Theoretical implications of the Model of Product-Concept Evaluation 125
        10.2.2 Theoretical implications for conjoint analysis 128
    10.3 Suggestions for further research 130
    10.4 Practical implications 130
        10.4.1 Practical implications for researchers 131
        10.4.2 Practical implications for product-designers 132
Summary 135
Samenvatting 139
References 143

Appendix 1:
An illustration of conjoint analysis: The Sunbeam case 157

Appendix 2:
Results of control questionnaire 159

Appendix 3:
The structure of studies on the reliability and the validity of conjoint-analysis results 167

Appendix 4:
Studies on the reliability of conjoint-analysis results 173

Appendix 5:
Studies on the concurrent validity of conjoint-analysis results 177

Appendix 6:
Studies on the predictive validity of conjoint-analysis results 181
1. Introduction

1.1 The aim of the study

New-product development is a strategic process. By developing new products an entrepreneur can attempt to gain (an increase in) market share. In order to be successful in the market place, it is important that the new product can compete with, or outperform, rival products.

In developing a new product, many decisions have to be made. The most crucial one is to decide what type of product will be made. Subsequently, and among other questions, the following needs to be answered: what characteristics should the new product have? In answering this question, it is necessary to obtain insight into the products which are already on the market. This insight comprises the extent to which these products fulfill consumer wishes.

By combining the strengths of the firm with opportunities in the environment, ideas about new products can be generated. Once a number of ideas have been generated, the most promising ones can be transformed into new-product concepts. An example of a new-product concept is presented in Figure 1.

```
without dripstop
with thermos flask
80 guilders
```

Figure 1: Example of a new-product concept

Even at this early stage of the new-product development process, an attempt can be made to integrate consumer preferences. By collecting evaluations of early product-concepts from potential buyers of the new product, the development of the product can be directed at consumer demand. As a result, better choices can be made with respect to the characteristics of the new product, substantially increasing the chance of a successful market introduction. Also, by integrating consumer preferences during the early stages of the development process, money wasted during the development, and the risk of introducing unwanted products, can both be minimized.

This idea of integrating consumer preferences as early as possible, is based on the implicit assumption that consumers are actually capable of evaluating early product-concepts. Whether or not this assumption is justified is the subject of this thesis. Only if consumers are capable of evaluating the early product-concepts does it make sense to integrate consumer preferences into early stages of the development process. If they
are incapable of doing so, the incorporation of consumer evaluations has to be postponed until later stages in the process.

1.2 Theoretical perspective

Whether or not consumers are capable of evaluating early product-concepts is investigated from a psychological perspective. Although much attention is given to the different consumer-research techniques which are applied in consumer research in general (particularly conjoint-analysis in concept optimization), the key concepts are borrowed from psychology. In our opinion, (consumer) psychology offers the most appropriate perspective to investigate the validity of the assumption above, since the subject of how consumers evaluate product stimuli has been intensively investigated from this perspective. Moreover, much attention has been given to the cognitive aspects of the evaluation process.

The insights obtained in consumer psychology are helpful in answering our (first) research question, i.e. in what respects are early product-concepts evaluated differently from actual products? In answering this question, first the nature of early product-concepts will be defined, and subsequently the available amount of information in them and their presentation format will be distinguished as the most important differences between early product-concepts and actual products. Next, different models of product evaluation will be discussed and integrated into the Model of Product-Concept Evaluation. In this model, a distinction will be made between the processing and the evaluation of new-product concepts, since new-product information has to be processed before it can be evaluated. Processing and evaluation constitute different stages which precede the evaluative response. Consumer characteristics are also part of the Model of Product-Concept Evaluation. In particular, the consumer characteristic of "product-category knowledge" is expected to influence the process of forming an evaluation.

In short, in determining the ways in which early product-concepts are evaluated differently from actual products, much attention will be given to early product-concepts as stimuli (consisting of the variables "amount of information" and "presentation format"), and to consumers as intervening variables with a certain level of "product-category knowledge". Different experiments are conducted to assess the effects of each of these variables on the evaluation of early product-concepts.

The variables which do affect such evaluations can be considered to be variables which potentially endanger the reliability and the validity of the evaluations. Whether or not they affect the said reliability and validity is investigated through answering the second research question: how reliable and how valid are evaluations based on early product-concepts? First, a review is presented of studies investigating the reliability and the validity of conjoint-analysis results. For those variables, for which insufficient insight is obtained, additional empirical research is conducted.

Our reason for first investigating the influences of stimulus and consumer variables on the evaluations of early product-concepts, and then investigating the consequences in terms of reliability and validity, instead of the other way around, is based on pragmatic reasons. Assessing differences in evaluations can be accomplished in a few months. By contrast, assessing differences in reliability and validity can take several years. As a consequence, we first concentrate on the evaluations. Only those variables which lead to differences in evaluations between early product-concepts and
actual products need to be included in a study on the reliability and the validity of the evaluations. In addition, we know which consumer characteristics have to be controlled for.

1.3 Plan of the thesis

In Chapter 2, the importance of consumer research for the development of new products is discussed. In addition, methods of consumer research are presented in relation to the different stages of the new-product development process. Practical reasons are given for concentrating on the concept stage. This stage starts when product-concepts are formulated.

Chapter 3 discusses the main problem which is investigated in this thesis and we elaborate upon the nature of early product-concepts. It is in this chapter that "amount of information" and "presentation format" are distinguished as stimulus characteristics.

In Chapter 4, we concentrate on the evaluation of early product-concepts. The Model of Product-Concept evaluation, an integration of different models of product evaluation, is presented as a starting point for describing and investigating the evaluation of early product-concepts. Various hypotheses are derived from the model and these hypotheses are tested in the succeeding chapters. The operationalizations which are used to test these hypotheses are discussed in Chapter 5. We select conjoint analysis as the technique for measuring evaluations.

In Chapter 6, the results are presented and an answer is given to the first research question (in what respects are early product-concepts evaluated differently from actual products). Those aspects which influence the evaluation of early product-concepts are included in Chapter 7 as factors endangering the reliability and the validity of the evaluations.

In Chapter 7, first the concepts of reliability and validity are briefly discussed. After this, the outcomes of many studies on the reliability and the validity of conjoint-analysis results are described. In doing so, special attention is given to the variables "amount of information", "presentation format", and "product-category knowledge". In addition, attention is given to the methodology applied in assessing the reliability and the validity of conjoint-analysis results. Since little research has been conducted on the influences of these factors, hypotheses are formulated for further investigation.

In Chapter 8, operationalizations are discussed to test these hypotheses, and in Chapter 9 they are tested. Also in Chapter 9, conclusions are drawn with respect to the reliability and the validity of evaluations based on early product-concepts. In this way the second research question is answered (how reliable and how valid are evaluations based on early product-concepts?). In the final chapter (Chapter 10), overall conclusions are drawn and the generalizability of the findings is discussed. In addition, the conditions are specified under which it makes sense to integrate consumer evaluations into the early stages of the new-product development process.
2. New-product Development and Consumer Research

2.1 Introduction

The key to achieving organizational goals consists of determining the needs and wants of target markets and delivering the desired satisfactions more effectively and efficiently than competitors. This is the definition of the marketing concept as it is stated by Kotler (1988, p. 17). In this definition, two main activities are distinguished. The first is the determining of needs and wants of target markets, and the second is delivering the desired satisfaction through products and/or services. The first activity is primarily performed by marketing and market-research departments, the second by those of research and development.

Both Kotler (1988) and Day & Wensley (1988) demonstrate that the delivery of the desired satisfaction can, in some instances, be realized by developing new products. Insight into the needs and wants of target markets can be used to create superior customer value (Day & Wensley, 1988). Therefore, to be successful in the market place, one has to be both capable of gaining insight into the product demands of potential buyers of the product, and of transforming these product demands into actual products. Moreover, in a competitive environment, one must outperform competitors in both activities.

A new product can be defined as a new way to satisfy (existing or new) consumer needs. This new way can, in the eyes of consumers, be more or less desirable than those already existing. Following Lancaster (1966; 1971), whose ideas have been very influential in marketing literature, we assume that a product can be considered as a bundle of attributes. Consequently, the desirability of a product is determined by its constituent attributes. This contribution by Lancaster (1971) has been very beneficial to new-product development and is still widely accepted. His definition of an attribute, however, has been broadened by other scholars.

In Lancaster's economic approach, it is assumed that attributes are objectively measurable, universal, and both quantitative and physical. By contrast, in psychological approaches (either cognitive or motivational), the perception of the attributes, not the attributes themselves, is seen to be of prime importance (e.g. Anderson, 1981; 1982; Holbrook, 1985). As a consequence, although physical attributes may be objectively measurable and universal, perceptions of them are subjective and can be largely idiosyncratic. Since different consumers may perceive the same attributes in different ways, it is important to know how the attributes are perceived. Only by knowing how an attribute (product) is perceived can the link between that attribute (product) and its desirability be well understood.

Another adaptation of the Lancaster's definition is given by Garner (1978). Garner distinguishes between quantitative and qualitative attributes, to which she respectively refers as dimensions and features. Quantitative attributes are variables at an ordinal or metric level. A filter coffee-maker, for instance, has quantitative attributes, such as "capacity" and "price" which are both metric variables. Qualitative attributes are variables at a nominal level (categorical variables), for example the presence or absence of a dripstop and a thermosflask. A consumer may buy a given filter coffee-
maker (rather than any alternative) because it consists of a particular combination of quantitative and qualitative attributes (e.g. a large capacity, inexpensiveness, and the presence of a thermosflask).

With respect to the adjective "physical" in Lancaster's definition of attributes, the following point of criticism can be made. Attributes are not always physical. Examples are the attributes "price" and "brand image". By strictly concentrating on physical attributes, one may forget the contribution of non-physical attributes to the desirability of a product.

Assuming that a product is a bundle of attributes, one can assert that a new product differs from existing products in one of the four following ways:

- The level(s) of one or more existing product attributes are changed (e.g. a revolutionary small video camera). This mostly concerns changes in quantitative attributes.
- One or more new1 attributes are added to (or taken from) an existing product (e.g. a timer is added to a filter coffee-maker). This mostly concerns changes in qualitative attributes.
- Both the level(s) of existing product attributes are changed, and one or more new attributes are added (e.g. a revolutionary small filter coffee-maker with a timer).
- A new combination of already existing attribute levels and attributes is created, (e.g. a drip coffee maker which has a large capacity, an on/off indication light, a thermosflask, and a rather low price).

From the perspective of new-product design, one only speaks of a new product if at least one physical change to the core product has been made. The core product consists of the physical attributes of the product itself (e.g. Leeftlang & Beukenkamp, 1987). In the augmented product, the core product is extended by attributes of the product as marketed. To create the augmented product no physical changes are made to the core product.

In marketing, in contrast to new-product design, a broader point of view is often taken. In the marketing perspective, mere changes to the augmented product, such as changes in the packaging of the product, or in the way it is communicated to its potential buyers, can be sufficient to warrant speaking of a new product. For instance, Crawford (1985) demonstrates how by creating differences in the positioning of a product in the market, a new product can be developed. In that case, the new-product changes do not consist of changes in physical attributes, but of changes in the communication of existing attributes. In this thesis, we take the perspective of new-product design, and only consider a product to be new if at least one physical attribute has been changed in any of the four mentioned ways.

In order to create superior customer value, the new product should, in the eyes of the consumers, be more desirable than existing products. Consumer research can be helpful in providing insight into consumer needs and subsequently in guiding the development of new products. More specifically, insights from consumer research can be transformed into such qualitative or quantitative new-product differences that superior value is obtained. In this respect, a series of studies investigating the dimensions of new-product success show the importance of market research for the development of new products (Rothwell et al., 1974; Cooper, 1979; 1982; Booz et al., 1982). Most successful firms use a market-oriented program to develop products with differential customer advantage (Cooper, 1986).
This thesis concentrates on the consumer-research activity of determining the respects in which new products should differ from already existing products. Therefore, attention will be given to the crucial questions as to which product attributes should be changed and in what directions. Major methodological problems in performing this consumer-research activity will be discussed. Before doing so, we will first give attention to the broader organizational context of new-product development.

2.2 Stages of the new-product development process and consumer research activities

In Section 2.1, different studies were mentioned stating the importance of consumer research in achieving new-product success. In this section, we will show that the informational needs of the new-product designer depend on the specific stage of the new-product development process. This is due to the fact that the product decisions to be made differ according to stage within the new-product development process (Box & Van Eyk, 1983). Correspondingly, the consumer research methods applied are different at the different stages. Therefore, before going into the methods of consumer research, first the new-product development process will be described.

A comprehensive model of the new-product development process, covering its essential activities, can be found in the recent book on the structure and methods for developing new products by Roozenburg & Eekels (1991). The essentials of this model are shown in Figure 2.

![Diagram of new-product development process](image)

**Figure 2:** Major components of the new-product development process

In Figure 2, we can see that the new-product development process consists of three levels. These levels comprise different parts of the development process: the strategic part, the planning part and the operations part. The strategic part is referred to as product policy-making and product idea-generation. In the strategic part, strategic issues are considered, such as starting a new-product development project, taking over competing firms, business alliances, or getting involved in market development. If the decision is taken to develop new products, the second part, the so-called planning part, can begin. In this part, a number of different activities are initiated simultaneously. These activities are concerned with answering the following questions: what specific type of product will be developed (product elaboration and design engineering), how will it be produced (production planning and development), and how will it be presented to the market (marketing-mix planning)? When the planning
part is completed, it can be turned into action in the operations part. This part is referred to as “realisation” in Figure 2. The integration of different activities, both in the planning and the operational part of the new-product development process, is the responsibility of the new-product designers.

The focus on needs and wants of target groups, and related market-research activities, is dominant in the planning part, both in product design and in the planning of the marketing mix. The activities which take place in these boxes are described in more detail in textbooks on new-product management (e.g., Booz et al., 1968; Wind, 1982; Hisrich & Peters, 1984; Crawford, 1987; Urban et al., 1987; Carels, 1990). In Figure 3, these activities are depicted in four different and frequently quoted sequential models. As can be seen, horizontal lines are drawn to establish the phases (A to E) which correspond with subsequent information needs of the product designer. For the sake of clarity, we have depicted the product-development activities of the different models which, in our opinion, are highly similar in these phases.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic</td>
<td>A</td>
<td>strategic analysis</td>
<td>organizational structuring</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>idea generation</td>
<td>idea stage</td>
<td>concept creation and development</td>
<td>opportunity identification</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>idea/concept screening</td>
<td>concept stage</td>
<td>concept evaluation</td>
<td>design testing</td>
</tr>
<tr>
<td>Planning</td>
<td>D</td>
<td>concept/ product development</td>
<td></td>
<td>product development stage</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>concept/ product evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>generation and evaluation of final product</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operations</td>
<td>E</td>
<td>introduction</td>
<td>commercialization</td>
<td>commercialization</td>
<td>introduction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>designing system of continuous evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Models of major product-development activities and phases of corresponding information needs
In Figure 3 one can observe that both Wind (1982) and Urban et al. (1987) start their models in Phase A. Phase A represents the product policy-making part of the upper box of Figure 2. The information needs in Phase A mainly concern (internal) strategic matters such as the financial position of the firm. The interest in (external) market-related matters concentrates on expectations of the market figures of current products. At this general level, market research can play a crucial role. By investigating market structures and wider external opportunities, management can become aware of the necessity to adapt its current product policy (Box and Van Eyck, 1983). Nevertheless, since in Phase A the decision has not yet been made to get involved in new-product development, no consumer research can be initiated with respect to the acceptance of new-product concepts. Therefore, bearing the goal of this dissertation in mind, Phase A does not receive further attention here.

Once the decision is made to start a new-product development project, the creative process can begin. The first activity of such a creative product-development process depends on the desired newness\(^2\) of the new product. If a totally new product is desired, one has to start from scratch (Phase B). The first choice which has to be made in Phase B is the selection of a specific product domain. This activity is referred to as product idea-generation in Figure 2, and as idea generation, idea stage, concept creation and development, and opportunity identification in Figure 3. The information needs in Phase B consist of finding unfulfilled consumer needs. The corresponding consumer-research activities, which can be executed in the chosen product domain, are referred to as need assessment. Phase B ends with the formulation of new-product ideas.

The Phases C and D are represented by product elaboration and design engineering in Figure 2. The ideas which have been generated in Phase B can be developed further in Phase C. In Phase C, the information needs concern the consumer acceptance of product-concepts. After defining the new-product concepts, they can be evaluated by potential buyers. This consumer-research activity is called concept testing. The most promising ones can be developed further into prototypes\(^3\). If a firm already has at least global ideas of how to change an existing product (or line of products) at the beginning of the project, the first activity of the development process can be located in Phase C.

Once prototypes have been constructed, the consumer acceptance of the product, as it functions, can be tested by means of product testing in Phase D. This may, or may not, lead to adaptations of the product.

In Phase D, finishing touches are made to the development process. Price, distribution channel, and promotional means for introducing the product into the market are determined. In consumer research, the test-market technique is sometimes applied for this purpose. This part of Phase D corresponds to the marketing-mix planning of Figure 2.

Since, in Phase E, no physical changes are made to the core product, this phase usually falls outside the scope of the (strict)\(^4\) new-product design perspective.

Below, the consumer research activities in each of the phases (with the exception of the Phases A and E) will be discussed in somewhat more detail. An attempt will be made to demonstrate how the consumer-research activities are embedded in the product development process. First, however, in Figure 4, a very short overview is given of the information needs, the corresponding consumer-research activities, and the outcomes in each of

---

\(^2\) In relation to the firm or to the market.

\(^3\) A prototype is a complete functioning version of the new product. This version is the result of craftsmanship, in contrast to the 0-number which is the first version from industrial production.

\(^4\) In Roozenburg & Eekels' (1991) terms.
the phases. In addition, the outcomes are presented in terms of product output in the different phases.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Information need</th>
<th>Consumer-research activity</th>
<th>Product output</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>market opportunities</td>
<td>Need assessment</td>
<td>ideas concepts</td>
</tr>
<tr>
<td>C</td>
<td>acceptance of concepts</td>
<td>Concept testing</td>
<td>concepts prototypes</td>
</tr>
<tr>
<td>D</td>
<td>acceptance of products (functions) market-entry strategy</td>
<td>Product testing</td>
<td>prototypes products</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marketing testing</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Information needs and corresponding consumer research activities in different phases of the new-product development process

Phase B: Need assessment

Need assessment is the consumer research part of the broader design activity called 'goal-finding' (e.g. Roozenburg & Eekels, 1991). Goal-finding seeks a global specification of consumer benefits which have to be provided by the new product. This specification comprises the design goal, or in other words, the product to be designed. Hypothetical examples of design goals could be a motorcycle which provides much protection in the case of accidents, or a very flat television set which occupies less space.

Goal-finding consists of different activities. The first activity is searching for one or more domains which suggest a potential for providing successful product ideas. Crawford (1987) refers to this first activity as formulating the 'new-product arena', and considers this activity to be part of organizational structuring. One method for formulating the new-product arena is the so-called 'search-area' method presented by Buijs (1987). Essentially, in this method, a matrix is constructed by figuring the internal strengths (e.g. highly educated technical personnel or a sound financial position) and external opportunities (e.g. individualization, new markets born of political changes) on two axes. Each combination in the matrix defines a search-area. The combinations showing most potential are given special attention, and in these, a closer look is taken at the market structures. This examination offers an insight into (trends in) the (lack of) customer satisfaction provided by existing products. In this step, consumer research is applied to the search for poorly fulfilled needs. The detection of specific poorly fulfilled needs can be the basis for a program of requirements for the new product, (Roozenburg & Eekels, 1991)\footnote{A program of requirements is, among other purposes, used by product designers to specify which product functions will have to be delivered by the new product in order to supply customer satisfaction.}

Different consumer-research techniques have been developed to detect poorly fulfilled consumer needs. These techniques differ with respect to the number of alternatives (existing products) which are taken into consideration, and with respect to level of product specificity. At a high level of product specificity (e.g. a metal comb), one can investigate the desirability of a particular attribute. At a low level of product specificity (e.g.
a visual communication device), one can only give attention to the global needs and value structures of consumers.

Descriptions of different consumer-research techniques can be found in Wind (1982), Box & Van Eyk (1983), Crawford (1987), and in Holt (1989). Wind (1982) presents techniques which focus on a large number of alternatives, such as market structure/gap analysis (based on multi-dimensional scaling), and on a limited number of alternatives, such as problem detection studies, and consumer complaints. In addition, some of the techniques mentioned by Wind are for a high level of product specificity, such as product-deficiency analysis. Other techniques are for a low level of product specificity, such as motivation research, or consumption-system analysis. For a more detailed description of consumer research at a low level of product specificity, we refer the reader to Ort et al. (1992).

The techniques mentioned by Holt (1989) and those by Wind (1982) overlap. Crawford (1987) adds scenario analysis to the techniques already mentioned. In Figure 5, the distinctions between different consumer-research techniques are summarized.

<table>
<thead>
<tr>
<th>number of alternatives</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>motivation research, scenario analysis</td>
<td>product deficiency analysis, consumer complaints</td>
</tr>
<tr>
<td>Large</td>
<td>consumption system analysis</td>
<td>multi-dimensional scaling</td>
</tr>
</tbody>
</table>

Figure 5: Distinctions between different need-assessment techniques

The appropriateness of a given consumer-research technique depends on the market structure (many versus few existing products), and on the level of product specificity of the search-area (high, e.g. pleasure in private transport; or low, e.g. sugar confectionery in cars).

After having formulated a global or specific design goal, the second activity of goal-finding can begin. This second activity is concerned with generating new-product ideas. Wind (1982) and Crawford (1987) describe a number of techniques for generating possible solutions to unfulfilled needs, such as brainstorming, synectics, the suggestion box, and morphological analysis. In relation to these techniques, Wind asserts that they should primarily be used by experts in the specific product-related areas. In this respect, it is important to note that it is much more difficult for consumers to generate solutions for specific problems than to merely discuss problems. For this reason, both Holt (1989) and Von Hippel (1978) feel the applicability of consumer-based techniques for this purpose to be restricted. They state that user cooperation in idea generation is only beneficial in situations where the would-be customer is overtly aware of his new-product need. Consequently, Holt excludes user-cooperation when technology-based products are involved. If, however, the customer has technological
expertise, user-cooperation can be utilised in the case of advanced and complex technology (Mantel & Meredith, 1986). The essential point to take from these comments is that the cooperating respondents need to have some orientation with respect to the new product to be developed.

**Phase C: Concept testing**

The outcomes of a need assessment, as shown in Figure 4, can be ideas or concepts. We define a concept as a new-product idea which is transformed into some kind of lasting medium (textually, pictorially, by means of software statements, etc.). A crucial characteristic of a concept is that its dynamic functions, apart from possible static aesthetic functions, are not yet operative. To illustrate, a concept of a car can not be used to drive.

The attractiveness of the concepts to potential buyers is tested in concept testing. This results in the selection and refinement of concepts. This consumer-research activity, to which we will refer as concept testing, is identical to the idea/concept screening, concept development and concept evaluation of Wind (1982), the concept stage of Hisrich & Peters (1984), concept creation, development and evaluation in Crawford (1987), and design and testing in Urban et al. (1987).

Central questions in concept testing are: what is the most profitable new product that can be made, and what target groups will optimize profits? Two different approaches can be used to answer these questions: concept screening and concept optimization.

**Concept screening**

Concept screening is the selection of concepts for further development. A number of concepts are evaluated on a number of criteria, and the most promising one(s) are chosen for further development. Concept screening can be conducted using either qualitative or quantitative techniques. When using qualitative techniques such as in-depth interviewing, a small number of concepts are presented to a small number of respondents. These qualitative techniques supply a lot of diagnostic information (e.g. see: Iuso, 1975; Yuspeh, 1975; Wilding, 1986). A disadvantage is that because of the participation of only a limited number of respondents, the results can be biased by the specific (idosyncratic) characteristics of those respondents, and that no segments of potential buyers with similar product demands (benefit segments) can be formulated. When using quantitative techniques, in most instances, a large number of concepts are presented to a large number of respondents. In this type of study, respondents are invited to rate (or rank order) each of the alternative concepts on a number of evaluative dimensions. As a result, the conclusion can be drawn as to which of the concepts is preferred most. In addition, some insight is obtained on the characteristics of the potential buyers of the product. A disadvantage of these quantitative screening techniques is that it does not become clear why a specific concept is preferred. For more detailed descriptions of the concept-screening approach, we refer the reader to Locander & Scamell (1976), Hisrich & Peters (1984), and to Baker & Albaum (1986).

**Concept optimization**

Concept optimization was introduced by Zufriden (1977) and Green et al. (1981a), among others. A general assumption in concept optimization is that a product consists of a bundle of attributes (Lancaster, 1966; 1971)\(^6\). The attractiveness of a product for a consumer (how much utility it supplies)
depends on its constituent attributes. In other words, preference is a function of the attributes. In new-product development, to supply the maximum amount of utility to the consumers, one is advised to construct a product which is as close as possible to the optimal combination of attributes.

In concept optimization, the concepts which are presented to the respondents consist of systematic combinations of the product attributes. By collecting evaluations of each of the concepts, the importance of each of the attributes, and the most preferred attribute-levels can be inferred. Patterns of these diagnostic insights can be traced to formulate benefit-segments. Finally, one or more new products having the optimal combination of product attributes can be defined. In this sense, concept optimization is applied for the further refinement of existing concepts.

Concept optimization shares the advantages of both the qualitative and the quantitative concept screening techniques, and avoids their disadvantages. Because of this, concept optimization is more useful than applying either a qualitative or a quantitative concept screening technique. Nevertheless, in some situations, combinations of concept optimization and concept screening are recommendable.

In Section 5.3, we will discuss the consumer-research techniques used in concept optimization in more detail.

Phase D: Product testing

After concept testing, one or more of the most promising concepts can be transformed into mock-ups, or even prototypes of real products. If a functioning product is created, we come to Phase D. In a functioning model (e.g. prototype) the main product functions are installed in the product (for instance, a car which can actually be driven). The consumer testing of the (actual) product, in contrast to testing concepts, therefore, concentrates on the actual use of the product. To test the consumer acceptance of the functioning of the product, potential or prospective buyers are invited to use and evaluate the product. Although prospective buyers are the most appropriate respondents in a product test (De Jonge & Oppendijk van Veen, 1982), for pragmatic reasons (prospective buyers are not always easy to find) mostly potential buyers are invited to participate in the product test. A product test can take place in the consumer's home (called an in-home use test) or in a laboratory. The advantage of a laboratory is that (ergonomic aspects of) the interaction between the product and the consumer can be observed. Batsell & Wind (1980) discuss some methodological issues with respect to applying product tests in a laboratory. Among other things, they stress the biasing influences of product presentation on respondent evaluations.

Also in Phase D, the consumer acceptance of the product-as-marketed is tested. In the product-as-marketed, product aspects such as packaging, distribution points, etc. are included. The main objectives of the research are the prediction of sales volumes, and the evaluation of the effectiveness of potential market-entry strategies. The consumer research technique which is mentioned in literature on this subject is the test-market technique. This technique consists of a field experiment in which an experimental design is followed to test the acceptance of different market-entry strategies. In the practice of market research, the use of the test-market technique is restricted to fast moving consumer goods. Therefore, in this thesis, no attention is given to this technique. For a discussion of the control
problems and cost considerations of the test-market technique we refer the reader to Churchill (1987).

2.3 Practical reasons for focusing on concept testing
In this thesis, the focus is on concept testing. We have several theoretical and practical reasons for its being so. The theoretical reasons will be elaborated upon in the next chapter. The practical reasons for concentrating on concept testing are that:
1. Decisions taken in concept testing have major implications for the subsequent development process. This is stressed by several authors who point out the dramatic consequences of wrong decisions in concept testing. According to Wind (1982), a failure to eliminate a poor idea may result in unnecessary costs and the diversion of resources from more promising ideas. He adds that the deletion of a promising idea may be associated with a lost opportunity cost. For these reasons, Johne (1985) states that it would be particularly advantageous to explore as many viable new-product concepts as possible, as a safeguard against expensive and embarrassing mistakes.
2. Concept testing is a way of obtaining market information at a relatively low cost. The creation of textual, pictorial, or rough three-dimensional concepts is much cheaper than the creation of prototypes. In particular, recent developments in computer graphics make it possible to generate large numbers of systematically varying product-concepts (De Bont, 1990). Such inexpensive concepts can be used in a consumer test and adaptations quickly and inexpensively made.
3. The choice has been made to focus on the redesign of existing products as opposed to major innovations. Redesign is a much more frequent design activity than the development of entirely new products (major innovations). In the case of redesign, Phase B of Figure 3 (need assessment) can be skipped, in most instances. On the basis of market knowledge which is already available, suggestions can be made for changing the existing product (or product line). As a result, concept testing, the consumer-research activity of Phase C, becomes more important.
4. Concept testing is a frequently used and a well-accepted market-research technique. The validity of the assumptions behind concept testing (e.g. the ability to evaluate new-product concepts), and the predictive validity of the obtained results, however, have received little attention from academic researchers.

2.4 Summary and conclusions
In this chapter, the concepts 'product' and 'new product' were defined. In essence, a product is defined as a bundle of attributes, and a new product as a product in which one or more (physical) attributes are modified. To introduce a potentially successful product into a competitive market, the attribute changes have to supply superior customer value. That is, the new product will have to be more successful than competitors in fulfilling consumer needs.

To illustrate the role of consumer research in the development of new products, the stages of the development process were depicted in a comprehensive model. In this model, three parts were distinguished: a strategic part, a planning part, and an operations part. Since most design activities are located in the planning part of the comprehensive model, a
closer look was taken at this part. To this end, four different models were presented indicating major design activities and corresponding phases (Phase A to Phase E) for obtaining required market information (information needs).

In Phase B, first an attempt is made to find unfulfilled (or poorly fulfilled) consumer needs, second to generate solutions (new-product ideas) to these need problems. The consumer research activity which concentrates on finding the unfulfilled needs is need assessment. Different techniques used in need assessment were presented. After one or more new-product ideas have been formulated as new-product concepts, Phase C can begin. In Phase C, the concepts are developed further. The information need consists of obtaining the consumer acceptance of the concepts. By means of concept testing, either the most promising concepts can be selected (concept screening), or an optimal combination of product attributes can be formulated (concept optimization). Following concept testing, a first prototype of the new product can be constructed and tested. In product testing, the ease of use of the new product receives much attention. Finally, when the last product changes have been completed, the new product is almost ready to be introduced into the market. In Phase E, the optimal marketing-mix combination is determined. In some instances, this is performed by using the test-market technique.

In this thesis, we concentrate on Phase C. The practical reasons for doing so are: (i) the vital importance of decisions in concept testing for the subsequent development process; (ii) the relatively low costs involved; (iii) the importance of redesign in new-product development; and (iv) the frequency of its use in market-research practice. In addition to these practical reasons, a number of theoretical reasons will be discussed in the next chapter.
3. The Problem of Testing the Consumer Acceptance of Early Product-Concepts

3.1 The nature of early new-product concepts

In this thesis a concept is defined as the formulation of new-product ideas. From the first moment, any new-product idea which is generated in some kind of development project is transformed into a kind of lasting medium like on paper, or in bytes or polystyrene, or through stereolithography. Such transformations are labelled concepts.¹ By transforming the new-product ideas into some kind of lasting medium, they can be communicated to the relevant persons (experts on materials, marketers, potential buyers, etc.) and comparisons can be made between the different new-product ideas. Once the new-product concepts are developed to the extent that they actually function, we do not label them as concepts any more. It is then that we speak of products instead of concepts.

In a study by De Bont & Looschelder (1992), in which a number of major Dutch market research agencies and major manufacturers of new products were interviewed, it was demonstrated that concepts can be of many kinds. Sometimes strictly textual concepts are being used. In that case, the new-product ideas are written down and no visual support is offered. In other instances, combinations of textual and pictorial information are presented to potential buyers. A particular example of these types of concepts, which are used rather frequently in the Netherlands, are the concepts which are constructed by advertising agencies and combine visuals of the product with textual claims.² Yet another way in which new-product ideas are communicated is by means of three-dimensional models. These models can still be very rough, or they can be highly realistic. In this context, realism refers to the extent to which the finished product is represented in the new-product concepts. Highly realistic, three-dimensional concepts, sometimes called mock-ups, are very expensive to construct. For this reason, when the choice has been made to construct mock-ups, only a limited number are developed.

The three presentation formats which are used most frequently in concept testing in the Netherlands are: (i) strictly textual descriptions, (ii) combinations of textual and pictorial information, and (iii) three-dimensional models (e.g. mock-ups). By comparison, Cattin & Wittink (1982) found that, when speaking of applications of conjoint analysis by American market-research firms, the most frequently used presentation formats were (what they called) verbal descriptions (46%), paragraph descriptions (23%), and pictorial presentations (17%).³

Due to the early stages of the new-product development process, the information comprised by new-product concepts differs from the information comprised by the actual products in a shop. A new-product concept, by definition, is unfinished, whereas a product in a shop contains information on all product attributes. Whereas in a new-product concept even the core product is incomplete, in a shop the core product is always complete. Furthermore, in a shop all aspects of the augmented product are shown (e.g. distribution outlet, in-store advertising). By contrast, in a product-

¹ "Lasting" is used to exclude the verbalization of the new-product idea. In that form, unless it is tape-recorded, the new product idea is not lasting, but volatile.

² These types of concepts are often used in the case of fast moving consumer goods, and much less so in the case of durables.

³ Paragraph descriptions are strictly textual. They differ from verbal descriptions in the sense that they do not explain the product-concept in short statements, but describe the product-concept in an entire paragraph.
concept, only in a limited number of instances are some aspects of the augmented product shown (e.g. brand name). In addition, when in a shop desired information about (an attribute of) a product is missing, questions can be posed to the sales personnel.

To describe differences between early product-concepts and actual products, we introduce the theoretical concept "degree of realism". As stated above, realism refers to the extent to which the finished product is represented in new-product concepts. We put 'degree of' in front of realism to indicate that large differences exist with respect to the realism of product-concepts.

Within one product-concept, differences can exist in the degree of realism in the information which is supplied on each of the product attributes. The information contained by the attributes can have a high degree, an intermediate degree or a low degree of realism.

With a high degree of realism, detailed (a lot of) information is given on all relevant aspects of an attribute, in such a way that a clear picture is obtained of how this attribute is represented in the final product. This can, for example, concern the attribute "form" of a product. At a high degree of realism, a realistic, three-dimensional mock-up is presented showing all aesthetic details.

At an intermediate degree of realism, information on an attribute is only partly present. Some information is given on an attribute, but a complete picture of how this attribute will appear in the final product is not formed. In the main, attributes at this intermediate degree are of two kinds. An attribute can be presented with little detail (e.g. a global model to represent the form of a product without aesthetic details), or an attribute can be presented in a different presentation format (e.g. textually or pictorially, instead of in three-dimensions). Different combinations of presentation format and amount of detail can be thought of. Information which is presented at an intermediate degree of realism can be regarded as ambiguous information. Consumers who are being confronted to product information at this level have to put much effort in interpreting the information. Due to the ambiguity different consumers may come to different interpretations of the same information.

At a low degree of realism, information on a given attribute is absent. Sometimes, some information on a relevant attribute can (correctly or erroneously) be inferred from existing information.

Considering the degree of realism in an early product-concept, it can be stated that the degree of realism in the attributes of concepts is lower than that in actual products in nearly all instances. Degree of realism has two aspects which convey the fundamental differences between early product-concepts and actual products: amount of information (amount of detail) and presentation format. The more information that is given, the higher the degree of realism. With respect to presentation format, generally it can be stated that the three-dimensional presentation format is the most realistic, followed by the pictorial presentation format, which is more realistic than the textual presentation format. Within a given presentation format, the amount of attribute information can vary to a large extent. A textual description of an attribute can, for example, consist of just a few words, or consist of a paragraph of several pages. For this reason, a globally presented pictorial concept can be less realistic than a textually presented concept containing much detail. In Figure 6, a line is drawn representing the shift from a low to a high degree of realism.
In most instances, information on some attributes of new-product concepts is given at an intermediate degree of realism, and at the same time information on other (relevant) attributes is given at a low degree of realism. A high degree of realism is seldom reached in concept testing.

In market-research practice, the degree of realism of the concepts depends on two aspects: the amount of information present at the time of a concept test (development stage), and the research budget (De Bont & Loosschilder, 1992). In the early stages of the new-product development process, the functional attributes in most cases can only be represented textually. The attribute form can be represented either textually, pictorially or even three-dimensionally. In most cases, however, this attribute will be presented without much detail. A new-product concept, therefore, can either be entirely textual, or consist of a combination of both textual and pictorial information, or of a combination of textual and three-dimensional information. In all three cases, there are substantial differences in the degree of realism between early product-concepts and actual products.

3.2 The problem of evaluating early product-concepts

In Chapter 2, we mentioned several practical reasons for concentrating on concept testing. The theoretical reason will be described in this section.

The testing of the consumer acceptance of product-concepts in the early stages of the new-product development process poses a major methodological problem. In the early stages of the new-product development process, by definition, product designers still have to decide on many aspects. Some decisions are taken early in the design process, such as those about global form, or the presence/absence of attribute-levels. Other decisions are taken later in the process, such as those about colour, position of attributes, and form details. As a result, in concept testing only a limited amount of information on the new product can be demonstrated to consumers. Moreover, the presentation formats of the new-product concepts often differ from those of actual products. New-product concepts are, for instance, presented as global sketches, textual descriptions, three-dimensional models, or combinations of these. An example is given in Figure 1. Consequently, the new-product information is presented either at an intermediate or at a low degree of realism. This fact

---

5 Theoretically speaking, animation is also a possibility for demonstrating a functional attribute. Because of practical reasons, such as complications both in creating and presenting the stimulus material, however, this option is seldomly used.
conflicts with a generally accepted rule of thumb in concept testing. This rule of thumb states that new-product concepts should be presented as realistically as possible. This requirement is mentioned by numerous authors, such as Sherak (1966), Green & Srinivasan (1978), Batsell & Wind (1980), Actio & Hustad (1981), Hills (1981), Wind (1982), Finn (1985), and Holbrook et al. (1985). The importance of this requirement is formulated by Finn (1985, p.37) in the following manner: “For the prediction to have value, the stimulus presented at the time of the concept test must convey to the subjects the same meaning that they would extract from a marketplace exposure to the product at a later time of launch”.

When consumers are confronted with concepts which only contain limited product information, they may face major difficulties in evaluating the specific new-product concepts. Possibly, the new-product concept is not understood well enough to grasp its benefits (e.g. Laric, 1979; Virzi, 1989). Triesscheijn (1982) states that sometimes consumers lack abilities to imagine how the new product will appear. In addition, Crawford (1987) asserts as a major weakness of concept testing that, because of the fact that product information is ambiguous, there are numerous opportunities for misunderstanding. As a result, the meanings conveyed by the concepts are likely to differ from the meanings which would be extracted in a later stage of the new-product development process (Finn, 1985). In that case, the early evaluations are of no value. When, nevertheless, these evaluations are interpreted as valid responses, they may endanger the new-product development process.

In conclusion, the nature of early product-concepts poses an interesting problem. This problem consists both of theoretical and practical aspects. With respect to the theoretical aspects, we are very much interested in the matter of how the early product-concepts are evaluated by consumers. To what extent can differences be found between the evaluations of early product-concepts and actual products? In this thesis we will give attention to the theoretical aspects by attempting to answer the first research question:

1. In what respects are early product-concepts evaluated differently from actual products?

   The practical aspects of the problem are related to the theoretical aspects. If consumers find it very difficult to evaluate the early product-concepts, the question arises as to what the value is of the results obtained in early product-concept testing. Possibly, the reliability and the validity of the evaluations which are made early in the new-product development process are very low. If this is true, it is not a wise thing to invest in product-concept testing. Before we draw such a negative conclusion (if at all), we will first attempt to answer the second research question:

2. How reliable and how valid are evaluations based on early product-concepts?

In Chapter 3, a major problem was introduced which consumers face when they are invited to evaluate product-concepts in the early stages of the new-product development process. The quintessence of this problem is how consumers evaluate early product-concepts, for as Finn (1985) states, early-product concepts may convey meanings that are different from those of actual products. Correspondingly, the first research question which was posed at the end of Chapter 3 is: in what respects are early product-concepts evaluated differently from actual products?

In this chapter, in order to answer the first research question, we take a closer look at the process of product evaluation. This process is discussed in full detail. In doing so, we will present different models of product evaluation. These models give some insight into the process of product-concept evaluation.

We will propose that the models of product evaluation are based on a very general so-called two-stage model. On the basis of an integration of the different models of product evaluation, we will determine a Model of Product-Concept Evaluation. In order to better understand the effects on evaluations of the nature of early product-concepts, the insights obtained from the models of product evaluation will be broadened using more general findings from theories of information processing. After this, following the Model of Product-Concept Evaluation, the processes of perceiving and evaluating early product-concepts are discussed. Finally, in order to answer the first research question, hypotheses will be formulated about (among others) the effects on evaluation of the variables we distinguished within the nature of early product-concepts, i.e. amount of information and presentation format. The formulation of these hypotheses is guided by the proposed Model of Product-Concept Evaluation.

4.1 Models of product evaluation

In the context of new-product development, different but related models of product evaluation have been proposed. In this thesis, four models have been selected to be discussed in more detail. These models are:

- Two-Stage Model of Evaluative Judgment (e.g. Holbrook, 1981)
- Model of Hierarchical Information Integration (Louviere, 1984)

The first two models have been very influential in the literature on product evaluation. The third model offers an interesting modification of the second model. The fourth model is less well-known in the literature, but is included because, among other aspects, it elaborates upon individual differences.
Below, a brief, and therefore incomplete description of each of the models is given. After this, in the next section, it will be shown that different terminology is used in these models for similar phenomena, and that the main differences are found in the phenomena which are stressed most.

**Model of Information Integration**

Anderson has developed the Information Integration Theory. Anderson (1981) distinguishes the following processes in this theory: stimulus valuation, integration, and response.

Stimulus valuation is the chain of processing that transforms the physical stimulus into its psychological counterpart. Valuation consists of two components: the perception of the product and the evaluation of single attributes. In the Information Integration Theory, physical stimuli are observables which can be controlled in experiments. The corresponding individual psychological interpretations, by contrast, can be of many kinds.

After valuation, the evaluations of the perceived attributes are integrated into an internal response. This is the integration process. The integration process constitutes the central topic of Information Integration Theory, since most attention is allocated to the matter of how different product attributes, each of which is evaluated (either positively or negatively), are integrated into an overall evaluative response. According to Anderson, this integration process takes place along simple algebraic rules (e.g. averaging, subtracting, multiplying). Anderson uses cognitive algebra as an analogy to this cognitive activity.

Finally, the overall internal response is expressed by the (overt) response.

Anderson introduced functional measurement to demonstrate that numerical values can be obtained for the algebraic representations. In essence, functional measurement shows a close resemblance to the concept-optimization technique called conjoint analysis¹ (e.g. Green and Srinivasan, 1978). In both techniques, the same procedure is followed. First, stimulus attributes are systematically varied in attribute profiles. Second, the profiles are evaluated by respondents. Third, the relationships, between attributes on the one hand and evaluations on the other, are numerically specified.

According to Stokmans (1991), however, a major difference between both techniques is that in functional measurement more attention is paid to the modelling of the integration function. While a linear additive model is often assumed in conjoint analysis, in functional measurement no specific model is presupposed to be operating. By means of fitting different models to the preference data, the best fitting model is determined empirically. Following Information Integration Theory, this best-fitting model represents the algebraic (integration) rule which has been used to reach an evaluative response.

In Figure 7, a schematic description of the Model of Information Integration is given.

![Diagram](image)

**Figure 7:** Schematic description of the Model of Information Integration

¹ This technique will be discussed in more detail further on.
Two-Stage Model of Evaluative Judgment

The Two-Stage Model presented by Holbrook (1981) pays most attention to the perception of product stimuli. The rationale behind this is the conviction that the meanings which an individual attaches to a stimulus determine his evaluation of that stimulus. In the Two-Stage Model of Evaluative Judgement, the process of perceiving a stimulus is described as inferential belief formation. Inferential belief formation is the process in which objective, physical attributes of a stimulus are perceived by an individual who transforms this information into subjective “percepts”, of different attributes in particular, and of the stimulus as a whole.

Inferential belief formation is a constructive (bottom-up) process in which a subset of all aspects of the stimulus (the so-called cues) are perceived. Especially striking aspects of the stimulus (e.g. contrasting colours), and attributes in which an individual is interested beforehand, attract attention. The perceived attributes are interpreted and beliefs about the stimulus are constructed. The percepts (or beliefs), therefore, can be regarded as inferences or interpretations of the objective information.

An example of inferential belief formation is when, from the objective attribute "small", in the case of a car, it is inferred that the car is made in Japan. In this example, the attribute (cue) is the size of the car, and the belief (percept) is that the car is Japanese.

Several models of inferential belief formation have been developed in the psychological literature. In this respect, we agree with Steenkamp (1989), that Brunswik’s Lens model has been very influential. Based on the Brunswik Lens model (Brunswik, 1943; 1955; 1956), two-stage models have been developed in new-product research (e.g. Huber, 1975; Hauser & Simmie, 1979; Holbrook, 1981; Hauser, 1984; Steenkamp 1989). The essence of the Two-Stage Model is that subjective attribute perceptions intervene between ‘objective’ product features and evaluative judgment. In the first stage (Stage One), on the basis of attributes (the objective, physical attributes), subjective beliefs (percepts) are formed. This is referred to as “psychophysical relations” by Holbrook (1981). In Stage Two, the perceptions are integrated into an evaluative response. In this second stage, some of the beliefs which have been formed are reflected in the evaluation, whereas other beliefs are given no, or very little, importance. Holbrook uses the term “attribute integration” to refer to the second step.

The Two-Stage model of evaluative judgment can be visualized in the following way:

![Diagram of Two-Stage Model of Evaluative Judgment]

Figure 8: Schematic description of The Two-Stage Model of Evaluative Judgment

Model of Hierarchical Information Integration

The Model of Hierarchical Information Integration presented by Louviere (1984) focuses on the first stage of the Two-Stage Model of Evaluative Judgment. Following the hierarchical approach, in order to simplify decision-making when there are many product attributes, consumers organize
individual decision attributes into clusters or subgroups. Consumers then evaluate and aggregate some property of each of the subgroups to reach an overall judgment.

An example of the Model of Hierarchical Information Integration is given by Louviere & Gaeth (1987). In evaluating a supermarket, one can think of many attributes, like prices of meats, price of dairy products, the range of products of meats, travel time, opening hours, the quality of dairy products, etc. In this example, four subgroups of attributes are created. These are: price, range, convenience, and quality. The subgroup “price” comprises prices of meat, produce, packaged goods, and dairy products. In evaluating a particular supermarket, first each subgroup is given an evaluative score. In the final evaluative response, the set of reduced evaluative scores is integrated.

The Model of Hierarchical Information Integration has led to applications of decompositional techniques (discrete-choice analysis) in which new-product concepts are presented which contain attribute-level information on more than twenty attributes (e.g. Oppewal, 1992). When compared to actual products, which possess information on a wide range of attributes, this number is not extraordinarily large. Nevertheless, to avoid the danger of information overload in the attribute profiles, in line with Green & Srinivasan (1978), we think that, when inviting consumers to evaluate attribute-profiles, it is better to restrict the attribute-level information to a maximum of 5 or 6 attributes.3

In Figure 9, a schematic description of the Model of Hierarchical Information Integration is given.

![Diagram](image)

Figure 9: Schematic description of the Model of Hierarchical Information Integration

Model of the Consumer Evaluation Process for New Products and New-Product Concepts

Finn (1985) presents what he calls a theory of the consumer evaluation process for new-product concepts. This theory is grounded in both economic concepts (e.g. level of consumer’s consumption technology, benefit production process) and psychological concepts (e.g. psychophysics, imagery).4 In line with the two-stage models, first some product cues (attributes) are perceived. In interpreting the product-concept, the consumer makes use of his prior (direct or indirect) experiences with respect to the same type of products (human capital). His current state in terms of possessions (consumption capital) is taken also into account. The combination of human capital and consumption capital, in evaluating new-product concepts, is labelled the consumption technology. In the model presented by Finn, product attributes, marketing controlled variables, and consumption technology explain whether or not a consumer actually buys a product. By then using the product, its benefits can be enjoyed.

In evaluating a product-concept, a consumer attempts to anticipate the benefits of the product. In addition, the evaluation is affected by the imagery

---

3 In this respect, studies by Scott and Wright (1976), and by Cattin and Weinberger (1980), showed that the number of attributes (6 versus 9) was negatively related to the validity of the results.

4 Finn (1985) does not explain what he means when discussing imagery.
of the product, which is the result of marketing variables and of the behavior of others. Through imagery, the attributes which constitute the so-called augmented product are included in the evaluation.

The Model of the Consumer Evaluation Process for New Products and New-Product Concepts proposed by Finn is the result of integrating insights from different disciplines. In our opinion, the most important contribution of Finn's model is that it relates individual differences to the process of product evaluation. The model explains why the amount of benefit perceived in a new-product concept differs between individuals.

With respect to this model, some points of criticism can be mentioned. To begin with, in our opinion, it is not correct to restrict imagery to marketing-controlled and situational variables. Product attributes such as form and colour can also give rise to imagery. Furthermore, the model is not empirically tested. Therefore, the insights it provides need to be tested before they can be accepted.

A schematic depiction of Finn's model is given in Figure 10.

![Diagram](image)

Figure 10: Schematic description of the Model of the Consumer Evaluation Process for New Products and New-Product Concepts

4.2 Integration of models of product evaluation into a Model of Product-Concept Evaluation

The insights into the perception and evaluation of new-product concepts, which we derived from the four models in the previous section, can be summarized in three observations.

1. All models of product evaluation are in a sense two-stage models: processing intervenes between stimuli (products) and the forming of an evaluation. The model presented by Holbrook (1981) comprises an overall framework.

To illustrate how the different models really only use different terminology to refer to similar phenomena, we present Figure 11.
<table>
<thead>
<tr>
<th>Processing product information</th>
<th>Integrating</th>
<th>Evaluative Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage One</td>
<td>Stage Two</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anderson (1981)</th>
<th>Valuation</th>
<th>Integration</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holbrook (1981)</td>
<td>Psychophysical relations</td>
<td>Attribute integration</td>
<td>Response</td>
</tr>
<tr>
<td>Louviere (1984)</td>
<td>Subgroup perception</td>
<td>Subgroup judgment</td>
<td>Response</td>
</tr>
<tr>
<td>Finn (1985)</td>
<td>Psychophysical relations</td>
<td>Benefit imagery</td>
<td>Overall evaluation</td>
</tr>
</tbody>
</table>

Figure 11: Models of product evaluation presented as two-stage models

To clarify Figure 11, Holbrook uses “inferential belief formation” and Finn uses "psychophysics" (c.f. Stevens, 1975) when referring to psychophysical relations.

From the discussion of the models, it becomes clear that they either stress the process of processing product information (Holbrook, 1981; Louviere, 1984) or the process of forming an evaluation (Anderson, 1981; Finn, 1985). Figure 11, nevertheless, shows that all four models distinguish both stages. Therefore, given the generality of the two-stage model, it is most likely that in the evaluation of early product-concepts both processing product information and forming an evaluation will take place.

2. An algebraic formula is followed in the integration of the valued attributes into a final evaluative response (Anderson, 1981). The Information Integration Theory pays most attention to the matter of how differently evaluated attributes are combined to give an overall preference score.

3. Both product attributes and individual characteristics of consumers, determine the process in which the attributes are perceived and interpreted in terms of benefits (Finn, 1985).

To conclude, the Two-stage Model of Evaluative Judgment sets out an overall framework which comprises all three other models. In the first stage, on the basis of attributes (the objective, physical attributes), subjective beliefs (percepts) are formed. This is referred to as "psychophysical relations" by Holbrook (1981). We prefer to speak of "processing of product information". By doing so, instead of restricting the first stage to psychophysical processes only, more general, cognitive notions of information processing can also be included (e.g. those of Greenwald & Leavitt, 1984).
In the second stage, the processed product-attributes are integrated to form an evaluative response. In this second stage, some of the beliefs which have been formed are reflected in the evaluation, whereas other beliefs are given no, or very little, importance. Holbrook uses "attribute integration" to refer to the second stage. Since not the attributes themselves, but its evaluative loadings are at stake, following Anderson, we prefer to speak of integration only. In addition, in line with Finn (1985), we think it is necessary to include consumer characteristics, since different consumers will evaluate the same stimulus in different manners. Due to the expectation that some consumers will be more capable of handling early product-concepts than others, the influence of consumer characteristics may be especially important when discussing the evaluations of early product-concepts.

To encapsulate the comments we have made with respect to the models of product evaluation, we postulate The Model of Product-Concept Evaluation. This model can be visualized in the following way:

![Diagram of the Model of Product-Concept Evaluation]

Figure 12: The Model of Product-Concept Evaluation; a schematic description

In our opinion, the Model of Product-Concept Evaluation suggests a more fruitful starting point than Holbrook's model for trying to understand the effects of the nature of early product-concepts on evaluations. We see two reasons for this.

First, by broadening the processing of product information from the theory of belief formation to a more cognitive orientation, general theories on information processing can be included to explain the processing of early product-concepts.

Secondly, by incorporating consumer characteristics into the model, attention can be given to individual differences in evaluating early product-concepts.

Below, we will discuss each of the two stages of the Model of Product-Concept Evaluation in more detail. In doing so, elements of the models of product evaluation will be complemented with other theoretical perspectives.

*Stage One of the Model of Product-Concept Evaluation: Processing product information*

Meyer (1987) stated that the process of concept evaluation is inseparable from more general principles of information processing. This correlates with the recognition of such in the first stage of the Two stage-models. We agree with this, and therefore, will start with a brief discussion of some basic principles of information processing.
The information-processing approach to communication was largely established by Hovland, Janis & Kelley (1953), and by McGuire (1969). The sequential steps presented by those authors are often only slightly adapted by later researchers, such as Edell & Staelin (1983), or Greenwald & Leavitt (1984). In their hierarchical processing model, Greenwald & Leavitt distinguish the following sequential steps: pre-attention, focal attention, comprehension, and elaboration. These steps are depicted in Figure 13.

![Diagram](image)

Figure 13: Hierarchical processing model of Greenwald & Leavitt (1984)

The information-processing model of Greenwald & Leavitt is hierarchical, meaning that only when significant message content is detected at lower levels of processing, is the next level of processing invoked. Below, we will briefly discuss each of the steps.

Pre-attention

In the pre-attention stage, a particular stimulus elicits an orienting response. The orienting response, requiring little capacity, consists of mild physiological arousal together with the physical orienting of receptors toward the source of stimulation. In this way, the stimulus, e.g. an attribute profile of a filter coffee-maker, is first detected.

Focal attention

For focal attention, a modest capacity is allocated to focus on one message source, and to decipher the message's sensory content into categorical codes. Thus, the observer identifies the stimulus, e.g. a particular attribute profile indicates that the filter coffee-maker has a dripstop.

Comprehension

During comprehension, further capacity is required for analyzing the message by constructing a propositional representation of it. In the Two-Stage Model of Evaluative Judgment, this process is described by the process of inferential belief formation. On the basis of what is seen (attributes), beliefs are constructed. In the Model of Hierarchical Information Integration, Louviere (1984) reports how different attributes, which are meaningfully related, are clustered. In order to simplify decision-making when there are many product attributes, consumers organize individual decision-attributes into clusters or subgroups. In this respect, we have already discussed an example of the Model of Hierarchical Information Integration from Louviere & Gaeth (1987). Essentially, during comprehension the observer attempts to understand what is meant by the information content, and thereby, beliefs are constructed; e.g. the attribute profile says the filter coffee-maker has a dripstop, meaning that when I take the can away, no coffee will drip on the heating element.

Elaboration

In elaboration most capacity is required to enable the integration of
message content with existing conceptual knowledge. The stimulus which is elaborated upon is interpreted within a larger framework. Possibly, comparisons are made with similar stimuli, and then related personal experiences which are stored in memory are activated.

The observer interprets the information content to find out what it means for his personal situation, e.g. because the dripstop prevents coffee pouring onto the heating element, it will not get dirty so easily. In other words, information content comes to be seen in terms of functional benefits.

In the model presented by Finn (1985), elaboration receives a lot of attention. In this model, in line with the Two-stage models, first some product cues (attributes) are perceived. After this, the perceived attributes are interpreted in terms of benefit. In Finn’s terminology, benefit refers to an individuals’ estimation of the amount of utility which will be supplied by a particular product. The amount of benefit deduced from a product or an attribute, apart from the intrinsic value of that product or attribute, depends on the individuals’ prior product experience (the ability to estimate the desirability of different aspects of the new product), and on declining marginal utility. In addition, imagery, which depends on society’s associations with the product and is a function of price, place and promotions, also influences the amount of benefit. Finn considers benefit to be an output of the consumption process. By obtaining, preparing, using, and disposing products, consumers have developed a frame of reference for evaluating the benefits of new products. In Reynolds and Gutman’s (1984) terms, such consumers are capable of inferring both the functional and psychophysical consequences of particular product-attributes.

Stage Two of the Model of Product-Concept Evaluation: Integrating processed attributes into an evaluative response

Stage Two is given more attention in the Information Integration Theory (Anderson, 1981) than in any of the other models. Although both stages of the Two-Stage Model of Evaluative Judgment are discussed in Anderson’s theory, the second stage is stressed most. Anderson distinguishes between valuation, integration, and response. Only valuation (the perception of the product and the evaluation of single attributes) is located in stage one.

The most important contribution of the Theory of Information Integration is to be found in the notion of integrating the processed attributes. After individual evaluation, the evaluations of the perceived attributes are integrated into an internal response. The central topic of the Information Integration Theory is about how different product attributes, each of which is evaluated (either positively or negatively), are integrated into an overall evaluative response. This is the process of stimulus integration. Anderson (1981, p. 2) states that “stimulus integration is the central concept”. The conviction discussed by Anderson, that thought and behavior depend on the joint action of multiple stimuli, explains the centrality of stimulus integration. Anderson tries to understand how effective stimuli are combined (or integrated) to produce the evaluative response. He also seeks to find out which are the effective stimuli.

According to Anderson, the integration process takes place along simple algebraic rules (e.g. averaging, subtracting, multiplying). He uses cognitive algebra to specify this cognitive activity. Empirical support for the use of a specific algebraic rule is supplied by Troutman & Shanteau (1975). These authors found that evaluations of disposable diapers and infant car seats
were formed by averaging the attribute beliefs.

Anderson does not predict which algebraic rule is applied in which situation, but to find this out he introduces the technique called functional measurement. To include the possibility of testing different integration models, full factorial designs\(^6\) are used in this technique.

Another technique which concentrates on the integration of processed attributes into an overall response is conjoint analysis. In conjoint analysis, in most cases, it is assumed that respondents use the linear additive model to integrate the evaluated attributes into the evaluative response. Empirical support for applying this model is supplied by several studies. We mention Slovic & Lichtenstein (1971), Steenkamp et al. (1986), and Stokmans (1991). According to Slovic & Lichtenstein (1971), the linear model accounts for all but a small fraction of predictable variance in judgments.

If one assumes that the linear additive model holds, one can apply fractional factorial designs. These designs exclude the possibility of detecting the interaction effects of specific combinations of attributes on the evaluative responses, and only estimate the main effects of the attributes. A practical advantage of fractional factorial designs, when compared to full factorial designs, is that more attributes and/or more attribute levels can be included. The reason for this is that when applying full factorial designs, including those with even a small number of attributes and a small number of attribute levels, the number of attribute profiles which have to be evaluated by respondents can be very large. To illustrate, in the case of three attributes with three attribute levels each, the number of attribute profiles is twenty-seven (3 \times 3 \times 3). When applying a fractional factorial design, however, the entire master design is not presented, meaning that with the same number of attribute profiles, more attributes and/or attribute levels can be presented. We feel that this is a real advantage in fractional factorial designs. So, in spite of elegance of Anderson’s theory, we assume that the linear additive model is valid.

4.3. The evaluation of early product-concepts: the formulation of hypotheses

In Chapter 3, it was mentioned that the two dimensions characterizing the nature of early product-concepts are “amount of information” and “presentation format”. Here we will describe the consequences of both these variables on the processing of early product-concepts and on the evaluative responses. In addition, possible effects of the consumer characteristic “product-category knowledge” will be discussed.

**Effects of “amount of information” on the processing of early product-concepts**

The model of Greenwald & Leavitt (1984) explains possible differences between the processing of attribute information at an intermediate degree of realism and the processing of actual products. In both instances, consumers will supply evaluative responses when they are invited to evaluate either product-concepts or actual products. Nevertheless, the manner in which the (new-)product information is processed may vary. More specifically, differences in the level of processing (from shallow to deep) might be observed. Only if a respondent in a concept test goes through each of the steps of information processing (from preattention to elaboration) when being confronted with new-product concepts, can it be said that a deep level of processing is followed.
Figure 14: Deep-level processing of early product-concepts

Sometimes, an attribute is comprehended but not elaborated upon. For instance, this can happen when a respondent cannot (or is not motivated to\textsuperscript{7}) think of how he would use the new product in his home environment, in spite of understanding the product attributes. In that case, beliefs are constructed, but there is no evaluation of the specific attribute levels. An intermediate level of processing is followed.

Figure 15: Intermediate-level processing of early product-concepts

Due to a low level of realism in early product-concepts, the motivation to go to higher levels of processing may be small. The information in early concepts will then only be processed at low levels, and the processing will (for the most part) not reach the comprehension or elaboration stages. If the existing attribute information is not comprehended, the consequence is that no beliefs are constructed. If the existing attribute information is miscomprehended, beliefs are constructed which do not correspond with the beliefs which should be constructed when the product is finished.

Figure 16: Shallow-level processing of early product-concepts

From the above, it can be expected that early product-concepts are processed at lower levels than actual products. To find out whether or not this is true, we formulated the following hypothesis\textsuperscript{8} (H1):

\textit{H1: If the amount of information contained in early product-concepts is small, a more shallow level of processing will be followed than if the amount of information is large.}

The distinctions between systematic and heuristic processing strategies (Chaiken, 1980), and between piecemeal and category-based processing

---

\textsuperscript{7} Petty et al. (1983) discuss the effects of lack of ability and/or lack of motivation on the processing of product information.

\textsuperscript{8} We use hypotheses although research questions would have been more appropriate. The reason is that we want to avoid confusion with the two research questions we already formulated.
(Sujan, 1985), closely resemble the distinctions between levels of processing that we have discussed here. Systematic processing strategies and piecemeal processing refer to a systematic investigation and thorough interpretation of the information content before responding. Heuristic processing strategies and category-based processing refer to the use of simple rules or cognitive heuristics based on the directly available information stored in memory. Systematic processing strategies and piecemeal processing are indicators of a deep level of processing. Heuristic processing strategies and category-based processing are indicators of a shallow level of processing. Wright (1975) introduced the "affect referral heuristic" to indicate a type of processing in which a consumer does not examine attributes or beliefs about alternatives. Although Bettman (1979) expects this heuristic to be operating either when a consumer has a lot of experience in buying a particular product, or in the case of low-involvement products, we think this heuristic may also operate if the available product information is not easily processable. In these instances, when invited to evaluate new product-concepts, we expect consumers either to use the bits of relevant information which are available in memory, or to rely on simple heuristics which have proved to be successful in other situations.

Effects of "amount of information" on the evaluative responses

In this section, two effects of "amount of information" are discussed. The first effect concerns that of an attribute which is presented with little detail on its evaluations. The second effect concerns that of presenting an attribute with little information, or indeed none at all, on the evaluations of the attribute "price".

According to Green & Srinivasan (1978) when little detail is shown in the concepts, much is left to the imagination of the consumer. When asked to evaluate early new-product concepts, consumers try to construct an image of the product by making inferences from the information which is present. These inferences may be different from the ones which would have been made if much more detail was presented. Consequently, the evaluations may be affected. Therefore, we formulated the following hypothesis (H2):

$$H2: \text{If the amount of information presented about an attribute is small, the evaluative responses of this attribute differ from responses of the same attribute presented with a large amount of information.}$$

At a low degree of realism, information on an attribute is missing. When, in the eyes of consumers, the missing attribute is relevant, inferences may be made from the present attributes to the relevant absent attributes (e.g. see Huber & McCann, 1982)\(^9\). As a consequence of different inferences about early product-concepts, final overall evaluative responses may be affected.

It has been found in many studies that the absence of important attribute information affects the overall evaluations (e.g. Huber & McCann, 1982; Johnson & Levin, 1985; Johnson, 1987). The overall evaluations of present attributes can be affected in two manners. First, the levels of an attribute A can be evaluated less positively when information on an attribute B is missing (discounting). Second, the marginal effect of changes in a present attribute A (sometimes referred to as "sensitivity") can be affected by the absence of information on attribute B (slope effect). To illustrate, Huber & McCann (1982) demonstrated that inferences influenced the marginal value.
of attributes in the directions they predicted. The study was conducted with
textual descriptions of beers. Information was given on one of the two
attributes (price and taste) or on both. Omitting price diminished the
marginal value of taste. Omitting taste information had an interactive effect
on the marginal value of price, increasing it at lower levels and decreasing it
at higher levels.

In contrast to the large number studies reporting either discounting or
slope effects, Simmons & Lynch (1991) found that few inferences were
made from present attributes to absent attributes when collecting process-
tracing data. A possible explanation may be found in the method which was
chosen by the investigators, i.e. retrospective comments. In our opinion,
considering hindsight bias and rationalisation processes, this method yields
results with questionable validity. We think, therefore, that inferences are
made from unambiguously presented attributes to missing or ambiguously
presented attributes.

An important contribution to the understanding of the inferences was
made by Rao and Monroe (1988). These investigators showed that
respondents having either a low or a high familiarity with a certain product
category displayed stronger positive price-perceived quality effects in
product evaluation than did moderately familiar respondents (slope
effects)\(^\text{10}\). The explanation given for this was that respondents with low
levels of familiarity consider as meaningful only extrinsic product information
(such as price or brand name). They cannot infer benefits from the intrinsic
attributes. In Greenwald and Leavitt's (1984) terms, respondents with little
product-category knowledge cannot elaborate upon the product
information. So, although information on intrinsic attributes is present, it
cannot be used in the evaluations. Respondents with high levels of
familiarity do know how extrinsic cues are related to the intrinsic cues, and
can therefore rely on the extrinsic ones.

In general, it can be stated that if relevant information is absent or
ambiguous\(^\text{11}\), as in the case of early product-concepts, one has to rely on
the cues which are present and which can be (easily) interpreted. This is
instrumental in reducing the risk involved in making a choice. In particular,
the attribute "price" appears to be used to reduce the risk involved in making
a choice. As such, inferences are expected from the attribute "price" to
relevant absent or ambiguous attributes. To find out whether or not this
expectation holds, we formulated Hypothesis 3.

\(^\text{H3: If information on a relevant attribute is missing or ambiguous, the levels of another relevant attribute (e.g. price) will be used as an indication of the level of the missing or ambiguous attribute.}\)

Effects of "presentation format" on the processing of early product-concepts

A number of studies concentrate on the effects of presentation format on
type of processing. When new-product information is presented pictorially,
this results in a more holistic processing according to Holbrook & Moore
(1981). Holistic processing refers to a simultaneous processing of all the
informational cues of a stimulus. The authors base this conclusion on the
observation that more statistically significant attribute interactions are found
in the evaluations of pictorial product-concepts than in the evaluations of
textual product-concepts. These results, however, are not supported by
Domzal & Unger (1985). In addition, Holbrook (1983) himself admits that
colourful words may trigger a holistic type of processing.
Even if the results of these studies had been in the same direction, we expect that the generalizability of the results is limited in practice. The reason for this is that the above-mentioned studies concentrated on differences between entirely textual and entirely pictorial product-concepts. In the practice of new-product concept-testing, the latter presentation format (entirely pictorial) is seldom used (De Bont & Looschelder, 1992). In most cases, the functional attributes one wants to include can only be represented textually in that early stage of the new-product development process. Consequently, the early product-concepts consist either of entirely textual information or of a combination of textual and pictorial information. Thus, we expect that an analytical, in contrast to a holistic type of processing will be generated, since in all these presentation formats textual information is present. Whether or not this expectation holds needed to be tested. Thus we formulated the following hypothesis (H4):

\[ H4: \text{If product-concepts consist of combinations of both pictorial and textual information, this will not lead to a more holistic processing than if product-concepts only consist of textual information.} \]

**Effects of "presentation format" on the evaluative responses**

The effects of "presentation format" on evaluations have been investigated by different researchers. More specifically, the evaluations based on entirely textual concepts have been compared with evaluations of concepts consisting of combinations of textual and pictorial information.

Taubner (1972) found that pictorial concepts were evaluated more positively than textual concepts. Since this study was not a conjoint-analysis study, however, the effects of presentation format on the importance and on the part-worths of the attribute levels could not be assessed. Oppedijk van Veen & Beazley (1977), who did perform a conjoint-analysis study, found that when an attribute was presented pictorially, this resulted in the assignation of more extreme part-worth utilities to that attribute than when this attribute was presented textually.\(^\text{12}\)

Louviere et al. (1987) studied the effects of presentation format (entirely textual attribute representations, or combinations of textual and pictorial attribute representations) using a conjoint-choice model. The main difference between conjoint analysis (as discussed in this dissertation) and the conjoint-choice model is that in conjoint analysis respondents are invited to evaluate profiles of new-product concepts, whereas in studies based on the conjoint-choice model respondents are invited to choose between different profiles of new-product concepts. Louviere et al. (1987) found minimal differences in the part-worths when comparing the results of both presentation formats. Statistically significant differences at the .05 level were found for only two out of twenty-one part-worth utilities.

Since, however, the findings of Tauber (1972) and Oppedijk van Veen & Beazley (1977) on the one hand, and the findings of Louviere et al. (1987) on the other hand, are not in the same direction, we still do not know whether or not various presentation formats influence product evaluations. Considering that the main purpose of the study by Louviere et al. (1987), however, was to find out the effects of presentation format on consumer evaluations (in contrast to the study of Oppedijk van Veen & Beazley, 1977), and that insights were obtained about part-worths (in contrast to Tauber, 1972), we are strongly inclined to generalize the outcomes of this study to our problem, rather than to generalize those of the other studies. We expect,
therefore, that presenting information on some attributes pictorially instead of textually does not affect the evaluations of those attributes. This expectation has led to Hypothesis 5 (H5).

**H5: If attribute information is presented textually, this will not yield evaluative responses which are different from those when presenting the same attribute information pictorially.**

**Effects of consumer characteristics on evaluative responses**

The Model of the Consumer Evaluation Process for New Products and New-Product Concepts by Finn (1985) explains that, in the case of minor innovations and depending on the amount of prior product-experience, evaluating early product-concepts may be difficult for some consumers but not for others. Following this line of reasoning, in the case of major innovations where not even a single consumer has prior product-experience, evaluating will be a difficult task for all consumers. A close parallel is found in literature on the effects of consumer knowledge on product judgments (cf. Alba & Hutchinson, 1987). Experts find attribute statements informative, whereas novices consider benefit statements informative (e.g. see: Conover, 1982; Walker, Celsi & Olson, 1987).

This difference can be explained by the inability of novices to infer benefits from attributes. Chi et al. (1981) suggest that experts are likely to elaborate upon the message information by evaluating it in relation to their prior knowledge, whereas novices are likely to represent message information more or less literally in memory. This is confirmed in a study by Maheswaran & Sternthal (1990), who demonstrate that experts process information on attributes in more detail than novices. In their study, Maheswaran and Sternthal found that the level of processing depends both on the way the product information is presented and the amount of product-category knowledge. The authors investigated the effects of presenting attribute information, either in terms of attribute statements or in terms of benefits, on the level of cognitive processing. These effects were determined for both experts and novices with respect to personal computers. The authors found that attribute statements for experts, and benefit statements for novices, were the best ways of presenting the attribute information, since those respective ways of presenting attribute information evoked a deep level of processing. In contrast to the novices, the experts were capable of inferring benefits from the attributes. On the basis of these findings, we expect that consumers with large amounts of product-category knowledge are better equipped to evaluate early product-concepts than consumers with little product-category knowledge. Consumers with a lot of product-category knowledge can infer benefits even from early product-concepts. Therefore, evaluations based on the attributes of early product-concepts, from consumers with a lot of product-category knowledge, will not differ to a large extent from their evaluations based on attributes of further developed concepts, or on those of products from the original early product-concepts. To test this expectation, we formulated Hypothesis 6.

**H6: If consumers have little (a lot of) product-category knowledge, they will demonstrate different (identical) evaluative responses of the same attribute presented with different amounts of information.**
4.4. Summary and Conclusions

In this chapter, four models of product evaluation were discussed. It appeared that the Two-Stage Model of Evaluative Judgment possesses a framework in which the other models and other theoretical notions can be integrated. This integration led to the Model of Product-Concept Evaluation. This model can also be considered a two-stage model. The most important differences between the Two-Stage Model of Evaluative Judgment and the Model of Product-Concept Evaluation are that in the latter, broader cognitive frameworks are included in the first stage, and that "consumer characteristics" is included as a factor which mainly affects the second stage. Each of the two steps of the Model of Product-Concept Evaluation were discussed in detail.

Next, the effects of presenting early product-concepts were discussed. The distinction between "amount of information" and "presentation format", which we made in Chapter 3, was applied in this chapter. By doing so, it was possible to concentrate on the separate effects of those factors, both on processing and evaluative responses. In addition, the effects of product-category knowledge on the evaluation of early product-concepts were considered. This discussion resulted in the following hypotheses:

**H1:** If the amount of information contained in early product-concepts is small, a more shallow level of processing will be followed than that if the amount of information is large.

**H2:** If the amount of information presented on about an attribute is small, the evaluative responses of this attribute will differ from responses of the same attribute presented with a large amount of information.

**H3:** If information on a relevant attribute is missing or ambiguous, the levels of the attribute "price" will be used as an indication of the level of the missing or ambiguous attribute.

**H4:** If product-concepts consist of combinations of both pictorial and textual information, this will not lead to a more holistic processing than that if product-concepts only consist of textual information.

**H5:** If attribute information is presented textually, this will not yield evaluative responses which are different from those when presenting the same attribute information pictorially.

**H6:** If consumers have little (a lot of) product-category knowledge, they will demonstrate different (identical) evaluative responses of the same attribute presented with different amounts of information.

These hypotheses are summarized in Figure 17.
Figure 17: Schematic representation of hypotheses about the Model of Product-Concept Evaluation

The hypothesized relationships are suggested by arrows in the figure. In the next chapter, several experiments are described that were conducted to test the hypotheses.
5. Operationalizations for Testing Hypotheses about the Evaluation of Early Products-Concepts

In Chapter 4, a number of hypotheses were formulated. In Figure 17, an overview of the hypotheses is given. In this chapter, attention will be given to the operationalization of the variables which were described in the Model of Product-Concept Evaluation. These variables are the following: product-concept information, processing, evaluative response\(^1\), and consumer characteristics. The variable "integrating" is not discussed because of the fact that we assume a linear additive integration process (see Section 4.2).

It is worth noting that no attempt was made to test the Model of Product-Concept Evaluation as a whole. This model was employed to formulate hypotheses about the processing and evaluation of early product-concepts. Whether or not these hypotheses held was what we wanted to find out.

The decision was made to test the hypotheses in a series of experiments in a laboratory. The main reason for doing this was the need for experimental control. Among other things, this comprised the systematic presentation of product-concepts, and the systematic registration of processing and evaluation. In addition to this, experimental control comprised the minimization of disturbing external effects.

5.1 Product-concept Information: choice of test product

In the experiments we conducted to test the hypotheses, we decided to use filter coffee-makers as stimuli. The reasons for using filter coffee-makers were that:

- the filter coffee-maker is a highly penetrated durable in the Netherlands. In the description of the household panel of the Delft University of Technology (Tan, 1989), one finds that almost eighty-seven percent (86.6) of the (557) households owns at least one filter coffee-maker
- filter coffee-makers are frequently used by different members of the household. For this reason, different persons in a household can be interviewed
- the replacement-purchase rate of filter coffee-makers (filter coffee-makers are rebought about every five or six years) is longer than that of most durables\(^2\).

As a result of these facts, it was relatively easy to identify potential buyers and potential users of new filter coffee-makers.

Before testing hypotheses in quantitative research, it is often recommended that qualitative research be conducted (Verhallen & Vogel, 1982). Qualitative research can, for instance, contribute to a better understanding of why particular attributes are important to consumers. Therefore, to gather some preliminary insight about filter coffee-makers, three group discussions (N=26) took place (De Bont, 1988). In these discussions, various issues were dealt with, namely the functions of coffee drinking, the functions of filter coffee-makers (as opposed to alternative methods of preparing coffee), and the relevance of specific attributes in

---

\(^1\) Sometimes also referred to as "evaluations".

\(^2\) The importance of the replacement-purchase period in testing the predictive validity of the evaluations is elaborated upon in Chapter 8.
usage and purchase situations. It was found that, when compared to alternative methods of preparing coffee, the filter coffee-maker was felt to be less labour intensive. Other activities, such as entertaining guests, can be performed while the coffee is being made. Another finding was that, when asked directly, respondents said the most important attributes were (in order of relevance) form, price, capacity, having a removable water reservoir, and having a thermosflask. These attributes were included, therefore, in the stimuli we used in our experiments.

The main differences between early product-concepts and actual products, as described in Chapter 3, are the amount of available information and/or presentation format. In the hypotheses which were formulated (H1 to H6), these differences are reflected. In these hypotheses, either the amount of information (H1, H2, H3, H6) or presentation format (H4, H5) constituted the independent variable. To be able to assess the effect of each of the variables, one variable was kept constant while the other was varied. Below, we will discuss how these variables were operationalized.

### 5.1.1 Product-concept information: amount of information

As was detailed in Figure 6 in Chapter 3, information on specific attributes in early product-concepts can have an intermediate degree or a low degree of realism. At an intermediate degree of realism, the attribute information is presented in less detail than, and/or in a different presentation format from that of actual products. At a low degree of realism, information on a particular relevant attribute is absent. Here, we will first discuss the operationalization of "amount of information" at an intermediate degree of realism.

**Amount of information: intermediate degree of realism**

To determine the effect of the amount of information, this variable was varied within a specific presentation format. In Figure 6 of Chapter 3, it is indicated that for each of the presentation formats (three-dimensional, pictorial, and textual), the amount of information can be small (little detail) or large (a lot of detail).

To test H1, within the pictorial presentation format, two types of stimuli were created: schematic pictorial stimuli and realistic pictorial stimuli. The main difference between both types of stimuli was that the schematic pictorial stimuli contained fewer form details than did the realistic pictorial stimuli. The schematic pictorial stimuli only illustrated the global form of the new product, and thus provided little detail.

The following steps were taken to create both sets of stimuli:

1. Seven filter coffee-makers, all available on the Dutch market, were bought. The products possessed a large variety of combinations of attributes, but all the filter coffee-makers were identical with respect to colour (white) and capacity (10 cups). The prices varied from thirty-five guilders to one-hundred-and-twenty guilders.³
2. Based on these seven filter coffee-makers, a student of the Faculty of Industrial Design Engineering⁴ created the schematic pictorial stimuli on a special computer system (Apollo 3000D) using the C.A.D.-package "Microsolid" (Horner, 1988). A description of the generation of product-concepts using Microsolid is given by De Bont & De Lind van Wijngaarden (1991).
3. Black and white photographs of both the real and the schematic filter

³ One guilder was equivalent to about fifty American dollar cents at the time of writing.
⁴ Mees de Lind van Wijngaarden
coffee-makers were taken at equal angles and with neutral backgrounds. Brand identification was disguised. The photographs of real filter coffee-makers represented realistic pictorial stimuli. Photographs of the schematic pictorial stimuli were taken directly from the computer screen. In both instances, the photographs were taken by a professional photographer\textsuperscript{5}.

In the figure below, the same filter coffee-maker is shown twice, but on the right-hand side with a lot of detail, and on the left-hand side with little detail.
Figure 18: The schematic pictorial stimuli (on the left) and the realistic pictorial stimuli (on the right)
To test H2 and H6, the amount of information on the attribute form was also varied within the three-dimensional presentation format. To construct the schematic three-dimensional stimuli from the real three-dimensional stimuli, the following steps were taken:

1. Three real filter coffee-makers were selected for their clear representations of different form structures (Muller, 1990).
2. On the basis of these three filter coffee-makers, three (three-dimensional) mock-ups were created by a student of the Faculty of Industrial Design Engineering\(^6\). This student received the instruction to build mock-ups which had the same form (and size) as the real products, but which did not show form details. The mock-ups were painted in the same colours as the real filter coffee-makers (white).

In Figure 19, pictures are shown of the three representations of levels of the attribute “form” for both sets of stimuli. (It is worth noting that respondents were not presented with the photographs, but with the real filter coffee-makers or with the mock-ups).

\(^6\) Maarten ten Houten
Douwe Egberts

Figure 19: Photographs of the schematic three-dimensional stimuli (on the left), and the realistic three-dimensional stimuli (on the right)

*Amount of information: low degree of realism*

For testing H3, the effect of the amount of attribute information was measured at a low degree of realism. The concern was here with the effects of missing information on a relevant attribute. To test Hypothesis 3, a comparison was made between one set of stimuli in which information on a relevant attribute was missing, and another set of stimuli in which all information on this attribute was given.

The first step in creating the first set of stimuli was to determine which attribute was highly relevant for filter coffee-makers. Here, we made use of the results of the group discussions on filter coffee-makers (De Bont, 1988). In these group discussions, the attribute "form" was mentioned most frequently when respondents were asked which product aspects they would consider when buying a new filter coffee-maker. From this, we concluded that form is a relevant attribute to consumers. Subsequently, in the stimulus set in which information on a relevant attribute was missing, no information was given on the attribute "form".

The attributes, which were included in the set of stimuli in which information on the attribute "form" was missing, were price, a dripstop, a thermostop, a removable water reservoir, and capacity. These attributes were also mentioned as relevant attributes in the study by De Bont (1988).

<table>
<thead>
<tr>
<th>With termosflask</th>
</tr>
</thead>
<tbody>
<tr>
<td>With removable water reservoir</td>
</tr>
<tr>
<td>120 guilders</td>
</tr>
<tr>
<td>Without high speed</td>
</tr>
<tr>
<td>Without dripstop</td>
</tr>
</tbody>
</table>

Figure 20: Example of a stimulus without information on the attribute "form"
5.1.2 Product-concept information: presentation format

The independent variable 'presentation format' (in H4 and H5) was varied with respect to the attribute "form" by using either the pictorial presentation format or the textual presentation format. To construct two matching sets of information on the attribute "form", one in the pictorial and one in the textual presentation format, a procedure was followed which has been used by Edell & Staelin (1983) and by Unnava & Burnkrant (1991). This procedure consisted of the following steps:

1. The seven schematic pictorial representations in Figure 18 were taken as a starting point.
2. Twenty-five respondents (members of the P.E.L.-household panel) were shown all seven pictorial representations. The respondents were asked to imagine that they had to describe the filter coffee-maker on a telephone to a friend who takes an interest in the shape of products. If the respondent had difficulties in completing this task, the interviewer played the role of the friend by asking what the filter coffee-maker looked like.
3. The descriptions given by the subjects were transcribed, resulting in a list of all expressions made per coffee maker.
4. For each filter coffee-maker, the five expressions that were made most frequently were taken out to serve as input for the textual descriptions of the attribute "form". In these descriptions, the expressions were ranked from general to specific in order to facilitate comprehension of the final product-concepts.
5. After having constructed a description for each of the seven pictures, we conducted a manipulation check to find out whether or not the descriptions properly matched the pictures. For this purpose, twenty-three respondents were recruited (members of the P.E.L.-household panel) who had not participated in a test on filter coffee-makers before (De Bont & De Graaf, 1989). The respondents were instructed to take a brief look at the descriptions of the seven pictures, plus three fake descriptions. Next, one of the seven pictures was chosen randomly, and the respondents were asked to say which of the ten descriptions matched the picture. This was repeated until all seven pictures were matched.
6. The three pictures which were matched correctly most of the time (21, 18, and 17 times out of 23, respectively) were selected for use in the main investigation. (One filter coffee-maker was matched correctly by all respondents. This filter coffee-maker, however, was not selected for the main investigation because of its highly deviant shape)\textsuperscript{7}.

\textsuperscript{7} If a highly deviant form is chosen, it is probable that respondents either strongly like or strongly dislike this level of the attribute. Consequently, artificially large importance of the attribute would be generated.

The selected pictures and the corresponding textual descriptions are presented in Figure 21.
- Modern machine, built from an old fashioned idea
- Edgy machine with edgy surface components
- Machine with many unintegrated components
- Round coffee jug with a fat belly

Philips

- Machine has the form of a cylinder, rusk jug
- Integrated form
- Moonshaped water reservoir at the rear
- Revolving filter
- Large, round handle

Braun

- Ordinary machine, traditional
- Machine has an oval form
- Integrated, because the filter and jug are entirely embedded in the machine
- Filter stands directly on the jug
- Rectangular, robust handle, bottom side open

Severin

Figure 21: Pictorial representations of the attribute “form” (on the left) and textual representations of that attribute (on the right)

5.2 Processing: level of processing and type of processing

In this dissertation, “processing” is the dependent variable for testing in hypotheses 1 and 4. Processing stands for “level of processing” in H1, and for “type of processing” in H4. Operationalization of both of these will be discussed.
Level of processing

To measure level of processing, Greenwald and Leavitt (1984) briefly discuss so-called orienting tasks. The "orienting-tasks" method consists of a number of tasks with increasing difficulty. These tasks are performed by subjects after they have seen a particular stimulus. The task with the highest degree of difficulty, for which correct answers are given, indicates the level at which a respondent processed the stimulus. A clear description of the different tasks, however, was not given by the authors. Consequently, this method was not used here.

Instead of using the orienting tasks, and in line with Sujan (1985) we decided to use the "cognitive response" method. The cognitive response method is a variation of verbal protocol analysis in which consumers are asked to describe what they are thinking or doing when they are being confronted with a particular stimulus. The verbal statements are coded, and hypotheses can be tested. Examples of verbal protocols can be found in Nisbett & Wilson (1977), Ericson & Simon (1980), and Wright (1980). According to Wright, verbalized thoughts are valid indicators of processing activities.

To assess the level of processing from the verbalizations, a coding scheme was determined. This coding scheme was, for the most important part, based on a scheme used to investigate the level of message-processing by Sujan (1985). Sujan investigated the effect of product-category knowledge on the processing of product-attribute information. To this end, she presented attribute information which matched expectations and attribute information which did not match expectations. She measured both the type and the number of verbalizations. Sujan developed the coding scheme by examining the protocols of the first few respondents. The types of verbalizations in Sujan's scheme were:

- Simple evaluative verbalizations (e.g. overall evaluations)
- Attribute-oriented verbalizations (e.g. attribute evaluation, attribute clarification)
- Categorization verbalizations (e.g. product is similar to other products in product category)
- Subtyping verbalizations (e.g. product is related to a more specific subcategory of product)
- Discrepancy verbalizations (e.g. product differs from other products in a product category)
- Other verbalizations (e.g. product imagery, disbeliefs).

In the case of matching information (category-based processing), more verbalizations related to the product category (categorization), fewer to product attributes, and fewer to subtypes than in the case of a mismatch with category knowledge (piecemeal processing). Sujan found that piecemeal processing (deep level of processing) corresponded with a high proportion of verbalizations\(^8\) related to message content (e.g. attribute-oriented verbalizations), and with a low proportion of verbalizations dealing with categorization or simple evaluations. Category-based processing (shallow level of processing), by contrast, corresponded with a high proportion of verbalizations dealing with categorization or simple evaluations, and with a low proportion of attribute-oriented verbalizations. Sujan concluded, therefore, that attribute information which matches prior expectations is processed at a more shallow level than non-matching attribute information, which is processed at a deep level. Subtyping

---

\(^8\) As a reminder, Sujan (1985) investigated category-based versus piecemeal processing.

verbalizations and discrepancy verbalizations were not directly related to level of processing.

From Sujan’s coding scheme, we borrowed the distinction between attribute-oriented verbalizations on the one hand and simple evaluations and categorization (referred to as total-product-oriented verbalizations) on the other. Attribute-oriented and total-product-oriented verbalizations respectively, comprised the first and the second main categories of our coding scheme.

In addition, when looking at the protocols of two respondents (N=2) in a pilot study, within both main categories of our coding scheme, the following subcategories needed to be formulated:

- need for clarification (e.g. what is the function of this device?)
- factual description¹⁰ (e.g. this filter coffee-maker is white)
- known/unknown¹¹ (e.g. I have seen such a filter coffee-maker before)
- evaluation (e.g. a nice thermosflask)
- statements on buying intention (e.g. I would never buy such a filter coffee-maker)
- associations (e.g. looks like a juke-box).

The subcategories were introduced to gain a more precise insight into the differences in inferences resulting from both types of stimuli. In particular, the numbers of both factual descriptions (beliefs) and evaluations were of interest.

Furthermore, and also based on the pilot study, we observed that a third main category was needed: reactions to the way the new product is presented.

So, the coding scheme used here was as follows:

<table>
<thead>
<tr>
<th>1 Attribute-oriented verbalizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>- need for clarification</td>
</tr>
<tr>
<td>- factual description</td>
</tr>
<tr>
<td>- known/unknown</td>
</tr>
<tr>
<td>- evaluation</td>
</tr>
<tr>
<td>- statements on buying intention</td>
</tr>
<tr>
<td>- Associations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 Total-product-oriented verbalizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>- need for clarification</td>
</tr>
<tr>
<td>- factual description</td>
</tr>
<tr>
<td>- known/unknown</td>
</tr>
<tr>
<td>- evaluation</td>
</tr>
<tr>
<td>- statements on buying intention</td>
</tr>
<tr>
<td>- Associations</td>
</tr>
</tbody>
</table>

| 3 Reactions to presentation of the product |

The coding of the verbalizations was performed by two judges. The judges were uninformed about the hypotheses and the treatment conditions. For each judge, a list was prepared containing all verbalizations. Disagreement was resolved through discussion, so that all verbalizations were fully coded.

The proportion of attribute-oriented verbalizations (a.o.v.) divided by the sum of this proportion and the proportion of total-product-oriented
verbalizations (p.o.v.), comprised the level of processing. The higher this coefficient \([\frac{a.o.v.}{a.o.v. + p.o.v.}]\), the deeper was the level of processing.

**Type of processing**

In testing H5, processing means type of processing. Type of processing (holistic or analytical) was operationalized via the same two measures as in the study by Holbrook & Moore (1981). The first measure consisted of counting the numbers of attribute interactions. The second measure was the so-called Hays’ omega-squared.

Counting attribute interactions

In this study, in line with the study by Holbrook and Moore (1981), a large number of statistically significant attribute-interactions would indicate a holistic type of processing; a small number would indicate an analytical type of processing.

To be able to assess attribute interactions, a conjoint analysis with a full factorial design was applied. In a full factorial design, all possible attribute combinations (the entire master design) are presented to the respondents. A multi-linear model is used to simultaneously estimate the main-effects and the interaction effects. In the case of three attributes, the multi-linear model can be described as follows:

\[
P_{ijk} = b_1(a_i) + b_2(a_i) + b_3(a_k) + b_4(a_{ij}) + b_5(a_{ik}) + b_6(a_{jk}) + b_7(a_{ijk}) + c,
\]

where:

- \(P_{ijk}\) = preference for product-concepts consisting of the attributes i, j, and k.
- \(b_i\) = regression weight i
- \(a_i\) = attribute i
- \(a_{ij}\) = interaction between attribute i and attribute j
- \(a_{ijk}\) = interaction between the attributes i, j, and k
- \(c\) = constant

An attribute interaction is statistically significant if the corresponding regression weight exceeds a specific critical value. The interactions between two attributes (e.g. \(a_{ij}\)) are referred to as first-order interactions, the interactions between three attributes (e.g. \(a_{ijk}\)) as second-order interactions.

**Hays’ omega-squared**

Holbrook & Moore (1981) introduced Hays’ omega-squared as a measure of holistic processing. This measure is calculated by dividing the strongest main effect by the strongest interaction effect, for each individual. To aggregate over individuals, the median ratio is taken. A low median ratio indicates that the attribute processing has been holistic, a high median ratio indicating an analytical type of attribute processing.

**5.3 Evaluative response: the choice for conjoint analysis**

The variable “evaluative response” was the dependent variable for testing in hypotheses 2, 3, 5 and 6. The operationalisation of evaluative response is determined by the choice of a particular concept-testing technique. In Chapter 2, the approaches “concept screening” and “concept
In addition to compositional and decompositional techniques, hybrid techniques have also been developed (see Green et al., 1981b). The hybrid techniques share both the advantages and the disadvantages of the compositional and the decompositional techniques.

A detailed description of the Expectancy-Value Model can be found in Van Raaij (1977).

Compositional techniques are sometimes referred to as the self-explicated model (e.g. Green and Srinivasan, 1978).

The Fishbein Attitude Model originated in social psychology and has become very influential in market-research and consumer psychology.

Optimization were introduced and briefly discussed. In the early stages of the new-product-development process, we prefer concept optimization to concept screening. Compared to concept screening, concept optimization has the advantage that a deeper analytical insight is obtained by directly relating the consumer evaluations to the attributes. Consequently, not only is a better understanding obtained of why certain concepts (attribute combinations) are liked or disliked, but also predictions can be made about the acceptance of new combinations of given attribute levels. Possibly, the most promising concept is a combination of attribute levels not yet presented to consumers. Therefore, in this thesis we will concentrate on the concept-optimization approach.

Consumer-research techniques in concept optimization aim to obtain an insight into the relationship between product attributes on the one hand and preference on the other. The outcomes are presented in preference models, such as the multi-linear model, which numerically specify the structure of these relationships. In describing consumer-research techniques in the concept-optimization approach, a major distinction can be made between compositional and decompositional techniques (e.g. see: Green & Srinivasan, 1978; Holbrook, 1981; Vriens & Wittink, 1990).

Compositional techniques

Compositional techniques comprise techniques which have been directly inspired by the Expectancy-Value Model of Rosenberg (1956). In the context of concept testing, applying compositional techniques takes place along the following lines:

1. A limited number of concepts is sequentially presented;
2. Combinations of questions are posed about the desirability of particular attribute levels, and on the importance of a particular attribute;
3. The overall utility of concepts is determined by calculating the weighted sum of the attribute-level ratings.

With compositional methods, the preferences for the attribute levels and the importance of the attributes are obtained from consumers by means of direct questioning. An important example of a frequently used compositional technique is the so-called Fishbein Attitude Model (Fishbein, 1975).

Decompositional techniques

With decompositional techniques, the procedure differs in certain respects from compositional techniques. To demonstrate these differences, we first present the steps which are followed when applying a decompositional technique. These are:

1. A number of systematically varied concepts are sequentially presented;
2. Consumers are invited to evaluate, or choose between, the concepts;
3. The overall preference ratings (or rankings) of the concepts are decomposed into the separate contributions of the attribute levels. That is, statistical procedures are applied to estimate the attribute importances (sensitivities), and the attribute-level preferences (part-worths), from the overall preference ratings (rankings).

An example of a decompositional technique is conjoint analysis (Green & Srinivasan, 1978; 1990).
Choosing between compositional and decompositional techniques

In Green & Srinivasan (1990), and in Vriens & Wittink (1990), a number of advantages of decompositional techniques are given, compared with compositional techniques. The most important ones are that:

- decompositional techniques supply information on the (relative) importances of attributes as a function of changes in attribute levels. This information is relevant when attribute-level changes are considered, for instance in a new-product development program. These insights make it possible to anticipate the effects on consumer acceptance of changing an attribute level.

- in contrast to compositional techniques, in decompositional techniques the direction of causality from attribute levels to preference is clear. In compositional techniques, the overall impression of the product may direct the attribute evaluations. Having a positive (negative) impression of a new-product concept affects the evaluations of single attributes in a positive (negative) way. This so-called 'Halo-effect' (e.g. see Beckwith & Lehmann, 1975) results in artificially high correlations between evaluations of different attribute levels of the same product.

In accordance with Green & Srinivasan (1990) and with Vriens & Wittink (1990), in this study we prefer decompositional techniques over compositional techniques.

Choosing a specific decompositional technique

The preference for decompositional techniques, however, does not imply that we think all decompositional techniques are suitable for the development of new products. For the purpose of selecting the most appropriate technique, we specify the following four pragmatic criteria:

1. Insight must be obtained on actionable attributes. (Actionable attributes are physical attributes which can be manipulated by the new-product designer).\(^\text{16}\)
2. The possibility must exist for incorporating both qualitative and quantitative attributes.
3. Insight into attribute importances and attribute-level preferences are needed at the individual level.
4. A lot of insight is required at a relatively low cost.

The first criterion excludes 'Joint-Space models' (e.g. PREFMAP; see Carroll, 1972). The output of these models consists of a number of concepts which are mapped onto a limited number of (mostly subjective) perceptual dimensions. On the basis of these maps, the researcher has to infer the dimensions by comparing the positions of the preferred alternatives with those of the non-preferred alternatives. These perceptual dimensions often consist of combinations of different aspects which cannot be easily manipulated.

The second criterion excludes both the ideal-point model and the vector model described by Green & Srinivasan (1978), in which only quantitative attributes (measured at the interval level) can be incorporated. The vector model even has an additional disadvantage, i.e. that curvilinear preferences for levels of an attribute cannot be detected. Only the part-worth model\(^\text{17}\) is capable of handling both qualitative and quantitative attributes. When applying the part-worth model, curvilinear preferences can be detected.

\(^\text{16}\) In addition to physical attributes, new product designers often attempt to design new products which satisfy consumer demand for subjective (non-physical) attributes. In most instances, these attempts concern the form (and colour) of the new product. The consumer-research technique "conjoint analysis" has been developed to determine the relative importance of a number of different attributes, which may or may not include subjective attributes.

\(^\text{17}\) An explanation of the part-worth model is given later in this section.
The third criterion is based on the assumption that different consumers have different preference structures. To account for heterogeneity in preferences, preference models should be estimated at the individual level (Timmermans, 1984). By doing so, one overcomes the fallacy of designing a product from averaged preference data. This fallacy is called the 'majority fallacy' by Kuehn & Day (1962). Based on similarities in individual preference models, sometimes different segments or target groups can be specified, and individual preference models can be aggregated.

This third criterion excludes techniques which estimate preference functions at the aggregate or at the segment level [see, for instance, Louviere & Woodworth (1983), and Louviere (1988)].

The fourth criterion excludes techniques which make use of very large numbers of consumers, or which restrict the amount of information that can be obtained (number of attributes and attribute levels). For this reason, functional measurement, which is advocated by Anderson (1981; 1982) and in which all possible attribute combinations (a full factorial design) have to be evaluated for the sake of determining the best fitting mathematical integration function, is excluded. The use of full-factorial designs either demands an increase in the number of consumers having to evaluate the concepts, or an undesirable reduction in the number of attributes and attribute-levels in the concepts.

Conjoint analysis (if restricted to the part-worth model) is the only technique which meets the above-mentioned criteria. Conjoint analysis was developed by mathematicians and mathematical psychologists (Luce & Tukey, 1964; Kruskal, 1965; Carrol, 1972; Srinivasan & Shocker, 1973). The first consumer-research oriented paper on conjoint analysis was written by Green & Rao (1971). Conjoint analysis has received an enormous amount of scientific and managerial attention.

An overview of conjoint analysis in consumer research is presented by Green and Srinivasan (1978). In this overview major issues on implementing conjoint-analysis are discussed. These issues are:

1. The model of preference (e.g. vector model, ideal-point model or part-worth model);
2. The data-collection method (e.g. trade-off method or the full-profile method);
3. Stimulus set construction for the full-profile method (e.g. fractional factorial design or full factorial design);
4. Stimulus presentation (i.e. presentation format);
5. The measurement scale for the dependent variable (e.g. rating, ranking);
6. The estimation method (e.g. MONANOVA, PREFMAP, LINMAP, JOHNSON's non-metric trade-off algorithm, OLS).

Wittink & Cattin (1989) demonstrate the popularity of conjoint analysis. Because of practical reasons, the authors advocate: (i) the use of the full-profile data-collection method, (ii) rating as measurement scale for the dependent variable, and (iii) ordinary-least-squares analysis (O.L.S.) as the estimation method. In Appendix 1, an illustration is given of an application of conjoint analysis in the context of new-product development.

To test H2, H3, H5 and H6, the evaluations of the attribute-levels of form were determined by applying the part-worth model. The part-worth model can be considered a regression model with dummy variables. The formula of a part-worth model appears as follows:
\[
P_{a_dj} = b_1(a_1) + b_2(a_2) + b_3(d_1) + b_4(d_2) + b_5(d_3) + c,
\]

where:

\[
P_{a_dj} = \text{preference for product-concepts consisting of the attributes } a \text{ and } d.
\]

\[
b_k = \text{regression weight of attribute-level i or } j \text{ (= part-worth utility)}
\]

\[
a_i = \text{level i of attribute a}
\]

\[
c = \text{constant}
\]

In this formula, attribute \( a_i \) consisted of three attribute levels, and attribute \( d_j \) consisted of four. In the regression analysis, to avoid multicollinearity, for each attribute \( n-1 \) attribute-levels are included in the model. The standardized regression weights represent the part-worth utilities. The part-worth utilities of the excluded attribute-levels are deducted from the ones which were included, since the sum of the part-worths of an attribute equals zero.

From the formula, it becomes clear that the attribute-levels are subjected to regression analysis as independent factors.

To test the hypotheses H2, H3, and H6, an O.L.S.-analysis was performed on the ratings of the full profiles (concepts). In this analysis, the part-worths for two of the three attribute levels were estimated.

5.4 Consumer characteristics: product-category knowledge

The effect of product-category knowledge on evaluations of early product-concepts was tested in Hypothesis 6. Here, we discuss the operationalisation of this consumer characteristic.

The measures of consumer product-category knowledge used in studies of consumer psychology fall into two categories. These categories are objective measures and subjective measures. Objective measures cover the amount, type, or organization of the information stored in memory (e.g. Brucks, 1985; Sujan, 1985; Rao & Monroe, 1988). The subjective measures comprise the individual's perception of how much he knows about a given product (e.g. Johnson & Russo, 1981; 1984; Alba, 1983; Srull, 1983).

Some authors proclaim the use of objective measures. According to Selnes & Grønhaug (1986), objective measures are preferable to subjective measures when the research goal is related to the consumer's ability to encode new information or to discriminate and choose between product alternatives. The authors state that a subjective measure reflects how confident a respondent is about having enough information to make a brand choice. Brucks (1985) and Selnes & Grønhaug, nevertheless, show that there is a considerable relationship between subjective and objective measures, (average correlation coefficients vary from .38, p ≤ .100, to .54, p ≤ .01). Cole et al. (1986) investigated the validity of different measures of prior knowledge on two game products. They did find high correlations (.47, .56, p ≤ .001) between objective and subjective measures in the case of a single product (convergent validity), and low correlations (.02, .10) in the case of different products (divergent validity). These results suggest that in most cases subjective knowledge implies objective knowledge. However, due to the fact that differences do exist between the results of both measures, but more importantly because the goal of our study was related to consumers' ability to encode new information, we preferred the use of objective measures. In our opinion, an objective measure would be the best.
indicator of the actual level of knowledge.

Our operationalisation of product-category knowledge was a modification of the measure used by Brucks (1986). The categories which were included in our measure were:

- product attributes: the number of attributes recalled (Q1)
- general and specific attribute evaluation: the number of attributes mentioned which discriminate between an expensive and a cheap product (Q2)
- brand facts: the amount of brands recalled (Q3), and the number of correctly identified brands from six photographs with camouflaged brand names (Q4)
- purchasing and decision-making procedures: price estimates of the cheapest and most expensive products on the market (Q5).

In the first question (Q1), the respondents were not only asked to recall as many attributes as possible, but were also invited to draw a coffee maker. An example of a drawing is presented below:

![Figure 22: Examples of drawings of a filter coffee-maker](image)

The attributes which were present in the drawings and not written down in the recall task were added to the total number of attributes recalled. This task was also performed in the study by Selnes & Grönhaug (1986). By including the drawing task, a possible disturbing factor (that some respondents mention trivial attributes when invited to recall as many attributes as possible (Q1), whereas others hesitate to mention those attributes) was controlled for. In the drawings, one possible trivial attribute, coffee jug, was drawn by almost all respondents.

When looking at the answers to the five questions, the first issue which attracted attention was that the respondents performed very poorly on Q4. Nearly fifty percent of the respondents did not succeed in correctly identifying even one of the brands from the six photographs. The discriminating power of this item was therefore very low. Moreover, those respondents who did correctly identify the brand of one of the photographs simply recognized their own filter coffee-maker, in most cases. Consequently, those respondents whose filter coffee-maker was not included performed worst, which means that there was a systematic bias in the results. We therefore excluded this item from further analysis. To test
the reliability of the remaining four items, we conducted a retest eight months after the test. The test-retest correlations found were: .47 (Q1), .49 (Q2), .61 (Q3), and -.03 (Q5). With the exception of Q5, all test-retest correlations were found to be statistically significant at the .05 level. Because of the low test-retest correlations we found for Q5, we also excluded this item from further analysis.

The answers to the remaining questions (Q1 to Q3) were used to create one product-category knowledge score for every individual. In order to give all questions equal importance, the individual scores on each of the questions were standardized. Subsequently, the standardized scores were summed. In a reliability test, standardized Cronbach's Alpha was found to be .75, which is acceptable in the early stages of predictive tests of a construct.\(^\text{18}\) The test-retest correlation of the overall product-category knowledge score was .59. From this, we concluded that our measure of product category-knowledge was reliable.

The final individual scores on prior knowledge were rank ordered and subsequently split into three equal-sized groups of respondents (trichotomized), representing the low, the moderate, and the high prior-knowledge groups. The reason for creating three groups instead of two was that we could determine more precisely at which level of product-category knowledge respondents evaluated early product-concepts identically to actual products (test of H6). In addition, by using three groups, we catered for the possibility of detecting curvilinear relationships.

5.5 Summary

In Chapter 5, attention was given to the operationalization of the variables which were described in the Model of Product-Concept Evaluation. These variables were "product information", "processing", "evaluative responses", and "consumer characteristics".

Product information: Amount of information and presentation format

The filter coffee-maker was selected to test the hypotheses because it is relatively easy to identify potential buyers and potential users of new filter coffee-makers. Early product-concepts were created for filter coffee-makers. The underlying dimensions (amount of information and presentation format) of early concepts were independently varied. Amount of information was determined at an intermediate degree of realism and at a low degree of realism. At the intermediate degree of realism, early product-concepts (the schematic pictorial stimuli and the schematic three-dimensional stimuli) only showed global information on the attribute "form", whereas actual products (the realistic pictorial stimuli and the realistic three-dimensional stimuli) showed all information on this aspect. At a low degree of realism, information on the attribute "form" was not included in one set ("without form" condition) but included in the other sets ("with form" condition).

The independent variable "presentation format" was varied with respect to the attribute "form" by using the pictorial presentation format and the textual presentation format. Two matched sets of information on the attribute "form" were constructed.

Processing: level of processing and type of processing

Processing stands for both level of processing and type of processing. To gain an insight into the level of processing of early product stimuli, verbal protocol analysis was used. To assess the level of processing of the
descriptions, a coding scheme was determined. The coding of the verbalizations was performed by two judges. The proportion of attribute-oriented verbalizations (a.o.v.), divided by the sum of this proportion and the proportion of total-product-oriented verbalizations (p.o.v.), comprised the observed level of processing. The higher this coefficient \( \frac{a.o.v.}{a.o.v. + p.o.v.} \), the deeper was the level of processing.

Type of processing (holistic or analytical) was determined in the same manner as in the study by Holbrook & Moore (1981): that is, by counting the number of attribute interactions, and by calculating Hays' omega-squared. A large number of statistically significant attribute interactions, and a low median value of Hays' omega-squared, would indicate an holistic type of processing. A small number of attribute interactions and a high median value of Hays' omega-squared, would indicate an analytical type of processing.

**Evalitative response: conjoint analysis**

To measure the evalulative responses we first chose a concept-testing technique. Because of a number of advantages in decompositional techniques (when compared to compositional techniques), the choice was made to concentrate on the use of decompositional techniques. For the purpose of selecting the most appropriate decompositional technique, four pragmatic criteria were defined.

Conjoint analysis (if restricted to the part-worth model) is the only technique which met the criteria.

To test H2, H3, H5, and H6, the evaluations of the attribute-levels of form were determined by applying the part-worth model. An O.L.S.-analysis was performed on the ratings of the full profiles (concepts). The standardized regression weights represented the part-worth utilities.

**Consumer characteristics: product-category knowledge**

An objective measure of product-category knowledge was selected to be the best indicator of actual knowledge levels. The objective measure of product-category knowledge used in this dissertation was a modification of the measure used by Brucks (1986). The categories which were included in our measure were "product attributes", "general and specific attribute evaluation", and "brand facts". In a reliability test, a standardized Cronbach's Alpha of this measure of product-category knowledge was found to be .75. All test-retest correlations (Q1 to Q3, and overall product-category knowledge score) proved to be very high.

Four experiments were conducted to investigate the processing and evaluation of early product-concepts. In Experiment 1, the level of processing of early product-concepts with little information was studied (H1). In Experiment 2, we concentrated on the effects of amount of information (H2) and product-category knowledge (H6) on the evaluations. In Experiment 3, the type of processing (holistic or analytical) of product-concepts was studied, in which the attribute "form" was presented pictorially. Also in this experiment, the evaluations made with respect to the pictorial representations of this attribute were compared with textual representations (H5). In Experiment 4, the effects of missing relevant attribute information on the evaluations of the attribute "price" were determined (H2). The reason for testing Hypothesis 2 after the other hypotheses was that, in order to test this hypothesis, a comparison had to be made between the results of Experiment 4 (in which information on the attribute "form" was missing) and those of the other experiments (in which information on the attribute "form" was present).

Since some elements were identical in all four experiments, to avoid redundancy as much as possible, we will discuss those aspects here. The identical elements were the experimental design and the respondents. (In some experiments, additional information was added to the information below).

Experimental design

A between-subjects design was chosen in all experiments. This means that the effects of the independent variables (amount of information, presentation format, and product-category knowledge) were studied by comparing the processing and evaluation of different groups of respondents. The reason for choosing a between-subjects design was that we expected that possible disturbing learning effects, which could occur in a within-subject design, would be avoided. The subjects were randomly assigned to the experimental conditions1. To check whether or not there were major differences between the two groups in terms of background variables (differences which might have explained the variance in the dependent variable), a questionnaire was included in all experiments, with questions on socio-demographic issues (sex, age, family size), and filter coffee-maker related issues (number of people in the family who drink coffee, ownership of filter coffee-makers). We will refer to this questionnaire as the control questionnaire.

Respondents

Only members of the P.E.L.2-consumer panel of the Delft University of Technology (Tan, 1989) participated in the investigation. The panel is an aselect sample of the population of the town of Delft and its immediate surroundings. The respondents were participating in an investigation on filter coffee-makers for the first time. Relevant characteristics of the

---

1 In Experiment 2, to create a maximum amount of variance in product-category knowledge, a prespecified number of respondents belonging to predefined populations were invited. In subjecting respondents to the conditions, however, population membership was not regarded.

2 Product Evaluation Laboratory.
respondents were measured by means of the so-called control questionnaire. The results from this control questionnaire are presented in Appendix 2.

All subjects received a small monetary incentive for their cooperation.

6.1 Experiment 1: The effect of "amount of information" on the processing of early product-concepts

Experiment 1 was conducted to investigate whether or not early product-concepts are processed at a more shallow level than actual products. Hypothesis 1 (H1) was tested.

Test of H1

H1: If the amount of information in product-concepts is small, a more shallow level of processing will take place than if the amount of information in the product-concepts is large.

Operationalization of the independent variable "Amount of information"

To be able to test the effects of different amounts of information in product-concepts, two types of stimuli were created: schematic pictorial stimuli and realistic pictorial stimuli (see Figure 18 of Chapter 5). In the schematic pictorial stimuli, the amount of information provided was small. In the realistic pictorial stimuli, the amount of information provided was large.

Operationalization of the dependent variable "Level of processing"

To gain an insight into the level of processing of early product stimuli, we used verbal protocol analysis. In this experiment, respondents were shown either the schematic pictorial stimuli or the realistic pictorial stimuli. The respondents were invited to verbalize what they saw in the stimuli. All verbalizations which were made were first tape-recorded, then later transcribed.

To illustrate, we present the verbalizations of respondent number 1. Respondent number 1 was shown, among other stimuli, stimulus number 4 (of the schematic pictorial stimuli).

Figure 23: Stimulus number 4

The verbalizations of Respondent number 1, when looking at stimulus number 4, are quoted below:

"Filter coffee-maker with flat sides, has a thermosflask, I mean a jug which stands on a small hotplate, above that one can see where
the coffee filter is put in, if I express it well though, a practical removable coffee jug of course, and a removable filter holder*. 

Experimental design

In the schematic pictorial condition, product-concepts with small amounts of information were shown. In the realistic pictorial condition, product-concepts with a lot of information were shown.

The respondents were randomly assigned to one of either of the conditions. In the schematic pictorial condition (n = 25), the respondents were shown a random subset of three out of the seven schematic pictorial stimuli. In the realistic pictorial condition (n = 26), the respondents were shown a random subset of three out of the seven realistic pictorial stimuli. Both sets of stimuli are illustrated in Figure 18 of Chapter 5. Each stimulus was shown at least seven times, and at most shown twelve times.

Procedure

All subjects completed two tasks in the same order. First, subjects were given a sequence of three stimuli out of the seven. In order to avoid order effects, a large number of different sequences was used. The stimuli were given to the respondents one by one. With respect to each of the three stimuli, the respondents were asked to verbalize what they saw. In the instructions, it was made clear that any comment, however peculiar, could be useful. Second, the respondents completed the control questionnaire. In general, the respondents had little difficulty in performing both tasks. Completing both tasks took between twenty and forty minutes in both conditions.

The verbalizations which were evoked by the stimuli were listed by the investigator and handed over to judges.

The verbalizations were categorized independently by two judges using the coding scheme (see Section 5.2). With respect to the coding of all verbalizations, the amount of agreement between both judges was 70.5 percent. The judges had some difficulty in differentiating between the subcategories “factual description” and “evaluation”3. The investigator, therefore, made a new list of verbalizations which only contained those verbalizations which were categorized differently by both judges. This new list contained 29.5 percent of all verbalizations.

Disagreement was smoothly resolved in a discussion between both judges, so that all verbalizations were coded. Once the judges agreed on the interpretations of the different categories, they could very easily classify the remaining 29.5 percent. It is important to note that although 29.5 percent was a substantial proportion, resolving disagreement on such a large number of verbalizations did not necessarily affect the validity of the test. The reason for this is that the judges were uninformed about the hypothesis we were testing (H1), and that therefore they could not know from which condition the verbalizations came.

Finally, from the entire list of categorized verbalizations, the investigator constructed two sublists of verbalizations: one for the verbalizations which were given in the schematic pictorial condition and another for the verbalizations which were given in the realistic pictorial condition.

Results of the control questionnaire

The results of the control questionnaire are presented in Table A of Appendix 2. This table shows that the respondents in both conditions

\[ \text{For instance, the statement that a drip coffee-maker is modern can be considered both a factual description and an evaluation.} \]
were very much alike in terms of socio-demographic background and filter coffee-maker related aspects. Only small, statistically non-significant differences were found with respect to family size and number of coffee drinkers.

Results of the test of H1

In Table 1, for both conditions, the frequencies of verbalizations are listed, corresponding to each of the codes of the coding scheme. The number of times that a verbalization is classified as belonging to a particular subcategory is summed over the seven different product-concepts of each condition. The results are given both as absolute figures and as percentages.

Table 1: Frequencies of verbalizations over subcategories evoked in the schematic pictorial condition (s.p.c.) and in the realistic pictorial condition (r.p.c.)

<table>
<thead>
<tr>
<th></th>
<th>s.p.c</th>
<th>r.p.c</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 26)</td>
<td></td>
<td>(n = 25)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>absolute</th>
<th>%</th>
<th>absolute</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>attribute oriented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-need for clarification</td>
<td>22</td>
<td>7</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>-factual description</td>
<td>62</td>
<td>19</td>
<td>86</td>
<td>26</td>
</tr>
<tr>
<td>-known/unknown</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>-evaluation</td>
<td>22</td>
<td>7</td>
<td>39</td>
<td>12</td>
</tr>
<tr>
<td>-statements on buying intention</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>-associations</td>
<td>1</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>107</td>
<td>33</td>
<td>151</td>
<td>46</td>
</tr>
<tr>
<td>total-product oriented</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-need for clarification</td>
<td>12</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>-factual description</td>
<td>49</td>
<td>15</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>-known/unknown</td>
<td>8</td>
<td>2</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>-evaluation</td>
<td>110</td>
<td>33</td>
<td>95</td>
<td>29</td>
</tr>
<tr>
<td>-statements on buying intention</td>
<td>14</td>
<td>4</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>-associations</td>
<td>24</td>
<td>7</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>217</td>
<td>65</td>
<td>168</td>
<td>51</td>
</tr>
<tr>
<td>Reactions to presentation</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Total verbalizations</td>
<td>337</td>
<td>100</td>
<td>327</td>
<td>99</td>
</tr>
</tbody>
</table>

The proportion of attribute-oriented verbalizations was .33 in the schematic pictorial condition \((107 / (107 + 217))\), and .47 in the realistic pictorial condition \((151 / (151 + 168))\). The proportions test for independent samples (see Edwards, 1969, p. 45-47) indicated that the proportions differed significantly (the corresponding z-score was equal to 3.42, which is
statistically significant at the 0.01 level). From this, we could conclude that product-concepts, in which small amounts of information are presented, are processed at a more shallow level than product-concepts in which large amounts are presented. Thus, we can accept Hypothesis 1.

Furthermore, one can see from Table 1 that factual descriptions and evaluations, both with respect to attributes and to the total product, were the subcategories to which most of the verbalizations have been ascribed. The subcategories "known/unknown", "statements on buying intentions", and "associations" were hardly used at the attribute level. At the total-product level, "known/unknown" was used more often in the realistic pictorial condition than in the schematic pictorial condition \( z = 2.10; p < .05 \). This difference is due to the fact that more respondents recognized the filter coffee-makers in the realistic pictorial condition than they did in the schematic pictorial condition.

Regarding the statements on buying intention, there was hardly any difference between both conditions. Associations were found more often in the schematic pictorial condition than in the realistic pictorial condition \( z = 3.09; p < .01 \). It is possible that in order to interpret the global product-concept information, attempts were made to refer to similar objects. Requests for clarification were present somewhat more often in the case of the schematic pictorial stimuli than in the case of the realistic pictorial stimuli \( z = .86; p = .19 \). This difference, however, was not found to be statistically significant. Finally, hardly any difference was found with respect to the number of reactions related to the way the stimuli were presented.

6.2. Experiment 2: the effects of "amount of information" and "product-category knowledge" on the evaluation of early product-concepts

In Experiment 2, both Hypothesis 2 (H2) and Hypothesis 6 (H6) were tested.

**H2:** If the amount of information presented about an attribute is small, the evaluative responses of this attribute will differ from those of the same attribute presented with a large amount of information.

**H6:** If consumers have little (a lot of) product-category knowledge, they will demonstrate different (identical) evaluative responses of the same attribute presented with different amounts of information.

**Independent variable: Amount of information**

For the conditions of Experiment 2, two sets of stimuli were created. One set of stimuli, the realistic three-dimensional stimuli, contained detailed information on the attribute "form". The other set of stimuli, the schematic three-dimensional stimuli, contained little information on the attribute "form".

In Figure 19 of Chapter 5, pictures are shown of the three representations of levels of the attribute "form", for both sets of stimuli. Here, we show one example of both sets of stimuli.

Besides the attribute "form", the following other attributes were included in the profiles: price, a dripstop, a thermos flask, a removable water reservoir, and high speed. The attribute "high speed", referring to the duration of the coffee-making process, was a new attribute for the product category of filter coffee-makers at the time of the experiment.

Only the attributes "form" and "price" had three attribute levels. For the levels of the attribute "form", three different existing forms were chosen. Price could be eighty guilders, one hundred guilders or one-
hundred-and-twenty guilders. All other attributes had only two levels, i.e. absent or present.

With the exception of the attribute "form", all attributes were presented textually. The attribute profiles, therefore, consisted of combinations of both three-dimensional and textual information.

Figure 24: Examples of a schematic three-dimensional stimulus and a realistic three-dimensional stimulus

Interaction effect: Amount of information x Product-category knowledge

To test Hypothesis 6, we determined the effect on the evaluations of the interaction between "amount of information" and "product-category knowledge". The reason for doing so was that we were not interested in the direct effect of product-category knowledge on the evaluations, but we were interested in whether or not consumers with a lot of product-category knowledge were more capable of evaluating early product-concepts than were consumers with little product-category knowledge. In this study, we conclude that consumers are capable of evaluating early product-concepts if the evaluations of schematic three-dimensional stimuli do not differ largely from those of realistic three-dimensional stimuli.

To measure the effect of the interaction between "amount of information" and "product-category knowledge" on the evaluations of the attribute "form", both "amount of information" and "product-category knowledge" were independently varied. How both variables are measured is described in Chapter 5.

Dependent variable: Evaluative responses

In order to measure the evaluative responses, a conjoint analysis was conducted. A fractional factorial design was constructed by means of "Conjoint Designer" (Bretton-Clark, 1986). This resulted in sixteen attribute-level combinations.

The evaluations of the three levels of the attribute "form" were represented by their part-worth utilities. In testing Hypotheses 2 and 6, we were interested in the differences between the evaluative responses of the same levels of the attribute "form" in the schematic three-dimensional condition and in the realistic three-dimensional condition.

Respondents

In order to obtain a maximum amount of variance in product-category knowledge, the respondents were selected by means of quota sampling.
Three quota samples were defined. These quota samples are described in Table 2.

Table 2: Definition of quota samples

<table>
<thead>
<tr>
<th></th>
<th>frequency of use</th>
<th>number of products owned</th>
<th>recency of purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample 1</strong></td>
<td>&lt; 1 a day</td>
<td>0 or 1</td>
<td>&gt; 6 months ago</td>
</tr>
<tr>
<td><strong>Sample 2</strong></td>
<td>&lt; 1 a day</td>
<td>≥ 2</td>
<td>&gt; 6 months ago</td>
</tr>
<tr>
<td><strong>Sample 3</strong></td>
<td>≥ 1 a day</td>
<td>≥ 2</td>
<td>or &lt; 6 months ago</td>
</tr>
</tbody>
</table>

The reason for applying frequency of product use, number of products owned, and recency of last product-purchase, in defining the quota samples, was the assumption that these aspects are related to the amount of knowledge people have about filter coffee-makers.

Respondents in Sample 1 were expected to have the smallest amount of product-category knowledge, and respondents in Sample 3 the largest. The respondents in Sample 2 were expected to have a moderate amount of product-category knowledge.

The respondents were selected by means of a telephone questionnaire interview. Sampling stopped at the moment when the smallest sample contained thirty respondents. Altogether, this resulted in ninety-seven respondents.

To test Hypothesis 6, the total sample of respondents (n = 97) was confronted with the measure of product-category knowledge (Q1 to Q3). Subsequently, the total sample was divided into three groups with different amounts of product-category knowledge. To this end, a “three-way split” was made of the individual scores of product-category knowledge. This resulted in the following: a large-knowledge group (n = 32), a moderate-knowledge group (n = 32), and a little-knowledge group (n = 33).

When considering the three groups together, product-category knowledge appeared to be highly correlated with age (-.40, p ≤ .001) and level of education (.40, p ≤ .001). The correlations between product-category knowledge and (i) frequency of use (.15), (ii) number of products owned (.06) and (iii) recency of purchase (.05), however, proved to be lower than we expected. Consequently, our attempt to create a maximum amount of variation in product-category knowledge on the basis of quota sampling did not yield the desired result.

**Experimental design**

A full 2x3 factorial design was created in which the variable “amount of information” had two levels: the schematic three-dimensional condition and the realistic three-dimensional condition. The variable “product-category knowledge” had three levels: little, moderate, and large product-category knowledge. In the full factorial design, the effect on the evaluations of the first-order interaction between amount of information and product-category knowledge could be tested.
Hypothesis 2 would be confirmed if different evaluations were obtained in the two conditions. Hypothesis 6 would be confirmed if statistically significant interaction effects were found which indicated that respondents with a lot of product-category knowledge gave identical evaluations of the same stimuli with different amounts of information, while respondents with little product-category knowledge gave different evaluations.

About one half of the respondents were shown the realistic three-dimensional stimuli (n = 50), the other half of the schematic three-dimensional stimuli (n = 48) \(^4\). Within each of these conditions, three groups of respondents with different levels of product-category knowledge were formed. The number of respondents in each of these subgroups was determined during the experiment. The reason for this was that, because the variable product-category knowledge was measured during the interviews, and not manipulated, it was not possible to equally divide the respondents into three knowledge groups within each of the two conditions beforehand.

**Procedure**

The procedure in Experiment 2 consisted of the following tasks: questionnaire on product-category knowledge (Q1 to Q5)\(^5\), conjoint task, control questionnaire, and magazine test. The conjoint task was a preference rating task. The respondents rated each of the eighteen attribute combinations on a 10-point scale.

**Results of the control questionnaire**

The results of the control questionnaire (see Appendix 2) are presented four times. In Table B, a comparison is made between the characteristics of the respondents in the schematic three-dimensional condition and those of the respondents in the realistic three-dimensional condition. In the Tables C, D, and E, at each of the levels of product-category knowledge, comparisons are made between the characteristics of the respondents in the different conditions of amount of information.

Table B shows that only small differences existed between the respondents in the schematic three-dimensional condition and the respondents in the realistic three-dimensional condition. Similar results are found in the Tables C, D, and E. In the little-knowledge group, the proportion of older respondents within both conditions was relatively large. This is not remarkable since the correlation between product-category knowledge and age was very strong (-.40).

In the moderate-knowledge group, within the realistic three-dimensional condition, there were (relatively) many male respondents, coming from relatively small households.

In the large-knowledge group, the most important difference was found with respect to sex. In the schematic pictorial condition, the number of male respondents was somewhat larger than that in the realistic pictorial condition. None of the differences, however, were found to be statistically significant at the .05 level.

**Results of the tests of H2 and H6**

To test Hypotheses 2 and 6, a multivariate analysis of variance (MANOVA) was conducted on a full factorial design for each of the three levels of the attribute "form". The factors were "amount of information" and "product-category knowledge". In addition, the first-order interaction between these factors was included in the analysis.

\(^4\) One respondent was excluded from the analysis because they had participated in an experiment on drip coffee-makers before.

\(^5\) Only Q1 to Q3 were included in the final measure of product-category knowledge, as already discussed in Chapter 5.
Figure 25: Graphical representation of mean part-worth utilities as a function of amount of information and product-category knowledge.

P = Philips
B = Braun
DE = Douwe Egberts
Since the part-worth utilities of the third attribute-level of form were derived directly from those of the first and the second, the dependent variables were related. As a consequence, if an effect was found with respect to one attribute level, this may have been the result of the sum of two smaller effects (in the same direction) on the other attribute levels. Therefore, to gain a better insight into the origin of the effects, first we graphically present the mean part-worth utilities given under condition.

In Figure 24 (see p. 65), one can see that the largest differences between the conditions, in terms of mean part-worth utilities, were found in the little-knowledge groups. In those groups, contrasting evaluations of the attribute levels "Philips" and "D.E." can be witnessed. In the realistic three-dimensional condition, the D.E. was preferred to both other attribute levels, whereas in the schematic pictorial condition, the D.E. was the least preferred attribute level. A possible explanation is that the respondents who had little product-category knowledge recognized the D.E. only in the realistic pictorial condition. After this recognition, the older respondents who were over-represented in this group, evaluated this attribute level positively since the D.E. brand is known for its durability. The (younger) respondents in the moderate- and large-knowledge groups preferred Braun (which has an integrated form), and strongly disliked the D.E. (which has an unintegrated form). An explanation for this reversal is that those respondents found the form of the product more important than its durability.

In the Table below, the results of the tests of H2 and H6 are presented.

Table 3: Part-worth utilities as a function of amount of information (A.o.I.), product-category knowledge (P.C.K.), and the interaction between amount of information and product-category knowledge (A.o.I. x P.C.K.) (N=96)

<table>
<thead>
<tr>
<th>Attribute level</th>
<th>Factor</th>
<th>SS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philips</td>
<td>A.o.I.</td>
<td>.09</td>
<td>1</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>P.C.K.</td>
<td>.06</td>
<td>2</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>A.o.I. x P.C.K.</td>
<td>1.00</td>
<td>2</td>
<td>4.86*</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>.25</td>
<td>1</td>
<td>2.41</td>
</tr>
<tr>
<td></td>
<td>Within cells</td>
<td>9.31</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Braun</td>
<td>A.o.I.</td>
<td>.01</td>
<td>1</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>P.C.K.</td>
<td>1.50</td>
<td>2</td>
<td>4.26*</td>
</tr>
<tr>
<td></td>
<td>A.o.I. x P.C.K.</td>
<td>.43</td>
<td>2</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>5.67</td>
<td>1</td>
<td>32.27**</td>
</tr>
<tr>
<td></td>
<td>Within cells</td>
<td>15.82</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>D.E.</td>
<td>A.o.I.</td>
<td>.18</td>
<td>1</td>
<td>.40</td>
</tr>
<tr>
<td></td>
<td>P.C.K.</td>
<td>2.13</td>
<td>2</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td>A.o.I. x P.C.K.</td>
<td>2.53</td>
<td>2</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>8.30</td>
<td>1</td>
<td>18.82**</td>
</tr>
<tr>
<td></td>
<td>Within cells</td>
<td>39.67</td>
<td>90</td>
<td></td>
</tr>
</tbody>
</table>

*: p < .05
**: p < .01

From Table 3, we can conclude that the variable "amount of information" does not have a significant effect on the evaluations of its own. Consequently, Hypothesis 2 should be rejected. In combination with
"product-category knowledge", however, "amount of information" had a statistically significant effect on the evaluations of the attribute-level "Philips". This effect is presented more clearly in Table 4. In this table, we present the average evaluations (and standard deviation) of the attribute level "Philips", for each of the (six) cells.

Table 4: Average evaluations of the Philips form as a function of amount of information and product-category knowledge (N=96)

<table>
<thead>
<tr>
<th></th>
<th>Small</th>
<th></th>
<th>Large</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Little (n=11)</td>
<td>Moderate (n=18)</td>
<td>Large (n=17)</td>
<td></td>
</tr>
<tr>
<td>Amount of information</td>
<td>mean (SD)</td>
<td>mean (SD)</td>
<td>mean (SD)</td>
<td>mean (SD)</td>
</tr>
<tr>
<td></td>
<td>.17 (.33)</td>
<td>.11 (.40)</td>
<td>-.03 (.30)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.13 (.28)</td>
<td>.01 (.29)</td>
<td>.19 (.32)</td>
</tr>
</tbody>
</table>
6.3 Experiment 3: the effects of presentation format on the processing of early product-concepts and on the evaluative responses

In Experiment 3, both Hypothesis 4 (H4) and Hypothesis 5 (H5) were tested. Both hypotheses could be tested in one experiment because the stimuli (presentation format), the designs, the respondents, and the procedures needed for these tests were identical. The only difference between them was the dependent variable (processing of the information in the concepts or the evaluative response).

H4: If product-concepts consist of both pictorial and textual information, this will not lead to a more holistic processing than if product-concepts only consist of textual information.

H5: If attribute information is presented textually, this will not yield evaluative responses which are different from those when presenting the same attribute information pictorially.

Independent variable: Presentation format

The independent variable in this investigation was the presentation format of the product-concepts, i.e., textual (all attribute levels, including form, expressed textually) or pictorial (the attribute "form" presented pictorially and the other attributes textually).

Both types of presentation format comprised the same attribute combinations. The attributes were form (3 levels), price (3 levels), dripstop (2 levels), and thermostool (2 levels). For the levels of the attribute "form", three different existing forms were chosen. Price could be eighty guilders or one-hundred-and-twenty guilders. How the three levels of the attribute "form" were transformed into the textual format has been described in Chapter 5.

Dependent variable (H4): Type of processing

In testing H4, "type of processing" (holistic or analytical) was determined in the same two ways as in the study by Holbrook & Moore (1981). A large number of statistically significant attribute interactions would indicate an holistic type of processing, a small number would indicate analytical processing. Furthermore, a low median value of Hays' omega-squared would indicate an holistic type of processing, whereas a high median value would indicate an analytical type of processing.

To be able to assess attribute interactions, a full factorial design was applied in which 36 product-concepts were presented.

The evaluations of the attribute-levels of form were determined by performing a MANOVA on the ratings of the concepts. For every individual subject, a MANOVA was performed. In this estimation procedure, all first-order interactions were included. The number of statistically significant interaction effects were counted and compared between both sets of stimuli. Since the number of attributes was four, the number of significant first-order interactions could be at most six.

Dependent variable (H5): Evaluative responses

The evaluative responses were determined by estimating the part-worth utilities using O.L.S.-regression. This estimation procedure gave the part-worth utilities of each of the attribute levels (for every individual respondent). For the purpose of testing H5, only the part-worth utilities of the attribute "form" were taken into consideration.

6 The coffee can in the photographs could be seen as an ordinary can, but also as a thermostool. See figure 21.
Experimental design

In Experiment 3, there were two conditions. In the textual condition, the respondents (n = 20) completed a concept test with textual stimuli. In the pictorial condition, the respondents (n = 21) completed a concept test with concepts consisting of combinations of both pictorial and textual information. The stimuli used in both conditions are shown in Figure 21 in Section 5.1. Here, we present one example from both conditions.

![Figure 26: Examples of stimuli from the textual condition and the pictorial condition](image)

Procedure

Depending on the condition, the subjects started with a concept test with either textual or pictorial concepts. All thirty-six full concepts were rated on a ten-point scale. The words "very attractive" and "very unattractive" were put at both extremes. After the concept test, a control questionnaire was completed.

Results of the control questionnaire

The results of the control questionnaire are presented in Table F of Appendix 2. As can be seen in this table, the respondents of both conditions were very much alike, except for age. In the pictorial condition, there were fewer respondents in the youngest group and more in the oldest group than in the textual condition. These differences are statistically significant (z = 2.12 and z = 2.27, respectively). Since in the literature, to our knowledge, no effects of age on type of processing are reported, we assumed that the differences in age between both respondent groups would not effect the test of Hypothesis 4. With respect to the test of Hypothesis 5, however, we did expect an effect of age on the evaluations of the attribute "form". This expectation was based on the study by De Bont et al. (1992), which showed that age is an important factor in explaining the acceptance of different forms of espresso machines. As a consequence, we controlled for the effects of age in testing Hypothesis 5 by using age as a covariate.

Results of the test of H4

In Table 5, the numbers of statistically significant first-order attribute interactions are given for both conditions. Since one individual could produce at most 6 statistically significant first-order interactions, theoretically these numbers lie between 0 and 180 for the textual condition (n=20) and between 0 and 186 (n=21) for the pictorial condition.
Table 5: Numbers of statistically significant attribute interactions as a function of presentation format (N=41)

<table>
<thead>
<tr>
<th>Presentation format</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>textual (n = 20)</td>
<td>pictorial (n = 21)</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
</tr>
</tbody>
</table>

In Table 5, one can see that the numbers of statistically significant first-order attribute interactions in both conditions are almost the same. We can report that, in both conditions, nine respondents produced least one attribute interaction. The only difference between the conditions was that in the pictorial condition there were two respondents with two first-order attribute interactions, whereas in the textual condition there was only one.

It is worth noting that seven out of the eleven first-order interactions in the pictorial condition included the attribute “form”, whereas in the textual condition this number was three out of ten. This difference, although not statistically significant, indicates that there might possibly have been a somewhat stronger inclination towards holistic processing when pictures were included.

The second measure of holistic/analytical processing was Hays’ omega-squared. This measure is calculated by dividing the strongest main effect by the strongest interaction effect for each individual. To aggregate over individuals, the median ratio is taken. In this experiment, the median values were 7.42 in the pictorial condition and 11.41 in the textual condition. From this, we concluded that the attribute information in the pictorial condition was processed somewhat more holistically than the attribute information in the textual condition. However, when comparing both median values with the median value obtained in the pictorial condition of the study by Holbrook & Moore (1981), i.e. 1.06, the median values obtained in our experiment reflected an analytical type of processing much more.

When considering the results of both measures of type of processing, we can conclude that Hypothesis 4 can be accepted: that presenting the attribute “form” pictorially will not lead to a more holistic processing than presenting the attribute “form” textually.

Results of the test of H5
To test Hypothesis 5, evaluations of the three levels of the attribute “form”, expressed in part-worths, were compared between both conditions. We controlled for the effects of age by performing an analysis of covariance (ANCOVA) in which age was submitted as a covariate. Similar to the tests of Hypotheses 2 and 6, the dependent variables were highly correlated since the third part-worth utility was derived from the first and the second. This means that if an effect was found with respect to one attribute level, this may have been caused by other attribute levels. For this reason, and to facilitate the interpretation of the ANCOVA, first we graphically present the mean part-worth utilities of both conditions.
P = Philips
B = Braun
S = Severin

Figure 27: Graphical representation of the mean part-worth utilities in the textual and the pictorial conditions

In Figure 27, one can see that the mean part-worth utilities were quite different in both conditions. In particular, the attribute-level "Severin" was evaluated quite differently. Whether or not these differences are statistically significant can be seen in Table 6.

Table 6: Mean part-worth utilities of the levels of the attribute form as a function of presentation format and age (N=41)

<table>
<thead>
<tr>
<th>Attribute level</th>
<th>Presentation format</th>
<th>mean</th>
<th>SD</th>
<th>mean</th>
<th>SD</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phillips</td>
<td>textual (n=20)</td>
<td>-.25</td>
<td>.26</td>
<td>.01</td>
<td>.41</td>
<td>2.38</td>
<td>3.03</td>
</tr>
<tr>
<td></td>
<td>pictorial (n=21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.33*</td>
</tr>
<tr>
<td>Covariate Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Braun</td>
<td>textual (n=20)</td>
<td>-.12</td>
<td>.32</td>
<td>.13</td>
<td>.21</td>
<td>2.38</td>
<td>7.38*</td>
</tr>
<tr>
<td></td>
<td>pictorial (n=21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.90</td>
</tr>
<tr>
<td>Covariate Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severin</td>
<td>textual (n=20)</td>
<td>.37</td>
<td>.54</td>
<td>-.12</td>
<td>.50</td>
<td>2.38</td>
<td>6.69*</td>
</tr>
<tr>
<td></td>
<td>pictorial (n=21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.62</td>
</tr>
<tr>
<td>Covariate Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: p < .05

We can conclude from Table 6 that the differences in evaluations of the three levels of the attribute "form" are statistically significant. Consequently, we should reject Hypothesis 5 since it appears that presenting attribute information textually does yield different evaluative responses than
presenting the same attribute information pictorially.

The covariate "age" had a statistically significant influence on the evaluation of the attribute-level "Philips". To facilitate the interpretation of the effect of age on the evaluative responses, in Table 7 we show the correlations between age and each of the attribute levels in both conditions.

Table 7: Correlations between age and the evaluative responses of the levels of form in the textual and pictorial conditions

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Textual</th>
<th>Pictorial</th>
</tr>
</thead>
<tbody>
<tr>
<td>level of form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philips</td>
<td>.37</td>
<td>.20</td>
</tr>
<tr>
<td>Braun</td>
<td>.07</td>
<td>-.01</td>
</tr>
<tr>
<td>Severin</td>
<td>-.22</td>
<td>-.16</td>
</tr>
</tbody>
</table>

The evaluations of the levels of the attribute "form" correlate more strongly with age in the pictorial condition than in the textual condition. This indicates that, in the textual descriptions of the attribute-level "Philips", certain words triggered positive connotations in the case of older respondents, whereas the textual descriptions of the attribute-level "Severin" triggered negative connotations.7

Furthermore, it is noteworthy that presentation format had a statistically significant influence on the importance of the attribute "price". Price was found to be more important in the pictorial condition. In the pictorial condition a low price was clearly preferred to a high price. In the textual condition, however, only eleven respondents preferred the cheapest alternative, whereas eight respondents preferred the moderately priced alternative and two the most expensive one. In the pictorial condition, these figures were fifteen, three and three respectively. Related to Experiment 4, it is possible that because the information on the relevant attribute "form" was somewhat ambiguous in the textual condition, many respondents used price as an indicator of the quality of that attribute.

6.4 Experiment 4: The effects of missing information about a relevant attribute on the evaluation of the attribute "price"

In Experiment 4, Hypothesis 3 was tested.

H3: If information on a relevant attribute is missing or ambiguous, the levels of the attribute "price" will be used as an indication of the level of the missing or ambiguous attribute.

Operationalization of the independent variable "missing" or "ambiguous" information on a relevant attribute

In testing Hypothesis 3, a distinction was made between missing and ambiguous information. To test the effect of missing information, a comparison was made with respect to evaluative response of the attribute "price" between the evaluations of respondents who did not receive any information on the attribute "form" and those of respondents who received as much information as possible on this attribute. How the first set of stimuli was created is described in Chapter 5. In Figure 20 of that chapter, an example is shown. The second set consisted of the realistic three-
dimensional stimuli which were used in Experiment 2. These stimuli presented information on the attribute "form" with a high degree of realism.

Ambiguous information was presented at an intermediate degree of realism when some but not all information was presented on an attribute. In Experiments 2 and 3, different sets of stimuli were shown at the intermediate degree of realism: the schematic three-dimensional stimuli (Experiment 2), the textual stimuli (Experiment 3), and the pictorial stimuli (Experiment 3). To test the effect of ambiguous information about the attribute "form" on the evaluations of the attribute "price", three comparisons were made. The first comparison was between stimuli presented at a low and at a high degree of realism. The second comparison was between the textual and the pictorial stimuli, the third between the schematic three-dimensional stimuli and the realistic three-dimensional stimuli. The reason for making the last two comparisons was that in both comparisons a difference exists with respect to the amount of information on the attribute, and that the sets of stimuli used in each comparison belonged to the same experiment. Because of the last reason, differences in the evaluations of the attribute "price" can only be attributed to the differences in the degree of realism (ambiguity) of the attribute "form".

**Operationalization of the dependent variable "Price as an indicator of the level of a relevant absent attribute"

When applying classical utility theory, one expects that (utility-maximizing) consumers will prefer a low-priced alternative to a high-priced alternative, given that the alternatives are identical. In some instances, however, when consumers cannot judge the attractiveness of different alternatives very well, a more expensive alternative may be preferred in order to reduce the risk involved in purchase (for instance, see the study by Rao and Monroe, 1988).

In this experiment, the attribute "price" had three levels: high, moderate and low. The attribute "price" could be considered to be an indicator of the level of an absent attribute, in which case the moderate price level or the high price level should be preferred to the low price level. Whether or not this is the case was assessed by comparing the part-worth utilities of the levels of the attribute "price".

We expected that respondents who were to be shown the stimuli which do not include information on the attribute "form" (or ambiguous information only) would prefer the moderate and the high price levels more often than respondents who were to be shown stimuli which did include (all or at least unambiguous) information on the attribute "form".

**Experimental design**

To test the effect of missing information on the attribute "form", two conditions were created: the "no information" condition and the "all information" condition. In the "no information" condition, the respondents evaluated the product-concepts without information on the attribute "form" (no information). In the "all information" condition, the respondents evaluated product-concepts with a maximum amount of information on the attribute "form" (all information).

To test the effects of ambiguous information about the attribute "form", the conditions which were created were: the textual condition, the pictorial condition, the schematic three-dimensional condition, and the realistic three-dimensional condition. The first three conditions provided some
information on the attribute “form”. The last condition was equivalent to the “all information” condition, described above, in which all information was present. The stimuli for the first and the second condition were constructed in Experiment 3, those for the third and the fourth in Experiment 2.

Procedure

Since the procedure has already been described for all conditions except the “no information” condition, here we restrict ourselves to a description for that condition only. The procedure for the respondents in the “no information” condition consisted of the following tasks: conjoint task, control questionnaire, and magazine test.

Using “Conjoint Designer” (Bretton-Clark, 1986), a fractional factorial design was constructed consisting of sixteen attribute-level combinations. The conjoint task was to rank the sixteen attribute profiles according to preference. In the magazine test, respondents were invited to choose one filter coffee-maker from a number of magazines on household durables which included filter coffee-makers. This task was included for validation purposes and will be given more attention in the second part of this thesis.

Results of the control questionnaire

The results of the control questionnaire are presented in Table G of Appendix 2. The characteristics of the respondents in the conditions with information on the attribute “form” are already shown in Appendix 2. In Table G, it can be seen that there are more female than male respondents in the “no information” condition, as in both conditions of Experiment 3. In the realistic three-dimensional condition of Experiment 2, by contrast, slightly more male than female respondents participated (the differences are not statistically significant).

Results of the test of H3

In Table 8, the numbers of respondents who preferred a particular level of price are presented for all conditions.

Table 8: Numbers of respondents preferring a particular level of price as a function of the amount of information on the attribute “form”

<table>
<thead>
<tr>
<th>Amount of information about form</th>
<th>No information</th>
<th>Some (ambiguous) information</th>
<th>All information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>textual (n=20)</td>
<td>pictorial (n=21)</td>
<td>3D8 (n=46)</td>
</tr>
<tr>
<td>low</td>
<td>18 (44%)</td>
<td>15 (71%)</td>
<td>14 (30%)</td>
</tr>
<tr>
<td>moderate</td>
<td>8 (20%)</td>
<td>3 (14%)</td>
<td>18 (40%)</td>
</tr>
<tr>
<td>high</td>
<td>15 (37%)</td>
<td>3 (14%)</td>
<td>14 (30%)</td>
</tr>
</tbody>
</table>

The proportions of respondents who preferred a moderate or a high level of price were 56% (23 / 41) in the “no information” condition and 48% (24 / 50) in the “all information” condition. The small difference between them is not statistically significant. Therefore, we should reject one part of Hypothesis 3 since it appears that if information on a relevant attribute is
missing, the levels of the attribute "price" are not used as an indicator of the level of the missing attribute. To some extent the rejection of H3 can be explained by the fact that in the "no information" condition (5) the number of attributes was smaller than in the "all information" condition (6). In addition, "all information" in the "no information" condition was presented in the same presentation format whereas in the "all information" condition information was presented both three-dimensionally and textually. As a result, the respondents in the "no information" condition, in spite of lacking information on the attribute "form," may have felt more secure than the respondents in the "all information" condition.

When taking a closer look at the differences between the sets of stimuli in Experiments 2 and 3, we find stronger effects. The proportions of respondents who preferred a moderate or a high level of price were 45% in the textual and 29% in the pictorial presentation format. The corresponding z-value is 1.06. This is also not statistically significant. The proportion of respondents who preferred a moderate or a high level of price was 70% in the schematic three-dimensional condition, and only 48% in the realistic three-dimensional condition. The corresponding z-value is 2.19. This is statistically significant at the .05 level. From this, we can conclude that the second part of Hypothesis 3 can be accepted, i.e. that ambiguous information on the attribute "form" does affect the evaluation of the attribute "price".

6.5 Summary and conclusions

Four experiments were conducted to investigate the processing and evaluation of early product-concepts. First, we give a summary of these experiments, and then draw some conclusions with respect to the first research question, i.e. in what respects are early product-concepts evaluated differently from actual products?

Experiment 1

Experiment 1 was conducted to investigate whether or not early product-concepts are processed at a more shallow level than actual products, and thus Hypothesis 1 (H1) was tested. In this experiment, respondents were shown either photographs of early product-concepts (schematic pictorial condition) or photographs of actual products (realistic pictorial condition). The respondents were invited to verbalize what they saw in the photographs. In the schematic pictorial condition (n = 25), the respondents were shown a subset of three out of the seven schematic pictorial stimuli. In the realistic pictorial condition (n = 26), the respondents were shown a subset of three out of the seven realistic pictorial stimuli.

The verbalizations which were evoked by the stimuli were listed by the investigator and handed over to judges. The judges independently categorized all verbalizations using the specified coding scheme. The proportion of attribute-oriented verbalizations was found to be .33 in the schematic pictorial condition, and .47 in the realistic pictorial condition. The proportions test for independent samples indicated that the corresponding z-score is statistically significant at the .01 level. From this, we concluded that the information contained in early product-concepts is processed at a more shallow level than information contained in actual products. Thus, Hypothesis 1 was accepted.
Experiment 2

In Experiment 2, both Hypothesis 2 (H2) and Hypothesis 6 (H6) were tested. Hypothesis 2 concerned the effect of the amount of information about an attribute on its evaluative responses. Hypothesis 6 concerned the effect of product-category knowledge on the evaluation of an attribute which is presented with a small amount of information.

For the conditions of Experiment 2, two sets of stimuli were created. One set of stimuli, the realistic three-dimensional stimuli, contained detailed information on the attribute “form”. The other set of stimuli, the schematic three-dimensional stimuli, contained little information on the attribute “form”. To test Hypothesis 6, the total number of respondents (n = 97) was divided into three groups with different amounts of product-category knowledge, namely the large-knowledge group (n = 32), the moderate-knowledge group (n = 32), and the little-knowledge group (n = 33).

A full 2x3 factorial design was created, in which the variable “amount of information” had two levels: the schematic three-dimensional condition and the realistic three-dimensional condition. The variable “product-category knowledge” had three levels: little, moderate, and large product-category knowledge. In the full factorial design, the effect on the evaluations of the first-order interaction between “amount of information” and “product-category knowledge” was tested.

“Amount of information” did not appear to have an effect on the evaluations. Therefore, we concluded that Hypothesis 2 should be rejected. The factors “amount of Information” and “product-category knowledge” appeared to have a combined effect on the evaluation of the level “Philips”. The difference in evaluations of different amounts of information presented for the attribute-level “Philips” appeared to be very large for the respondents with little product-category knowledge (.30). The difference was smaller for the respondents with moderate knowledge (.10) and those with a large knowledge (.22). Therefore, we came to the conclusion that Hypothesis 6 could be accepted. If Consumers have little product-category knowledge, they demonstrate different evaluative responses of the same attribute presented with different amounts of information. Consumers with moderate and large product-category knowledges evaluate attribute levels which are presented with a small amount of information identically to the same attribute levels presented with a large amount of information.

Experiment 3

In Experiment 3, both Hypothesis 4 (H4) and Hypotheses 5 (H5) were tested. H4 stated that attribute profiles in which the attribute “form” is presented pictorially, and all attributes presented textually, will not lead to a more holistic processing than profiles in which all attributes (including form) are presented textually. H5 stated that presenting attribute information textually will not yield different evaluations from those when presenting the same attribute information pictorially.

The independent variable in this investigation was the presentation format of the product-concepts, such that all attribute levels, including form, were expressed textually, or the attribute “form” was presented pictorially and the other attributes textually.

In testing H4, type of processing (holistic or analytical) was determined in the same ways as it was in the study by Holbrook & Moore (1981), i.e. via the number of statistically significant attribute interactions, and Hays' omega-squared. In testing H5, the evaluations of the levels of the attribute
form" were determined by estimating their part-worth utilities using O.L.S.-
regression.

The results from a group of respondents (n = 20) completing a textual
concept test (textual condition), both in terms of type of processing (H4) and
in terms of evaluations (H5), were compared with the results of a different
group (n = 21) completing the concept test using the pictorial presentations
(pictorial condition).

The results for type of processing indicated that there were only small,
non-significant differences between both conditions. From this, we
concluded that Hypothesis 4 could be accepted: that presenting the attribute
"form" pictorially does not lead to a more holistic processing than presenting
the attribute "form" textually.

The results also indicated that the evaluations of the three levels of the
attribute "form" differ to a large extent between both conditions.
Consequently, we rejected Hypothesis 5 since it was found that presenting
attribute information textually does yield different evaluations from those
when presenting the same attribute information pictorially.

Experiment 4

In Experiment 4, Hypothesis 3 was tested. Hypothesis 3 concerned the
effect of missing or ambiguous information about a relevant attribute on the
evaluation of the attribute "price". In testing Hypothesis 3, a distinction was
made between missing and ambiguous information. To test the effect of
missing information, a comparison was made between the evaluations of
respondents who did not receive any information on the attribute "form" and
respondents who received as much information as possible on this attribute.
To test the effect of ambiguous information about the attribute form on
evaluations of the attribute "price", two comparisons were made. The
first comparison was between the textual and the pictorial stimuli, the
second between the schematic three-dimensional stimuli and the
realistic three-dimensional stimuli.

The proportions of respondents who preferred a moderate or a high level
of price were 56% (23 / 41) in the "no information" condition and 48% (24 / 50)
in the "all information" condition. The small difference between them is
not statistically significant. Therefore, we rejected one part of Hypothesis 3
since if information on a relevant attribute is missing, the levels of the
attribute "price" are not used as an indication of the level of the missing
attribute.

With respect to the effect of ambiguous information, different results were
found. The proportions of respondents preferring a moderate or a high level
of price were 45% in the textual and 29% in the pictorial presentation format,
the corresponding z-value being 1.06. This is not statistically significant.
The other proportions were 70% in the schematic three-dimensional
condition and only 48% in the realistic three-dimensional condition, the
corresponding z-value being 2.19. This z-value is statistically significant at
the .05 level. From this, we concluded that the second part of Hypothesis 3
could be accepted. Ambiguous information on the attribute "form" does
affect the evaluation of the attribute "price".

Conclusions

On the basis of the findings of the four experiments, an answer can be
given for the first research question. With respect to the evaluation of early
product-concepts, the following conclusions can be drawn:
1. Early product-concepts are processed at a more shallow level than actual products. With respect to this finding the facts that (a) the consumers were invited to take a look at the stimuli instead of evaluating them, and that (b) less attributes could be seen in the schematic pictorial stimuli than in the realistic stimuli, may have influenced the results. Nevertheless, the point to be made is that schematic stimuli evoked a shallow level of processing. If respondents use this shallow level of processing when evaluating schematic product-concepts, and evaluative responses are made without elaborating upon and/or comprehending the product-concept as such, then the reliability and the validity of these responses must be questioned.

2. "Amount of information" does not affect the evaluative responses of an attribute in the case when respondents have moderate or large product-category knowledges, but does in the case when respondents have little product-category knowledge. Evidently, respondents with a moderate or a large product-category knowledge are better equipped to evaluate early product-concepts. As a consequence, one can expect that respondents with moderate or large product-category knowledges generate more reliable and valid data than respondents with little product-category knowledge.

3. If information on a relevant attribute is presented ambiguously, this affects the evaluations of the attribute "price". In those instances, respondents are more inclined to accept higher prices, and in a large number of cases even prefer high prices to low prices. Having found this, we have to be very careful in interpreting the part-worth utilities and importances of the attribute "price" when information on a relevant attribute is presented ambiguously.

4. Pictorial information does not seem to lead to an holistic type of processing. The combination of pictorial and textual information in product-concepts guarantees that the information is processed analytically. Since, in a shop too, products like filter coffee-makers are presented both textually (features) and pictorially (package), we expect that this type of processing does not differ from that in a shop. In addition, the finding that the information is processed analytically supports the linear additive model in which only main effects are determined.

5. Presenting attribute information pictorially leads to different evaluative responses than presenting attribute information textually. Evidently, the carefully created textual descriptions yield connotations that are different from their matching pictorial counterparts. From this, we can conclude that one should be restrictive in using textual descriptions of the form of a product in concept testing.

These conclusions indicate that the nature of early product-concepts (small amount of information, different presentation formats) has consequences for their processing and evaluative responses. A lower level of processing, the use of the attribute "price" as an indicator of missing attribute information, and differences in evaluations when using different presentation formats, give rise to questions about the reliability and the validity of evaluations which are based largely on early product-concepts.

These matters are investigated in the second part of this thesis. In assessing the validity and the reliability of evaluations which are based on early product-concepts, the influences of the amount of available
information and the presentation format will have to be determined. In addition, product-category knowledge needs to be included as an important mediating variable.
7. The Reliability and the Validity of Conjoint-Analysis Results: A Review of Literature

In Chapter 6, we found that early product-concepts are processed at a more shallow level than actual products. The amount of available information affected evaluations by respondents who had little product-category knowledge. The amount of information further affected the evaluations of the attribute “price”. We also found that presentation format affected the evaluations. These findings may have consequences for the reliability and the validity of conjoint-analysis results. This brings us to the second research question: how reliable and how valid are evaluations based on early product-concepts? Since we determine evaluations by means of conjoint analysis (see Section 5.3), the second research question can be reformulated as follows: how reliable and how valid are conjoint-analysis results (part-worth utilities, importances) based on early product-concepts?

Before answering the second research question, in this chapter we will first discuss what is meant by reliability and validity, in general and in the context of conjoint analysis in particular. Next, we will describe studies which have already been conducted to assess the reliability and the validity of conjoint-analysis results, and which have been reported in scientific journals. The outcomes will be discussed in detail.

Much attention will be given to factors influencing the reliability and the validity of conjoint-analysis results. Two kinds of factors will be discussed. Firstly, because of the findings on the processing of early product-concepts which we reported in Chapter 6, special attention will be given to the effects of the nature of new-product concepts (amount of information and presentation format), and to the intervening effects of consumer characteristics (product-category knowledge). Secondly, attention will be given to the factors which are part of the structure of a conjoint-analysis study (Green & Srinivasan, 1978; 1990). These factors are the model of preference, the data-collection method, stimulus-set construction, stimulus presentation, the measurement scale for the dependent variable, and the estimation method. These factors connote important choices which have to be made when applying conjoint analysis, and which have received much attention in validation studies. The established findings from studies on the influences of these factors should be respected when conducting a conjoint-analysis study. The outcomes of several such studies of the reliability and the validity of conjoint-analysis results are presented in the Appendices 4 to 6.

The established findings from studies on the influences of amount of information, of presentation format, and of product-category knowledge on the reliability and the validity of conjoint-analysis results, will be used to draw conclusions with respect to the second research question. However, methodological shortcomings, and omissions with respect to the factors in which we are interested, give rise to a need to empirically (re)investigate the reliability and the validity of conjoint-analysis results, and to include the omitted and potentially important factors. In this chapter, hypotheses for further research are formulated.
It is important to note that since we concentrate on the validity of the data at the individual level, no attention is given to studies about the validity of different techniques which can be used to aggregate the individual data. On this subject, we refer the reader to the study by Stokmans (1991).

7.1 Reliability and validity in general

Reliability and validity are crucial aspects of empirical research. Drenth (1980) states in broad terms that reliability concerns the amount of (non-systematic) error in measuring a variable, or in measuring the relationship between two (or more) variables. Validity concerns the extent to which the observed variable represents the concept in which we are interested.

Reliability

When discussing reliability, True-score theory is dominant (Cronbach, 1984). In True-score theory, an observed score in a test is the sum of the true score and the error of measurement. In terms of True-score theory, for a test to be reliable, the amount of unwanted variation (error of measurement) has to be substantially smaller than variation in the true score.

True-score theory asserts that, in observing relationships between different variables, it has to be assured that the observed covariation is stable over time, that the covariation is not the result of the measurement procedure itself, and that measurement is accurate.

A second theory on reliability is the theory of generalizability (Cronbach et al., 1972). In essence, this theory focuses on the factors which determine variance in the observed scores. These factors cannot only be the independent variable(s) in which one is interested, but also elements of generalization, such as time and situation. The focus of the theory of generalizability is often also covered by theories about the validity of the results. To illustrate, in the framework proposed by Cook and Campbell (1979), generalization is covered by external validity. For this reason, when we refer to reliability we stick to True-score theory.

Reliability is mainly determined by examining the degree of agreement between two efforts to measure the same concepts through maximally similar methods. Another method is data splitting, in which observations are divided into two (or more) groups. Reliable results are observed if highly similar relationships are found in each of the different subsamples.

Validity

Because of the heterogeneity of orientations, goals and subsequent methods applied in empirical research, different definitions of validity have been developed (Schwager, 1988). For example, the distinction between internal and external validity (e.g. Cook & Campbell, 1979; Christensen, 1988) is often made when discussing experiments. Internal validity refers to the extent to which one can accurately state that an independent variable produced an observed effect in another variable. One aspect of internal validity, namely accurately observing (or measuring) a relationship, overlaps with reliability. External validity is the extent to which the results of an experiment (causal relationship) can be applied to and across different people, settings and times (Cook & Campbell, 1979). External validity is a major issue in laboratory research, due to the fact that behavior in experimental settings, in most instances, largely differs from behavior in natural settings.

\[\text{Example 1} \quad \text{developed by Spearman (1910), and is often referred to as the classical theory of reliability.}\]

\[\text{Example 2} \quad \text{Examples of such laboratory experiments can be found in introductory books on social psychology or on game theory.}\]
In the context of consumer research, Zaltman et al. (1973) define different types of concept validity, such as:

- observational validity (the degree to which a concept is reducible to observations)
- content validity (the degree to which an operationalization represents the concept about which generalizations are made)
- construct validity\(^3\) (the extent to which an operationalization measures what it purports to measure)
- criterion-related validity\(^4\) (the degree to which the concept under consideration enables one to predict the values of some other concept which constitutes the criterion)
- systematic validity (the degree to which a concept enables the integration of previously unconnected concepts and/or the generation of a new conceptual system)
- semantic validity (the degree to which a concept has a uniform semantic usage)
- control validity (the degree to which a concept is manipulable and capable of influencing other variables).

With the exceptions of systematic validity and semantic validity, which refer to theoretical reasoning instead of to empirical content, Zaltman et al. (1973) show how each of the above-mentioned types of validity can be related to a particular research goal. If the goal of a study is to describe a phenomenon (except reliability), observational validity, content validity, and construct validity are the most important. If the goal is explanation and/or prediction, criterion-related validity is most important; if the goal is control, then control validity. It is important to note that acceptable degrees of reliability and validity, for lower level research goals (e.g. description), are a necessary condition for accomplishing high-level research goals (explanation, prediction or control).

7.2 Reliability and validity of conjoint-analysis results

In studies which use conjoint analysis, respondents are invited to evaluate stimuli consisting of sets of attribute levels (or attribute-level combinations). What are really measured are the relationships between attribute levels on the one hand, and evaluations on the other. These relationships serve as input to a statistical estimation procedure, yielding (at the individual level) the part-worth utilities corresponding to each of the attribute levels. The part-worth utilities, and the importances which can be deduced from the part-worth utilities, constitute the individual preference structures. In this thesis, we use conjoint-analysis results to refer to these individual preference structures.

The most important goal of a conjoint-analysis study is to determine individual preference structures. Based on a large number of preference structures, one or more optimal attribute-combinations can be determined. Subsequently, the decision can be taken to develop a particular concept (combination of attributes). Implicitly, these decisions are based on the assumption that if a product is introduced into the market which consists of the optimal combination of attributes (for a particular group of respondents), it will be preferred to all other alternatives on the market.

In terms of research goals, it can be stated that conjoint analysis aims to describe (by measuring the relationship between attribute levels and
evaluation), to explain (where the assumption is made that the attribute levels explain the evaluations), and to predict (where the preference structures are supposed to predict future preferences and purchases). Consequently, after Zaltman et al. (1973), reliability, observational validity, content validity, construct validity, and criterion-related validity are at stake. Reliability is concerned with measuring the relationships between attribute levels and evaluations. Observational validity indicates the degree to which evaluations can be observed. Content validity is concerned with the choice of the attributes and attribute levels, and of the way that these are communicated in the attribute profiles. Construct validity is concerned with whether or not the evaluations indicate (individual) preference. (A potential problem with respect to construct validity is that, in some instances, evaluations reflect saliency. In addition, sometimes respondents' evaluations are guided by anticipations of the needs of others.) Criterion validity is concerned with how well predictions can be made from the preference structures.

In the literature on conjoint analysis, much attention is given to the reliability and the validity of the results. Different factors influencing the reliability and the validity of conjoint-analysis results have been discussed. Major problems, however, in interpreting the results of those studies, are posed by the fact that so many different measures of reliability and validity have been used, and moreover that the conjoint-analysis studies which served to test reliability and validity have many different underlying structures. These problems make it difficult to compare the findings directly.

In Appendix 3, we give an overview of 52 studies found in the literature. This overview includes those with measures of both reliability and/or validity and the underlying structures.

Although a large number of types of validity apply to conjoint analysis, only one category is used in validation research: criterion-related validity. This type of validity will be discussed in terms of both of its components, namely concurrent validity and predictive validity. The difference between these types of validity is whether or not the criterion and the predictor are measured at the same time. In the case of concurrent validity, both are measured at the same time; in the case of predictive validity they are not.

We assign each of the studies investigating the reliability and the validity of conjoint-analysis results to one of the following categories: reliability, concurrent validity, or predictive validity. Sometimes, more than one category is discussed in a single study. In those cases, the studies are put in both, or sometimes in all three, categories.

7.2.1 The reliability of conjoint-analysis results

Measures of reliability

The reliability studies (see Appendix 4) deal with stability over time, stability over methods, or with the accuracy of the estimation method.

Stability over time can be assessed by inviting respondents to participate in a retest. In the retest, the respondents' task is identical to the test (the task in the first study) in most instances. In some instances, an orthogonal set of stimuli, different from the master set, is used in the retest. Only when the stimuli in the test and the retest are identical it makes sense to calculate correlations between the input data (raw data, or direct evaluations) of both tests. If this is not the case, only correlations at the level of part-worths and importances can be calculated.
With stability over methods, conjoint-analysis results are compared either with the results of other conjoint-analysis studies with different structures, or with the results of other methods of determining the attribute importances, mostly direct methods based on the self-explicated model. Correlational measures (in most instances Pearson correlations) are used to draw conclusions about the stability over methods. Instead of correlational measures, it is also possible to use an F-test to check the extent to which the part-worth vectors of the test and retest correspond (e.g. see Reibstein et al., 1988).

The accuracy of the estimation method is determined by Monté-Carlo studies. In Monté-Carlo studies, the data are self-generated. For this reason, one sometimes speaks of synthetic studies (Malhotra, 1986). The self-generated data can be considered as true preferences. These data can then be submitted to different types of estimation methods (e.g. LINMAP, OLS) under different conditions (e.g. degree of fractioning, amount of error). The results of the estimation method (part-worths, importances) can be compared with the original true preferences (or preference scores). High correlations indicate that the estimation method is accurate.

So, in studies on the reliability of conjoint-analysis results, reliability comprises stability over time, stability over methods, and accuracy. The studies which have been conducted on the reliability of conjoint-analysis data can be classified by these aspects. In Appendix 4, an overview is given of these studies. The contents of this Appendix (the numbers and types of respondents who participated, the measure of reliability, and the test product) are helpful in interpreting the findings. In addition, and also to facilitate the interpretation of the findings, information is supplied on the structure of each specific conjoint study (Appendix 3).

**Stability over time**

From Appendix 4, it becomes clear that most studies on the reliability of conjoint-analysis data either deal with their stability over time, or with their stability over method. From the studies which have been conducted on the stability over time, we conclude that part-worths and attribute importances are stable, assessed by means of conjoint analysis. This is found consistently in most of the studies (e.g. Parker & Srinivasan, 1976; McCullough & Best, 1979; Segal, 1982; Scott & Keiser, 1984). Only in the study by Leigh et al. (1981) were the part-worths not found to be very stable.

Stability over time has been found irrespective of the number of attributes (Cattin & Weinberger, 1980). A decrease in the reliability of the results is caused by fractioning (e.g. see Leigh et al., 1981, and Reibstein et al., 1988). Reibstein et al. (1981) observed a decrease in reliability if combinations of the trade-off method and fractionated designs were applied. In the study by Cattin & Weinberger (1980), however, the degree of fractioning did not influence the reliability of the results. The (potential) problem of fractioning can be overcome by using the full-profile method (Reibstein et al., 1988). In line with this, Segal (1982) found that the full-profile method generated data which were more stable over time than those using the trade-off method.

**Stability over methods**

The outcomes of the studies on the stability over methods point in different directions. High stability over methods was found in the study by Oppedijk van Veen & Beazley (1977). Oppedijk van Veen & Beazley found that the
data of a conjoint-analysis study were stable over different data-collection methods (trade-off and full profile) and over different estimation methods (MONANOVA and PERMUT). Low stability over methods was found in the studies by Scott & Wright (1976), Heeler et al. (1979), Jain et al. (1979), Leigh et al. (1981), Wittink et al. (1982), Akash & Karagolakar (1983), Jaccard et al., (1986), and Wittink et al. (1989a). Low stability over methods is contributed to by the different study factors, i.e. the model of preference, stimulus-set construction, the data-collection method, and the estimation method. With respect to the model of preference, contrasting results have been found. According to Heeler et al., the self-explicated model is more reliable than the part-worth model, whereas Akash & Karagolakar found the opposite. Scott and Wright (1976) observed differences between the results of the self-explicated model and those of the partworth model, but do not mention which method yielded the most reliable results. Jain et al. (1979) found different patterns of importance ranks when using different estimation methods. With respect to stimulus-set construction, Leigh et al. found that fractioning, and Wittink et al. (1982) and Wittink et al. (1989a) that using different numbers of attribute levels, has a negative influence on the reliability of the data. Leigh et al. also found that comparative judgments (e.g. pair-comparisons and graded pair-comparisons) generate more reliable data than non-comparative judgments, like ratings.

Accuracy of the estimation method

With respect to the accuracy of the estimation method, a small number of studies used the so-called Monté-Carlo method. The outcomes of these Monté-Carlo studies indicated that conjoint analysis is an accurate procedure for measuring the effect of attribute levels on preference. Carmone et al. (1978) found that conjoint analysis is robust over models of preference (additive or interactive), stimulus-set construction (degree of fractioning), data-collection methods (ranking, rating), and estimation methods (MONANOVA or ANOVA). With respect to estimation methods, however, Darmon & Rouziès (1989) found that OLS yielded more reliable data than LINMAP or MONANOVA. In line with this, Mishra et al. (1989) conclude that if precision is of principal interest, OLS should be used. In the study by Darmon & Rouziès-Ségalla (1990), both OLS and MONANOVA proved to be superior to LINMAP and JOHNSON.

Factors influencing the reliability of conjoint-analysis results

Amount of information

The factor "amount of information" did not receive attention in the studies we analyzed. Nevertheless, we expect that if a lot of information is given on a relevant attribute, this attribute will be better understood than if little information is given on that attribute. In the latter case, the respondent may have to use his imagination to interpret the attribute information. This process of imagining can be different on a second occasion. We expect, therefore, more reliable part-worth utilities and importances of a particular attribute if a lot of information is given on that attribute than if little information is given.

In addition, in Chapter 6 we found that presenting information on a relevant attribute in an ambiguous manner resulted in more positive evaluations of the moderate and high levels of "price". In those instances, price is used as an indicator for the overall quality of the alternative. The evaluations of the levels of the attribute "price", then, do not reflect real
preferences in many instances. When, in a retest, less uncertainty about the attractiveness of the product-concepts is felt, or a different strategy is chosen to reduce the uncertainty, a change is expected in the preferences for the attribute levels. It is expected then that low prices are preferred to moderate and high prices. Thus, we expect that ambiguous information on a relevant attribute has a negative effect on the reliability of the part-worth utilities of the attribute "price".

Presentation format

When looking at Appendix 3, one can see that in nearly all studies on the reliability of conjoint-analysis data, textual concepts have been used. This implies that the conclusions of those studies are restricted to the textual presentation format.

We expect, however, an effect for presentation format on the reliability of the results. (The part-worths resulting from the textual and the pictorial presentation format were found to be different in our study. See Chapter 6). The findings of Smead et al. (1981) give some idea about the direction of the effects. These authors found that respondents considered the making of choices between textual representations less difficult than making choices between real products. The explanation given by Smead et al. is that the textual presentation format provides a more manageable, organized structure of data than actual products. In our terminology, since in the textual presentation format (and especially when using short textual descriptions) the attributes which can be used in forming an evaluation are already selected by the researcher, the processing of the product information (Stage One) is substantially facilitated.

The pictorial presentation format also contains more, and less organized, information on the concepts than the textual presentations, although to a lesser extent than real products. We expect, therefore, that the textual presentation format will generate more reliable data than the pictorial presentation format.

Product-category knowledge

In the studies on the reliability of conjoint-analysis results, the factor "product-category knowledge" received no attention. In our opinion, this is a severe omission since we expect that respondents with a lot of knowledge about a given product class may provide more stable evaluations of new-product concepts than respondents with little product-category knowledge. In this respect, the so-called "stability effect" may operate. The stability effect refers to the finding that respondents with superior product knowledge are more stable in their product evaluations than respondents with less product knowledge. The stability effect was found by Srull (1983), who showed that the evaluations from respondents with superior knowledge were not influenced by frequency of exposure or by mood-induction procedures. The stability effect, if operative in the case of sufficient product-category knowledge, will probably affect the reliability of the results.

Methodological considerations

In interpreting the outcomes of the studies, we have to take into account some methodological aspects. First, all studies on stability over time have been based on a short interval period (from the same session to two months). In terms of new-product development programs, this is a very short period. Secondly, in most empirical studies9 on the reliability of conjoint-

---

9 In contrast to Monté-Carlo or synthetic studies.
analysis data, only students have participated. This may have had a biasing effect on the outcomes of the studies since students have some atypical characteristics, such as low age, low spending power and high intelligence. Young and educated people may be more competent in performing a conjoint task than older, uneducated people.

Thirdly, the numbers of respondents who were invited to participate in the studies we mentioned are large. Finding statistically significant test-retest (or split-half) correlations is much easier when using large numbers of respondents, than when using small numbers of respondents. In our opinion, the strength of the effect is more important. This strength, however, is not mentioned in a number of studies.

Fourthly, according to some authors, the widespread use of correlational measures in conjoint-validation research is a major problem because the number of stimuli and the variances of the evaluations, largely determine the degree of correlation. For example, with respect to attribute importances, when a subject attaches approximately equal values to all attributes, it is possible to match evaluations closely and still produce a low test-retest correlation (Leigh et al., 1984). To overcome this problem, an F-test solution is presented by Reibstein et al. (1988). This measure tests whether or not different vectors (of part-worth utilities) are similar enough to be pooled. Wittink et al. (1989b) have some problems with this measure. They found that the measure proposed by Reibstein et al. (1988) did not decrease as measurement error increased. Moreover, Green & Srinivasan (1990) show that correlations get smaller when variances become smaller. This properly reflects the amount of true variance. So, in spite of its limitations, correlational methods are preferable to the F-test solution presented by Reibstein et al..

Fifthly, no attention is given to ACA\textsuperscript{10} as a factor with the potential for affecting the reliability of conjoint-analysis results. ACA is a data-collection method developed by Sawtooth (1986), consisting of a computerized version of the hybrid model\textsuperscript{11}. In ACA, respondents respond to questions on a computer screen by using a computer keyboard. The ACA-software package comprises an interactive interview procedure. This means that the answers given by the respondents, which are processed directly, guide the subsequent profile presentations. This makes the conjoint task both more meaningful and less demanding for the respondents.

\textbf{Conclusions}

In conclusion, the overall reliability of conjoint-analysis results seems satisfactory. Some methodological shortcomings of the studies, however, provide reasons for being careful in drawing this conclusion. In order to assess the reliability of conjoint-analysis results well, we think it is necessary to use longer interval periods (more than six months), to invite "ordinary consumers" instead of students, to concentrate on the strength of an effect, and to use correlational methods.

In addition, two major aspects in which we are interested have not received any attention until now. These aspects are "amount of information", and "product-category knowledge". Nevertheless, we expect more reliable part-worth utilities and importances of a particular attribute if a lot of information is given on that attribute than if little information is given. We also expect that missing information on a relevant attribute results in less reliable part-worth utilities of the attribute "price". Furthermore, because of the stability effect we expect that respondents with a lot of product-category
knowledge will produce more reliable results than respondents with little product-category knowledge. Also, based on the findings of Smed et al. (1981), we expect that structured information (e.g., textual concepts) will yield more reliable results than unstructured concepts (e.g., actual products). Finally, we expect that the data-collection method "ACA" will generate more reliable results than traditional conjoint-analysis data-collection methods.

7.2.2 The concurrent validity of conjoint-analysis results

Measures of concurrent validity

As we stated at the beginning of this chapter, for concurrent validity, the predictor and the criterion are measured at the same point of time. In the context of conjoint analysis, we speak of concurrent validity in the case of studies dealing with goodness-of-fit, or with hold-out predictions (see Appendix 5). In addition, similar to studies on the reliability of conjoint-analysis results, Monté-Carlo studies are also used to determine the concurrent validity of the results. Since we have already discussed how Monté-Carlo studies are conducted (in Section 7.2.1), this technique will not be discussed here.

Goodness-of-fit is a measure of the amount of variance in individual evaluations that can be explained by the attribute levels. This measure can be compared to measures of stress or badness-of-fit. (Whereas goodness-of-fit is used in OLS, badness-of-fit is used in MONANOVA). A high goodness-of-fit indicates that the evaluations are to a large extent determined by the attribute levels.

A measure related to the goodness-of-fit measure consists of calculating the ratings or rankings of the concepts from their individual part-worths, and comparing these outcomes with the observed preference scores.

Hold-out predictions are made using procedures in which, apart from the stimuli to calibrate the conjoint model, respondents evaluate an additional hold-out set of stimuli. The hold-out stimuli consist of new combinations of the attributes and attribute levels used in the calibration stimuli, and are presented in the same presentation format. From their individual part-worth utilities, predictions can be made with respect to the preference order (or rating scores) of the hold-out concepts. These predicted scores can be compared with the (actually) observed scores. Predictions can also be made of the most preferred hold-out concept (first-choice predictions, or hits). The results are presented as correlational measures, as mean squared-differences, and as percentages. High correlations, small mean squared-differences, and large percentages of correct first-choice predictions, all indicate that the data are concurrently valid.

Several studies have been conducted to assess the concurrent validity of conjoint-analysis data. An overview of these studies is given in Appendix 5. Similar to the studies on the reliability of conjoint-analysis data, both empirical and synthetic studies (Monté-Carlo studies) have been conducted. To be able to interpret the outcomes, it is necessary to know the number (and type) of respondents who participated, the measure of concurrent validity, and the test product. This information is also given in Appendix 5. Information on the underlying structure of these specific conjoint studies, e.g., presentation format and estimation method, can be found in Appendix 3.

Goodness-of-fit

The outcomes of the studies in Appendix 5 indicate that the results of
conjoint-analysis studies possess concurrent validity. The goodness-of-fit is high in most studies (e.g. Srinivasan et al. 1981; Scott & Kelser, 1984; Krishnamurti, 1988), although Acito & Jain (1980) found rather high stress values. Further support for their concurrent validity is provided by the fact that the correlations between observed and predicted rankings appear to be high (Oppedijk van Veen & Beazley, 1977; Huber & Moore, 1979; Wright & Kriewal, 1980; Akaah & Korgaonkar, 1983).

The goodness-of-fit is influenced by the model of preference (Green et al., 1981b; Moore & Holbrook, 1982), and by stimulus set construction (Scott & Wright, 1976; Cattin, 1980). According to Green et al. (1981b), the goodness-of-fit increases by adding the main-effects model to the self-explicated model, and increases even more by adding the interaction-effects model. Moore & Holbrook (1982) find that the part-worth model shows better predictive fits than the joint-space model\(^\text{12}\). With respect to stimulus set construction, Scott & Wright (1976) conclude that the goodness-of-fit decreases when there are more attributes (2, 3 or 6). This finding, however, is contradicted by Cattin (1980). In addition, Cattin finds that the use of fractioned designs has a negative influence on the internal validity of the data.

\section*{Hold-out predictions}

With respect to hold-out correlations (between observed and predicted scores) and hold-out predictions (of the first choice), most studies have found that the data are significantly better than could be expected by chance (e.g. Acito & Jain, 1980; Moore, 1980; Akaah & Korgaonkar, 1983). The hold-out correlations and predictions of the part-worth model have been compared with that of other models. Akaah & Korgaonkar (1983) found that the part-worth model and the hybrid model outperform the Huber-hybrid model\(^\text{13}\) and the self-explicated model. Green et al. (1981) found similar outcomes, with the exception that the Huber-hybrid model was not included. Huber (1975) finds that the predictive power of the part-worth model is as good as that of ideal-point models or Two-stage models\(^\text{14}\).

In a number of studies, attention is given to the concurrent validity of results which were generated by means of Adaptive Conjoint Analysis (ACA). Research on the concurrent validity of ACA data has been carried out by Klein (1986), by Boecker & Schweikl (1988), and by Green et al. (1988). The first two studies concentrated on the aspect of excluding completely unacceptable levels. Including this ACA-option of excluding completely unacceptable levels implies a shift from a compensatory to a conjunctive heuristic. This option did not seem to have an impact on the concurrent validity (Klein, 1986; Green et al., 1988). In line with this, Huber et al. (1991) conclude that ACA performs better than the full-profile method. Somewhat surprisingly, Argarwal & Green (1991) found that the self-explicated model outperformed all of the ACA models.

Boecker & Schweikl (1988) demonstrated the use of individualized attribute sets, meaning that a subject is only confronted with attributes in the conjoint task which are relevant to him/her. The authors found a better fit and more correct first-choice predictions using individualized attribute sets. Boecker & Schweikl (1988), nevertheless, do not indicate how to solve the problem of integrating the individualized attribute sets.

\section*{Monte-Carlo studies}

The synthetic studies on the concurrent validity of conjoint-analysis

---

\(^{12}\) A joint space is formed by constructing a perceptual map of stimuli and by fitting ideal points or preference vectors into that spatial representation.

\(^{13}\) For a description, we refer the reader to Akaah and Korgaonkar (1983).

\(^{14}\) In the Two-stage model, preferences are predicted as a function of subjective interpretations of the attributes, and objective attributes are linked to these subjective interpretations.
results focus attention on the estimation methods. It is concluded that the main-effects (part-worth) model is robust, even in the presence of interactions (Carmone & Green, 1981). Wittink & Cattin (1981) conclude that ANOVA is the most preferred estimation method for compensatory models, and that LINMAP provides the best concurrent validity for the non-compensatory models with a single dominant attribute.

**Factors influencing the concurrent validity**

**Amount of information**

In the literature on the concurrent validity of conjoint-analysis results, the effects of the amount of available information are not investigated. In our opinion, missing information on a relevant attribute, or ambiguous information on a relevant attribute, can affect the concurrent validity of conjoint-analysis results. More specifically, because of the difficulties in evaluating early product-concepts which lack relevant information, the corresponding logical consistency of the evaluations will probably be lower.

**Presentation format**

Since, almost without exception, the concurrent validity of conjoint-analysis results has been studied using the textual presentation format, we know little about the possible effects of other presentation formats. The study by Anderson (1987), however, which incorporated the earlier reported findings of Smed et al. (1981) indicated that written product-descriptions provided the highest averages of fit (when compared with actual products). Anderson concluded that information which is well structured, such as that in the textual presentation format, yields more concurrently valid results than poorly structured information (such as in actual products). Consequently, one can expect that a textual presentation format will also generate results which are more concurrently valid than the pictorial or the three-dimensional presentation formats.

**Product-category knowledge**

The individual factor of product-category knowledge was not measured in the studies on the concurrent validity of conjoint-analysis results. It can be expected that respondents with a lot of product-category knowledge are better equipped to interpret the early product-concepts than respondents with little product-category knowledge. An explanation of this is given by the so-called facilitation effect (e.g. Johnson & Russo, 1981; 1984; Alba, 1983; Sruil, 1983)\(^{15}\). The facilitation effect concerns the finding that individuals who already have a reasonable amount of knowledge with respect to a certain subject are more able to assimilate new facts about that subject than respondents with little knowledge. Because of this effect, one could expect that respondents with a lot of product-category knowledge will be more logically consistent in their evaluations of the early product-concepts than will respondents with little product-category knowledge, thereby yielding more internally valid data. Whether or not this is true needs to be investigated.

Another aspect of the nature of early product-concepts, which did receive some attention, is the aspect of environmentally-correlated attributes (e.g. price and capacity). The reason for investigating this aspect is that, because of the fact that in the real world some attributes are related, systematically varying the attributes in the attribute profiles might result in unrealistic concepts, possibly leading to low respondent motivation. Investigating this
effect, Moore and Holbrook (1990) found that including environmentally-correlated attributes did not affect the concurrent validity of conjoint-analysis results.

Methodological considerations
Similar to the studies on the reliability of conjoint-analysis data, a methodological comment can be made with respect to the respondents. Students instead of "ordinary" consumers have been used.

Conclusions
In conclusion, the concurrent validity of conjoint-analysis data seems satisfactory. In order to assess the reliability of conjoint-analysis data well, however, we think it is necessary to invite "ordinary consumers" instead of students.

In addition, the factors "amount of information", "presentation format", and "product-category knowledge" have received little or no attention. We expect that smaller amounts of information on relevant attributes will yield results which are poorer in terms of concurrent validity. With respect to the factor "presentation format", we refer to the study by Anderson (1987). This study dealt with the effect of presentation format on the concurrent validity of conjoint-analysis results. In this study, however, Anderson simultaneously varied presentation format and amount of information. From Anderson's study, therefore, no direct generalizations can be drawn with respect to the isolated effect of presentation format on the concurrent validity of conjoint-analysis results. Nevertheless, because of this study, we can better formulate expectations with respect to the direction of the effect. We expect that structured information (e.g., entirely textual concepts) will yield results which have a higher concurrent validity than unstructured concepts (e.g., partly pictorial concepts).

In addition, because of the facilitation effect, we expect that respondents with a lot of product-category knowledge will produce results which have a higher concurrent validity than those from respondents with little product-category knowledge.

Finally, some extra research is needed with respect to ACA. ACA facilitates practical fieldwork. Studies investigating the concurrent validity of its results, nevertheless, come to opposing conclusions. In our opinion, therefore, it is necessary to investigate the concurrent validity of conjoint-analysis results obtained by means of ACA. Due to the fact that ACA makes the conjoint task both more meaningful and less demanding for the respondents, it is expected that ACA generates results which have a higher concurrent validity than those from using traditional conjoint data-collection methods.

7.2.3 The predictive validity of conjoint-analysis data
Measures of predictive validity
Because it is an element of criterion-related validity, predictive validity concerns the degree to which the concept under consideration enables one to predict the value of some other concept which constitutes a criterion. In studies on the predictive validity of conjoint-analysis results, preference models based on early product-concepts (predictors) are compared with evaluations (or choices) of actual products (criteria). This begs the question as to whether or not the causal attribute-level preference relationships, measured early in the new-product development process, still hold at the
moment that the respondents can buy the product on the market.

In contrast to concurrent validity, assessing the predictive validity of conjoint-analysis results requires a time interval between measuring the criterion and measuring the predictor. To be able to draw conclusions about the predictive validity of conjoint-analysis results, the length of the time interval should correspond (as much as possible) to actual lead times in real new-product development programs. In most instances, these lead times vary between one and several years. Preferably, the time interval should correspond to these lead times.

The measures which are used to determine the predictive validity of conjoint-analysis results are similar to the hold-out profile measure which was discussed with respect to concurrent validity. The main difference, however, is that in research on predictive validity, the hold-out profiles are replaced by actual products. Respondents are invited to evaluate (or choose between) actual products. The observed scores are compared with the predicted scores. To illustrate, if a respondent’s preference model indicates large part-worth utilities both for the lowest price level and for the attribute-level “with thermos flask”, one expects this respondent to evaluate, in a positive way, actual filter coffee-makers which possess these attribute levels.

Buying a product in a shop differs in a great number of ways from the conjoint task in the laboratory. Due to these differences, the predictive validity of the results may be negatively affected (Ajzen & Fishbein, 1977). The four major aspects which differ between “actual” behavior and behavior in test situations, mentioned by Ajzen & Fishbein (1977) are the action, the target, the context and the time.

In new-product concept testing, the action is the conjoint task. This can be the ranking of new-product concepts according to preference. In actual buying behavior, the action is buying the product. The target in product-concept testing is comprised by the early product-concepts, whereas actual products are considered in buying behavior. In most instances, the context of new-product concept testing is a laboratory. Actual buying behavior, by contrast, takes place in a shop. Finally, when a particular new-product development project is in the concept stage \( T_0 \), it will still take some time before the new product will be introduced onto the market \( T_1 \).

Considering these aspects, it is evident that there are large differences between carrying out a conjoint task and actual buying behavior. The differences are summarized in Table 9.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Conjoint task</th>
<th>Buying behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>Evaluating</td>
<td>Buying</td>
</tr>
<tr>
<td>Target</td>
<td>Early product concept</td>
<td>Actual product</td>
</tr>
<tr>
<td>Context</td>
<td>Laboratory</td>
<td>Shop</td>
</tr>
<tr>
<td>Time</td>
<td>( T_0 )</td>
<td>( T_1 )</td>
</tr>
</tbody>
</table>

Reviewing many studies of attitude-behavior consistency, Ajzen & Fishbein (1977) found that the higher the correspondence of aspects between the predictor and the criterion, the more consistent the attitude-behavior relationship. In later work, this aspect of correspondence is
referred to as the principle of compatibility (e.g. Ajzen, 1988).

In assessing the reliability of conjoint-analysis results, the aspect “time” is the only aspect which differs. In assessing their predictive validity, all four aspects differ. Differences in these aspects, between the predictor and the criterion, may affect the predictive validity of conjoint-analysis results.

With respect to action, Johnson & Russo (1981; 1984) found different types of information processing for a choice task (e.g. buying a product) and an evaluation task. In a choice task, a consumer can quickly eliminate the number of alternatives. In an evaluation task, by contrast, a consumer is asked to consider each of the alternatives. This means that there is a low correspondence on the action between a conjoint task and actual buying behavior.

Under the heading “target”, the central problem of this thesis is considered, i.e. the nature of early product-concepts. Targets are both new-product concepts as predictors and actual products as criteria. Changes to the product-concept in the production stage, and decisions with respect to the introduction of the new product, may explain discrepancies between preferences for actual products and preferences for product-concepts.

The behavior of consumers in the context of a laboratory differs from their behavior in natural settings. Cook and Campbell (1979) describe a large number of differences (e.g. hypothesis guessing). The time interval between \( T_0 \) and \( T_1 \) can be from a couple of months to several years. During such a long period, many interrupting events (Bettman, 1979) may occur. Needless to say, the longer this interval (difference between \( T_0 \) and \( T_1 \)), the more interrupting events that are likely to occur. These interrupting events can be both internal to the individual (e.g. forgetting) or external to the individual. With respect to the latter, all kinds of situational influences can be mentioned (see Belk, 1975).

Considering the above-mentioned differences in aspects between the predictor and the criterion, it will be difficult to obtain results which have a high predictive validity.

Below, we will first discuss studies which have been conducted on the predictive validity of conjoint-analysis results. For information on the number (and type) of respondents who participated in them, the measure of predictive validity, and the test product, we refer the reader to Appendix 6. (Information on the structure of these specific conjoint studies can be found in Appendix 3).

### Predictions of choices and evaluations of actual products

From the outcomes of the studies of the predictive validity of conjoint-analysis data (see Appendix 6), positive conclusions can be drawn. In all studies the predictions of the choices of actual products, on the basis of a conjoint study using early product-concepts, were better than could be expected by chance (e.g. Parker & Srinivasan, 1976; Wittink & Montgomery, 1979; Wright & Kriewal, 1980). In two studies, the outcomes are expressed in terms of the increase in market share (Page & Rosenbaum, 1987; Woodside & Pearce, 1989).

### Factors influencing the predictive validity

Amount of information

In the studies of the predictive validity of conjoint-analysis results, no attention has been given to the amount of information which was present in the new-product concepts. Following the principle of compatibility (e.g.
Ajzen, 1988), one would expect that a high degree of correspondence between the targets in the predictor and the criterion would lead to results which have higher predictive validity than that when targets have a low degree of correspondence.

In addition, it is interesting to have an insight into the predictive validity of part-worth utilities of the attribute "price", in relation to the amount of information which is presented on another relevant attribute. It was found in Chapter 6 that if ambiguous information was presented on the attribute "form" for filter coffee-makers, in a large number of instances, moderate and even high levels of the attribute "price" were preferred to low levels of the attribute "price". Since we reasoned that those preferences are the result of feeling insecure about evaluating the alternatives (and not of considering how much one is willing to pay), we expect that the corresponding predictive validity of the part-worth utilities of the attribute "price" will be low.

Again, as in the studies on the reliability and the concurrent validity of conjoint-analysis results, almost all studies have been conducted using the textual presentation format. Consequently, we do not know what the predictive validity is of conjoint-analysis results which are based on pictorial or three-dimensional presentations. This needs to be further investigated.

**Product-category knowledge**

As before, no attention has been given to individual characteristics in the studies on the predictive validity of conjoint-analysis results. One can expect that consumers with a lot of product-category knowledge generate results which have higher predictive validity than consumers with little product-category knowledge.

**Methodological considerations**

All studies on the predictive validity of conjoint-analysis results discuss the prediction of choices or evaluations of actual products. The other aspects mentioned by Ajzen & Fishbein (1977) do not differ between the predictor and the criterion in all studies. First, in some studies the aspect of time is reduced to zero (T0 = T1) and present behavior (instead of future choice-behavior) is predicted from the preference models. For example, in the study by Parker & Srinivasan (1978), the respondents' current physician was predicted. Wittink & Montgomery (1979) and Srinivasan (1988) predict which job has been chosen out of the jobs which were offered in the past. In our opinion, the outcomes of these studies are not very informative since the direction of causality is ambiguous. The fact that a certain physician or a certain type of job had already been chosen may have itself influenced the behavior in the conjoint task. In the study by Green et al. (1978), preferences for actual vacation sites were measured in the same session in which the part-worth models were determined. In these studies, the potential effect of interrupting events was minimized. This may have resulted in artificially good predictions. The generalizability of these findings, therefore, may not be very wide.

**Conclusions**

In conclusion, when one considers the large number of differences in aspects between predictors and criteria, the predictive validity of conjoint-analysis results nevertheless seems satisfactory. A large number of the studies, however, does not introduce a substantial time interval.
In addition, no attention has been paid to either amount of information or presentation format, or to product-category knowledge. Following the line of reasoning of Ajzen & Fishbein (1977), we can expect that a high degree of correspondence between targets as predictors and those as criteria, will lead to a higher predictive validity than results which are obtained with a low degree of correspondence. Therefore, if a lot of information is presented on relevant attributes, and if the pictorial presentation format is used, it is likely that results will be obtained which possess higher predictive validity than those if little information is presented on relevant attributes, or if the textual presentation format is used. In addition, in those instances where moderate and even high levels of the attribute "price" are preferred to low levels of the attribute "price", we expect that the predictive validity of the part-worth utilities of the attribute "price" will be low.

It has already been reasoned that product-category knowledge probably has a positive influence on the reliability (stability effect) and on the concurrent validity (facilitation effect) of conjoint-analysis data. From this, we expect that product-category knowledge will also positively influence the predictive validity of the results. Further support for this expectation is given by the finding that respondents whose attitudes are based on direct behavioral experience with the attitude object, more often behave according to their stated attitudes than do respondents who lack the behavioral experience (Fazio & Zanna, 1981).

Finally, we expect that ACA generates results which are higher in terms of predictive validity than traditional conjoint-analysis data-collection methods. The explanation is similar to that for the variable "product-category knowledge", i.e. positive scores with respect to reliability and concurrent validity provide better chances of obtaining results which have predictive validity than negative scores on these indices.

7.3 Summary and conclusions

In this chapter, studies investigating the validity of conjoint-analysis data have been discussed. Before doing so, the concept of validity was disentangled into three indices: reliability, concurrent validity and predictive validity. Reliability concerns the observation of covariation between two or more research variables. In order to measure the covariation itself in a reliable way, the covariation must be stable over time, stable over method, and must be accurately measured.

In conjoint analysis, the concurrent validity of the results can be assessed by determining the goodness-of-fit or hold-out predictions or Monté-Carlo studies. In studies on predictive validity, generalizations are made from findings on early product-concepts to actual products.

With respect to the studies which have been conducted with respect to each of the indices of validity, it can be concluded that conjoint-analysis results are reliable, and score moderately to highly on concurrent and predictive validity. This conclusion, however, is drawn with some methodological drawbacks in mind. The most important ones are that short interval periods have been used in studies on reliability and on predictive validity, and that, almost without exception, only the textual presentation format has been used. In studies on predictive validity, sometimes predictions are even made of past behavior. In those instances, rationalisation may have been crucial in determining the relationship between the independent variable (preference structure) and the dependent variable (e.g. choice of a job).
Both methodological drawbacks may have resulted in higher correlations and in better predictions.

With respect to the factors which were found to be important in the perception and evaluation of early product-concepts (amount of information, presentation format, product-category knowledge), no straightforward conclusions can be drawn on the basis of the studies which have been presented. Almost all the studies have used the textual presentation format and used students as respondents. Because of these facts, we still know very little about the reliability, the concurrent validity, and about the predictive validity of conjoint-analysis results based on presentation formats other than the textual presentation format (i.e. pictorial and three-dimensional presentation formats).

Based on the findings of Smead et al. (1981) and Anderson (1987), we expect that textual presentations will produce more reliable results with higher concurrent validity than those from using pictorial presentations.

Following the principle of compatibility (Fishbein & Ajzen, 1977), by contrast, it can be expected that the predictive validity of the part-worth utilities of attributes which are presented with a lot of information will be higher than that of attributes on which the information is either missing or ambiguous. So, the three-dimensional and the pictorial presentation format will yield results which have higher predictive validity than the textual presentation format because of closer resemblance to actual products.

This poses an interesting problem since, as we stated earlier, reliability and concurrent validity are necessary conditions for obtaining results which have predictive validity. We think, however, it is possible that textual representations, which we might expect to give greater reliability and predictive validity than pictorial and three-dimensional representations, give results which have less predictive validity than both latter presentation formats. Our explanation for this is that because of the large discrepancy between concepts and actual products, the textual representations will perform very poorly in terms of predictive validity. The pictorial presentation format and the three-dimensional presentation format, by contrast, will generate results which have a rather high (though not optimal) predictive validity because of their closer resemblance to actual products. (Since the reliability and the concurrent validity are not optimal, the predictive validity cannot be optimal either.)

Furthermore, based on other studies, it can be reasoned that product-category knowledge has a positive influence on the reliability (stability effect) and on the concurrent validity (facilitation effect) of conjoint-analysis data. It can also be expected that product-category knowledge will positively influence the predictive validity of the data.

In studies on concurrent validity, it was found that ACA is a promising new data-collection method. We think, however, that because of the substantial differences in the results which were obtained, some additional insight is needed into the concurrent validity of the results. In addition, insight is also needed on the reliability and the predictive validity of ACA-generated results. Because of the interactive nature of this new data-collection method, it is expected that more reliable results, with higher concurrent and predictive validity, will be obtained by using this data-collection method rather than by using more traditional data-collection methods.
7.4 Hypotheses tested

In Chapter 7, a large number of expectations were formulated. In our opinion, each of the expectations needed to be tested empirically. In this thesis, however, a selection has been made from those expectations. The most important criteria for selecting specific expectations for empirical testing were the results found in the first part of this thesis (Chapter 6). In this chapter, it was stated that the variables "amount of information", "presentation format", and "product-category knowledge" had an effect on the perception and/or the evaluation of early product-concepts. Thus, it was these variables that were to be subjected to further analysis in terms of reliability and validity. In addition, attention was to be given to the promising new data-collection method ACA.

An attempt was made to test all expectations about these factors with respect the reliability and the (concurrent and predictive) validity of conjoint-analysis results. It is important to note that the factor "amount of information" did have an effect on the evaluations of the attribute "price", and did not affect the evaluations of the attribute which was manipulated in terms of amount of information. Consequently, the only expectations tested were those about the effects of amount of information on the reliability and the validity of the part-worth utilities of the attribute "price".

Two expectations were not tested. Firstly, the testing of the expectation about the effects of amount of information, on the concurrent validity of the part-worth utilities of the attribute "price", posed a problem. The reason for this was the fact that the results of different experiments would have to be compared. Since, in the different experiments, different numbers of concepts and different numbers of attributes were used, it is not possible to compare the results of their respective goodness-of-fit measures. Secondly, the effect of the data-collection method on the concurrent validity of conjoint-analysis results was not tested either. The reason for this was that ACA does not provide a comparable measure of goodness-of-fit. ACA has a built-in measure for assessing the goodness-of-fit based on hold-out profiles, but the hold-out profiles are, as is made clear to the respondents, presented in the order of increasing overall utility. The evaluations, expressed in terms of buying intentions, are only accepted if they are at least as large as the preceding one(s). This results in artificially high correlations for the hold-out predictions, which cannot be compared with correlations of hold-out predictions from using traditional data-collection methods.

The expectations that were tested empirically were formulated in terms of hypotheses. We present these hypotheses below, ordered according to the dependent variables.

Hypotheses about the reliability of conjoint-analysis results

H7: If information on a relevant attribute is ambiguous, this results in less reliable part-worth utilities of the attribute "price" than those if the attribute information is unambiguously presented.

H8: If the textual presentation format is used, more reliable results are generated than if the pictorial presentation format is used.

H9: Respondents with a lot of product-category knowledge will produce more reliable results than those from respondents with little product-category knowledge.
H10: If the ACA data-collection method is used, more reliable results will be generated than if a traditional conjoint analysis data-collection method is used.

In Figure 28, these hypotheses about the reliability of conjoint-analysis results are presented graphically.

Figure 28: The hypotheses about the reliability of conjoint-analysis results

Hypotheses about the concurrent validity of conjoint-analysis results

H11: If the textual presentation format is used, this will yield results which are higher on concurrent validity than those if the pictorial presentation format is used.

H12: Respondents with a lot of product-category knowledge will produce results which are higher on concurrent validity than those from respondents with little product-category knowledge.

In Figure 29, the hypotheses about the concurrent validity of conjoint-analysis results are summarized.

Figure 29: The hypotheses about the concurrent validity of conjoint-analysis results

Hypotheses about the predictive validity of conjoint-analysis results

H13: If moderate or high levels of the attribute "price" are preferred to low levels of the attribute "price" (as a result of ambiguous information on a relevant attribute), the predictive validity of the part-worth utilities of the attribute "price" will be lower than that if the low level is preferred.
H14: If the pictorial presentation format is used, results which are higher on predictive validity will be obtained than that if the textual presentation format is used.

H15: Respondents with a lot of product-category knowledge will produce results which are higher on predictive validity than that from using respondents with little product-category knowledge.

H16: If the ACA data-collection method is used, this will yield results which are higher on predictive validity than those if a traditional conjoint-analysis data-collection method is used.

In Figure 30, the hypotheses about the predictive validity of conjoint-analysis results are presented.

Figure 30: The hypotheses about the predictive validity of conjoint-analysis results
8. Operationalizations for Testing Hypotheses about the Reliability and the Validity of Conjoint-Analysis Results

8.1 Reliability

To assess the reliability of the conjoint-analysis data, we conducted a retest. The data from a retest can be compared in at least three different ways with the data of the original test. These ways correspond to the degree to which the data are processed.

At the lowest level of data-processing, the evaluations are compared directly. In that case, the Pearson correlation between the evaluations of the set of product-concepts in the test and the evaluations of the same set in the retest are calculated for each respondent.

At an intermediate level of data-processing, the vectors of part-worth utilities from the test and the retest, resulting from a (dummy-type of) regression analysis, are compared. In these part-worth vectors, all levels of each of the attributes are included, minus one for each of the attributes (the complementary attribute level). For each respondent, the Pearson correlation between the vectors of part-worth utilities of the test and the retest can be calculated.

At the highest level of data processing, vectors of importances are compared. After having assessed the part-worth vectors, the importances are determined in two subsequent steps. First, the part-worth utilities of the complementary attribute levels are determined. Second, for each of the attributes, the ranges between the most positive and the most negative part-worth utilities are determined. These ranges comprise the importances. The importance-vector of a respondent is the list of importances corresponding to each of the attributes. Pearson correlations can then also be calculated for each respondent on the basis of the importance-vectors of the test and the importance-vectors of the retest.

Whenever possible, we made use of all three measures of retest-reliability because, although they are highly correlated, the outcomes are interpreted differently. By calculating correlations between the raw data of the test and the retest, one can directly observe the stability in consumer evaluations. These evaluations are not inflated by potential unreliability in the estimation procedure which may be present in the part-worths and importances. By comparing the part-worth vectors of the test and the retest, insight is obtained into the stability of attribute-level preferences. By comparing the importance-vectors of the test and the retest, insight is obtained into the stability of attribute importances. Whereas the first measure focuses on the input for performing a conjoint analysis, both last measures focus on the output of conjoint-analysis, i.e. the conjoint-analysis results.

Another measure of reliability concentrates on the stability of preference regarding a single attribute. In this measure the numbers of switches in the most preferred attribute-level are counted.

The actual execution of the retest, in all experiments, was more than six
months after the test. In the experiments, members of the general public were recruited instead of students.

8.2 Concurrent validity
To test concurrent validity, we use the goodness-of-fit measure. This measure is obtained by subjecting all the evaluations of the stimuli, included in the fractional factorial design, to a conjoint analysis. The goodness-of-fit is measured by determining the adjusted $R^2$ (i.e., the goodness-of-fit measure when the OLS estimation procedure is used) for each respondent.

A second method, which is frequently used in assessing the concurrent validity of conjoint-analysis results, is the use of hold-out predictions (see Section 7.2.2). With this method, in addition to the stimuli which are needed to estimate the model (the calibration set), an extra set of stimuli is presented to the respondents (the hold-out set). To avoid respondent fatigue, however, no use was made of hold-out predictions in the tests conducted here.

8.3 Predictive validity
The measures of predictive validity used in most studies do not take actual buying behavior as the criterion, (see Section 7.2.3). The main reason for this is the time lag between participating in a concept test and the actual buying of the particular product. In the case of consumer durables, this period may be very long. We were aware of this problem, but nevertheless wanted to assess the predictive validity of the conjoint-analysis results in this way. One possible way to overcome the time-lag problem was to invite as many respondents as possible. Following the law of large numbers, at least some of the respondents would have bought the particular product after some (relatively short) time. As an illustration, if a filter coffee-maker is bought every six years, and the number of respondents is two hundred, one can expect that after one year about thirty-three respondents would have bought a new filter coffee-maker. By choosing this approach, it would be possible to gain some insight into the predictive validity of conjoint-analysis results. Testing the hypotheses, however, may have become more difficult since only small numbers of respondents would appear in the subgroups. For this reason, in those instances where the numbers of respondents in one of the subgroups was less than ten, we introduced the magazine test. First, however, we will discuss how we conducted the actual buying test.

Actual buying test
The respondents who participated in Experiments 2, 3 and 4 were sent a very short postal questionnaire. In this questionnaire, the question was asked as to exactly when the filter coffee-maker which was used most often was bought. In this questionnaire, some cues were given to facilitate correct responses (i.e., "check your financial administration, and guarantee certificates"). To find as many respondents as possible who had bought a new filter coffee-maker, the questionnaire was sent out as late as possible. As a result, the time interval between conducting the experiments and sending out the questionnaire varied from about six months to about eighteen months.

The respondents who had bought a filter coffee-maker after they had participated in the experiment were contacted for a telephone-interview. In this interview, the respondents were asked to respond to questions about
the attributes of their recently bought filter coffee-makers.

**Magazine test**

The magazine test was a simulation of the buying of the particular product. In the magazine test, the respondents were confronted with commercial magazines with information on actual products, and were invited to indicate which product they would buy if their current product had to be replaced.

In both tests, the actual buying test and the magazine test, the following procedure was used to assess the predictive validity of the part-worths. First, each of the alternatives (bought on the market or selected from the magazines) were described in terms of its constituent attribute-levels. Second, at each attribute level, the correspondence was determined between the observed score (largest positive part-worth utility) and the predicted score (the attribute level which was chosen). In the situation where the attribute-level with the largest positive part-worth utility was reflected in the selected (or bought) alternative, we could speak of a match. For example, if a particular respondent had scored a part-worth utility of .34 for the level 'present' of the attribute "thermos flask" (and therefore -.34 for the level 'absent' of that attribute), we could speak of a match if this respondent then selected a filter coffee-maker which had a thermos flask. A large proportion of matches for attributes would indicate that the predictive validity of the part-worth utilities was high. In testing hypotheses, the numbers of matches for the attributes were compared between different conditions.

To determine whether or not the form of the filter coffee-maker one has chosen (in the magazine test or in the actual buying test) matches the most preferred attribute level, we made use of the study by De Bont & De Graaf (1989). In this study, the underlying preferential dimensions, which consumers use when evaluating filter coffee-makers, were determined by means of joint-space analysis. These dimensions were "integrated form versus unintegrated form" and "simple technology versus advanced technology". The product which was chosen was described in terms of these dimensions, as were the levels of the attributes in the profiles. If the product chosen possessed the attribute level with a positive part-worth utility larger than both other attribute levels, we could speak of a match.

To assess the predictive validity of importances, a comparison was made between the group of respondents who selected the level of a particular attribute with the largest positive part-worth utility (match), and the group of respondents who did not (mismatch). The comparison made was with respect to the mean importance of that particular attribute. The importances were regarded as predictively valid if the mean importance was substantially higher in the group with the matches than in the group with the mismatches. The rationale behind this was the assumption that if one of the attributes in a concept test was found very important for a particular respondents (as indicated by a large importance), this respondent would do their best to guarantee that the most preferred level of this important attribute was reflected in the product they would buy. In the case of an unimportant attribute, this effort would be much smaller.

In order to have enough respondents of each of the sub-groups (matches and mismatches), the predictive validities of the importances were not compared between conditions in testing the hypotheses. This test is added to enlarge insight into the predictive validity of conjoint-analysis results per se.
8.4 Data-collection method

To test the effect of ACA on the reliability of the data, a fifth experiment was conducted (Experiment 5). In Experiment 5, the ACA software-package of Sawtooth (1986) was used. The ACA-procedure consisted of four parts. In part one, the respondents were invited to rank-order the levels of each attribute (5 altogether) according to preference. In the second part, respondents were invited to indicate how important it was to them that each of the most preferred attribute levels is present when buying such a product. In the third part, respondents were confronted with pairs of stimuli and were invited to make graded comparisons. In the full-profiles, attribute levels were specified for each of the (5) attributes. The respondents were shown five different pairs of profiles. In the fourth part, the respondents were asked to indicate their buying intentions with respect to a number of sequentially presented stimuli (5).

To be able to attribute differences in data to the different data-collection methods used, the attributes and the attribute levels applied in Experiment 5 were identical to the ones which were used in Experiment 4. The data-collection method which was chosen as the traditional conjoint analysis data-collection method in Experiment 4 was full-profile ranking.
9. Testing Hypotheses about the Reliability and the Validity of Conjoint-Analysis Results

In Chapter 7, a number of studies investigating the reliability and the validity of conjoint-analysis results were discussed. In these studies, very little attention was paid to the variables "amount of information", "presentation format", and "product-category knowledge". We think, however, that these variables of the Model of Product-Concept Evaluation should not be disregarded since, as we demonstrated in Chapter 6, these variables affect product-concept evaluations. As a consequence, these variables should be included in answering the second research question (how reliable and how valid are evaluations of early product-concepts?). To find out what the consequences of these variables are in terms of reliability and validity, we formulated a number of hypotheses. In addition to the variables already mentioned, the influence of another variable, data-collection method (i.e. ACA) was to be investigated.

In this chapter, we will present the results of the test of each of the hypotheses. For an important part, the experiments discussed in this chapter are based on Experiments 2 to 4, as discussed in Chapter 6. In addition, one new experiment (Experiment 5) was conducted. In the table below, we show which experiments were involved in testing the effects of which factor(s).

Table 9: Overview of which variables are involved in which experiments (indicated by x)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
<th>Experiment 4</th>
<th>Experiment 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of information</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation format</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product-category knowledge</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Data-collection method</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

Since the study of the reliability and the validity of conjoint-analysis results was for the most important part based on experiments already described, no discussion will be made in this chapter of the socio-demographic or coffee-drinking related issues of the respondents.

In line with the structure of Chapter 7, in this chapter the reliability and the validity (both concurrent and predictive) of the results are discussed separately.

9.1 An empirical study of the reliability of conjoint-analysis results

With respect to the reliability of conjoint-analysis results, four hypotheses (H7 to H10) were formulated. Each of these will be separately discussed. In Chapter 5, among other issues, we described the operationalizations of the variables "amount of information in the stimuli", "presentation format" and
"product-category knowledge". The experimental design, and the respondents, were discussed in Chapter 6. Since these aspects were identical in the retests, they will not be discussed in this chapter. The operationalizations of reliability and of data-collection method were discussed in Chapter 8.

9.1.1. The effects of amount of information

In Experiment 4 (see Section 6.4), we found that if information on the attribute "form" was presented in an ambiguous manner (e.g. by means of schematic three-dimensional stimuli), this resulted in a large proportion of respondents preferring a moderate price, or even a high price, rather than a low one (test of H3). Large differences in such preferences were found between the schematic three-dimensional stimuli and the realistic three-dimensional stimuli (70% versus 48%). To a lesser extent, differences were also found between the textual stimuli and the pictorial stimuli (45% versus 29%).

We expected that, in a retest less uncertainty would be felt about the form of filter coffee-makers than in the test. The reason for this was that, because of having participated in a concept test before, the respondents would feel more confident about evaluating the new-product concepts. As a result, the need to reduce this uncertainty by preferring moderate or high prices to low prices would not be very strong any more. Thus, H7 was grounded on this assumption.

H7: If information on a relevant attribute is ambiguous, this results in less reliable part-worth utilities of the attribute "price", than if the attribute information is unambiguously presented.

To test Hypothesis 7, test-retest data from Experiments 2, 3, and 4 were employed. In Experiment 4, information on the attribute "form" was missing. In Experiment 3, information on the attribute "form" was presented either textually or pictorially, and in Experiment 2, information on the attribute "form" was presented three-dimensionally. In the retest of Experiment 2, only the schematic three-dimensional stimuli were used. In different terminology, information on the attribute "form" is at a low degree of realism in Experiment 4 and at an intermediate degree of realism in the Experiments 2 and 3. In line with Figure 6, it can be stated that the realism of the information increases from no information, through textual and pictorial information to three-dimensional information. The ambiguity of the product information decreases following this sequence.

If in the retest a level of the attribute "price" different from that in the test was preferred, one could state that there was a change in the most preferred price level. Many changes in the most preferred price level would indicate a low reliability in the part-worth utilities of that attribute. In the table below, a comparison is made of the numbers of changes between the test and the retest in terms of the most preferred level of the attribute "price". The number of changes are compared between the concepts with different amounts of information about the attribute "form".
Table 10: Number of changes between the test and the retest in most preferred price level, as a function of amount of information on the attribute "form" (N=75)

<table>
<thead>
<tr>
<th>Amount of information about form</th>
<th>No information</th>
<th>Some information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>schematic</td>
</tr>
<tr>
<td>(n=19)</td>
<td></td>
<td>3D^8</td>
</tr>
<tr>
<td>textual</td>
<td>(n=8)</td>
<td>(n=36)</td>
</tr>
<tr>
<td>pictorial</td>
<td>(n=12)</td>
<td></td>
</tr>
<tr>
<td>8 (42%)</td>
<td>2 (25%)</td>
<td>3 (25%)</td>
</tr>
<tr>
<td></td>
<td>18 (50%)</td>
<td></td>
</tr>
</tbody>
</table>

In Table 10, it can be seen that large numbers of changes in the most preferred price level occur both in the "no information" condition and in the "schematic three-dimensional" condition. When looking carefully at these findings, one can without statistical testing conclude that the reliability of the part-worth utilities of the attribute "price" does not linearly increase with decreasing ambiguity. This is not surprising because in these conditions the proportions of respondents who preferred a moderate or a high price were found to be largest in the first place, as was shown in Table 8.

The large number of changes in the preference for a particular level of the attribute "price", however, cannot solely be attributed to the ambiguity of the information on the form. This is indicated by the finding that the numbers of changes were considerably lower in the textual and the pictorial conditions than those in the schematic three-dimensional condition, although in these conditions the information on the attribute "form" was presented even more ambiguously. A possible explanation may be that the number of attributes in the pictorial and textual conditions (4) was smaller than that in the schematic three-dimensional condition (6). A reduction in the number of attributes may have facilitated the conjoint task, leading to a reduction in insecurity.

We conclude, therefore, that Hypothesis 7 should be rejected. Nevertheless, we draw a second conclusion which states that if the evaluation task becomes more difficult for the respondents (e.g. by including many attributes in the profiles), they will be more inclined to accept moderate and high prices as opposed to low prices. We can add to this that the reliability of the evaluations of the attribute "price" is low when obtained under a lot of uncertainty.

9.1.2. The effects of presentation format
In Experiment 3 (see Section 6.3), large differences were found in the part-worth utilities of the attribute "form" between the textual and the pictorial conditions. In relation to this finding, because textual information is more structured than pictorial information, we formulated Hypothesis 8.

*H8: If the textual presentation format is used, more reliable results are generated than if the pictorial presentation format is used.*

A retest was conducted eight months after the test (Experiment 3). The task and the stimuli in the retest were identical to those in the test. In this section, we assess the effect of presentation format on the reliability of
conjoint-analysis results at all three levels of data processing, i.e. raw data, part-worth vectors, and importance vectors.

We add that the findings are based on small numbers of respondents. This means on the one hand that, on statistical grounds, it was more difficult to find statistically significant results, and on the other hand that in spite of randomisation procedures, the possibility increased for major differences to exist between the respondents in both conditions. These differences may have disturbed the test of the hypothesis. In addition, the results were likely to differ from the normal distribution, excluding the use of certain statistical testing techniques.

Results for the raw data
Table 11: Test-retest correlations for the raw data as a function of presentation format (N=20)

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Textual (n=8)</th>
<th>Pictorial (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SD</td>
<td>.73 .14</td>
<td>.60 .13</td>
</tr>
<tr>
<td>df F</td>
<td>1,18 4.87*</td>
<td></td>
</tr>
</tbody>
</table>

* p < .05

In Table 11, we can see that the test-retest correlations between the direct evaluations (raw data) were significantly higher if the textual presentation format was applied rather than the pictorial presentation format.

Results for the part-worth vectors
In the case of the part-worth vectors, the standard deviation (SD) was .55 in textual condition and .35 in pictorial condition. The distributions of the test-retest correlations were skewed, leading to large differences between the mean and the median (.54 and .83 in the textual condition, and .63 and .72 in the pictorial condition). Therefore, we conducted a median test instead of an ANOVA.

Table 12: Test-retest correlations between the part-worth vectors as a function of presentation format (N=20)

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Textual (n=8)</th>
<th>Pictorial (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Probability</td>
<td>.83 .65</td>
<td>.72 .65</td>
</tr>
</tbody>
</table>

The median test, the results of which are presented in Table 12, did not highlight substantial differences between the test-retest correlations between the part-worth vectors of both conditions.
Below, we focus on the stability of the part-worth utilities of the levels of the attribute "form". The evaluations of these attribute levels showed large differences between both presentation formats in Chapter 6. In the table below, figures are shown for the test-retest correlations between the part-worth utilities for each of the attribute levels of the attribute "form", and for both presentation formats.

Table 13: Test-retest correlations between the part-worth utilities of the attribute "form" as a function of presentation format (N=20)

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Textual (n=8)</th>
<th>Pictorial (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>level of form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philips</td>
<td>.35</td>
<td>.67</td>
</tr>
<tr>
<td>Braun</td>
<td>.55</td>
<td>.54</td>
</tr>
<tr>
<td>Severin</td>
<td>.53</td>
<td>.50</td>
</tr>
</tbody>
</table>

In Table 13, one can see that the largest difference exists with respect to the attribute level "Philips". To test whether or not this difference in correlations was statistically significant, a test was conducted in which the correlations were transformed into z-scores (see Hays, 1981 p.466-467). The difference did not appear to be statistically significant. The same holds for the two other differences.

Results for the importance vectors

In the case of the importance vectors, the standard deviation was .58 in textual condition and .33 in pictorial condition. This difference, however, was not statistically significant (Cochran C = .76, p = .10). The distributions were not skewed and only small differences existed between the mean and the median (.32 and .33 in the textual condition, and .56 and .59 in the pictorial condition). Therefore, we conducted an ANOVA.

Table 14: Test-retest correlations between the importance-vectors as a function of presentation format (N=20)

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Textual (n=8)</th>
<th>Pictorial (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>.32</td>
<td>.56</td>
</tr>
<tr>
<td>SD</td>
<td>.58</td>
<td>.33</td>
</tr>
<tr>
<td>df</td>
<td>1,18</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>1.14</td>
<td></td>
</tr>
</tbody>
</table>

Although the mean of the test-retest correlations between the importance vectors was somewhat higher in the pictorial condition, this difference was not statistically significant.

Overall, with respect to the effects of presentation format on the reliability of conjoint-analysis results, we conclude that only at the level of the raw
data statistically significant differences exist between both conditions. In this instance, as expected, the textual presentation format provided more reliable results than the pictorial presentation format. Nevertheless, since the same effect was not observed with the actual conjoint-analysis results (i.e. the part-worth vectors and the Importance vectors), we did not find enough evidence to accept Hypothesis 8.

9.1.3. The effects of product-category knowledge

The effects of product-category knowledge were tested by taking Experiment 2 as a starting point. The retest was performed eight months after the test. Thirty-six respondents, who were in the schematic three-dimensional conditions were reinvited to the laboratory. This group of respondents was equally subdivided into three groups on the basis of their scores on product-category knowledge in Experiment 2.

The hypothesis that was tested was the following:

H9: Respondents with a lot of product-category knowledge will produce more reliable results than respondents with little product-category knowledge.

Results from the raw data

Table 15: Test-retest correlations for the raw data as a function of product-category knowledge (N=36)

<table>
<thead>
<tr>
<th>Product-category knowledge</th>
<th>Little (n=12)</th>
<th>Moderate (n=12)</th>
<th>Large (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SD</td>
<td>.34 .22</td>
<td>.63 .18</td>
<td>.60 .16</td>
</tr>
<tr>
<td>df F</td>
<td>2.33 .45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .001

In Table 15, one can see that respondents with moderate and large product-category knowledges on average produced higher correlations at the level of the raw data than did respondents with little product-category knowledge.

Results for the part-worth vectors

Table 16: Test-retest correlations between the part-worth vectors as a function of product-category knowledge (N=36)

<table>
<thead>
<tr>
<th>Product-category knowledge</th>
<th>Little (n=12)</th>
<th>Moderate (n=12)</th>
<th>Large (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SD</td>
<td>.17 .40</td>
<td>.52 .33</td>
<td>.44 .41</td>
</tr>
<tr>
<td>df F</td>
<td>2.33 2.80*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .10
In Table 16, one can see a pattern of results similar to those in Table 15. Respondents with moderate and large product-category knowledges on average produced higher correlations at the level of part-worth vectors than did respondents with little product-category knowledge. Although the differences are not statistically significant, an effect in the expected direction can be observed.

Results for the importance vectors

Table 17: Test-retest correlations between the importance vectors as a function of product-category knowledge (N=36)

<table>
<thead>
<tr>
<th>Product-category knowledge</th>
<th>Little (n=12)</th>
<th>Moderate (n=12)</th>
<th>Large (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>.24</td>
<td>.38</td>
<td>.36</td>
<td>.43</td>
</tr>
</tbody>
</table>

In Table 17, it can be seen that the means of the moderate and the large product-category knowledge groups are higher than that of the little product-category knowledge group. This difference, however, is not statistically significant when applying an ANOVA as an integrated test of the differences. If, however, we use a t-test and focus on the difference between the little-knowledge group and the large-knowledge group, we do find a statistically significant difference (t = 1.29; critical value of t at the .05 level, df 11, is .7).

Because of this finding, because of the effect which was found at the level of the raw data, and because of the effect which was found at the level of the part-worth utilities, we can conclude that Hypothesis 9 can be accepted. Respondents with a lot of product-category knowledge produce more reliable results than respondents with little product-category knowledge.

9.1.4. The effects of data-collection method

With respect to the data-collection method, Hypothesis 10 was formulated to test the reliability of ACA-generated results.

H10: If the ACA data-collection method is used, more reliable results will be generated than if a traditional conjoint analysis data-collection method is used.

Hypothesis 10 was tested by reinviting the respondents from the "without-form" condition of Experiment 4 (test: full-concept ranking), as well as those who participated in Experiment 5 (test: ACA). In Experiment 5, the same procedure was followed as in Experiment 4. The results of the control questionnaire are presented in Table G of Appendix 2. Experiment 4 and Experiment 5 were conducted in the same month (both in December, 1988), as were the retests (both in October 1989). The interval between the test and the retest was ten months. In both experiments, the same attributes and attribute-levels were used.
Due to the fact that ACA is an interactive procedure, and that therefore, in most instances, the concepts differ between the test and the retest, no correlations could be determined for the raw data. We will, therefore, concentrate on the reliability of the part-worth utilities and the importances, for only on these indices could comparisons be made between ACA-generated results and results which were obtained by full-concept ranking.

Results for the part-worth vectors

Table 18: Test-retest correlations between the part-worth vectors as a function of data-collection method (N=40)

<table>
<thead>
<tr>
<th>Data-collection method</th>
<th>ACA (n=20)</th>
<th>Full profile ranking (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>.18</td>
<td>.43</td>
<td>.33</td>
</tr>
</tbody>
</table>

In Table 18, it can be seen that the means of the test-retest correlations between the part-worth vectors do not differ greatly between the data-collection methods. The mean correlation is slightly higher in the full-concept ranking condition than that in the ACA condition, in contrast to what we expected.

Results for the importance vectors

Table 19: Test-retest correlations between the importance-vectors as a function of data-collection method (N=40)

<table>
<thead>
<tr>
<th>Data-collection method</th>
<th>ACA (n=20)</th>
<th>Full profile ranking (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>.65</td>
<td>.31</td>
<td>.53</td>
</tr>
</tbody>
</table>

Similar to those of the part-worth vectors, the means of the test-retest correlations between the importance vectors do not differ greatly between the data-collection methods.

Combining the results of Table 18 and Table 19, we can conclude that Hypothesis 10 should be rejected since if ACA is used, the results are not more reliable than those if a traditional conjoint-analysis data-collection method is used.

9.2 An empirical test of the concurrent validity of conjoint-analysis results

With respect to the concurrent validity of conjoint-analysis results, two hypotheses (H11 and H12) were formulated. Both hypotheses will be discussed separately. The independent variables were "presentation format" (H11) and "product-category knowledge" (H12). In Chapter 5, among other issues, we described the operationalization of these variables. The
experimental design and the respondents were discussed in Chapter 6. The operationalization of concurrent validity was discussed in Chapter 8.

9.2.1 The effects of presentation format

Due to the (pre-)structuring of the information in the textual presentation format (as opposed to the pictorial presentation format), it was expected that the respondents would be more logically consistent when confronted with textual stimuli than when confronted with pictorial stimuli.

H11: If the textual presentation format is used, this will yield results which are higher on concurrent validity than those if the pictorial presentation format is used.

Results

Table 20: Goodness-of-fit as a function of presentation format (N=41)

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Textual (n=20)</th>
<th>Pictorial (n=21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SD</td>
<td>Mean SD</td>
<td>df F</td>
</tr>
<tr>
<td>.58 .19</td>
<td>.62 .17</td>
<td>1.39 .53</td>
</tr>
</tbody>
</table>

In Table 20, it is shown that in the textual condition, 58 percent of the variance in the evaluations could be explained by variation in the attribute levels. In the pictorial condition the percentage is 62. This means that there were only small differences in goodness-of-fit between the presentation formats. Consequently, we can conclude that H11 should be rejected. The textual presentation format does not yield results with higher concurrent validity than those from using the pictorial presentation format.

9.2.2 The effects of product-category knowledge

The expectation that, because they are better informed, respondents with moderate and large product-category knowledges are more logically consistent in their evaluations than respondents with little product-category knowledge, was formulated in Hypothesis 12.

H12: Respondents with a lot of product-category knowledge will produce results with higher concurrent validity than those produced by respondents with little product-category knowledge.

Results

Below, Table 21 shows that large differences in the goodness-of-fit existed between respondents with little product-category knowledge on the one hand, and respondents with moderate or large product-category knowledges on the other. Respondents with moderate product-category knowledge and respondents with a large product-category knowledge were logically consistent in their evaluations of the attribute profiles. This can be concluded from the finding that, respectively, 68 and 70 percent of the variance in the evaluations is explained by variation in the attributes. This was only 46% in the case of respondents with little product-category
knowledge. Consequently, Hypothesis 12 can be accepted. Respondents with a large product-category knowledge produce results with higher concurrent validity than those from respondents with little product-category knowledge.

Table 21: Goodness-of-fit as a function of product-category knowledge (N=97)

<table>
<thead>
<tr>
<th>Product-category knowledge</th>
<th>Little (n=33)</th>
<th>Moderate (n=32)</th>
<th>Large (n=32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>.46</td>
<td>.68</td>
<td>.70</td>
</tr>
<tr>
<td>SD</td>
<td>.38</td>
<td>.20</td>
<td>.20</td>
</tr>
<tr>
<td>d</td>
<td>2.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td>7.56*</td>
</tr>
</tbody>
</table>

* p < .001

9.3 An empirical test of the predictive validity of conjoint-analysis results

With respect to the predictive validity of conjoint-analysis results, four hypotheses (H13 to H16) were formulated. Each of these will be discussed separately. The independent variables were as follows: amount of information, presentation format, product-category knowledge, and data-collection method. In Chapter 5, we described the operationalizations of the first three of these variables. Those of predictive validity and of data-collection method were discussed in Chapter 8. The experimental design and the respondents were discussed in Chapter 6.

The results of the actual buying test are presented first. Overall, forty-eight respondents bought a filter coffee-maker after having participated in one of the Experiments (2, 3, 4 or 5). Only if the number of respondents in one of the subgroups was smaller than ten, are the results of the magazine test presented after the results of the actual buying test.

First, attention will be given to the part-worth utilities; second, to the importances. Because of the small numbers of respondents in the subgroups of the actual buying test, the test of the predictive validity of the importances could only be conducted with respect to the magazine test.

In the magazine test of Experiment 2, we only made use of the attributes “form”, “price” and “thermos flask”. The other attributes, i.e. high speed, a removable water reservoir, and a dripstop, did not vary enough in the magazines to establish differences. The level “present” of the attributes “high speed” and “removable water reservoir” was not reflected in the alternatives in the magazines, nor was the level “absent” of the attribute “dripstop”. In Experiment 3, we only concentrated on the attribute “form”. In the Experiments 4 and 5, the attributes “price”, “dripstop”, “capacity” and “thermos flask” were included in the magazine test.

9.3.1 The effects of amount of information

In Experiment 4 (see Section 6.4), Hypothesis 3 was tested. It was found that ambiguous information on the attribute “form” affected the evaluations of the attribute “price”. Since we considered this finding to be the result of uncertainty in the respondent when s/he was confronted with early product-concepts. We expected that this strategy for coping with uncertainty (i.e. preferring moderate or high prices to low prices) would not appear in a real-
choice situation in which all information on the relevant attributes was present. To test this expectation, we formulated Hypothesis 13.

H13: If the moderate or high level of the attribute "price" are preferred to low levels of this attribute (as a result of ambiguous information on a relevant attribute), the predictive validity of the part-worth utilities of the attribute "price" will be lower than that if the low level is preferred.

In line with the results presented in Chapter 6, and in order to test H13, a comparison was made between the results of the schematic three-dimensional condition and the realistic three-dimensional condition of Experiment 2. Since only seven respondents in Experiment 2 actually bought a new filter coffee-maker after having participated in the experiment, H13 could only be tested on the basis of the results of the magazine test.

Results of the magazine test

Table 22: Percentages of respondents with matches on the attribute "price" as a function of amount of information about "form" at the different most-preferred levels of "price". The numbers of respondents preferring a particular level of price in the conjoint task are given first (N=178).

<table>
<thead>
<tr>
<th>Amount of information about form</th>
<th>No information</th>
<th>Some information</th>
<th>All information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>schematic</td>
<td>pictorial</td>
<td>realistic</td>
</tr>
<tr>
<td>Level</td>
<td>No info. (n=41)</td>
<td>Some info. (n=20)</td>
<td>Some info. (n=46)</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>61%</td>
<td>55%</td>
<td>47%</td>
</tr>
<tr>
<td>moderate</td>
<td>8</td>
<td>25%</td>
<td>57%</td>
</tr>
<tr>
<td>high</td>
<td>15</td>
<td>53%</td>
<td>100%</td>
</tr>
<tr>
<td>Total</td>
<td>51%</td>
<td>60%</td>
<td>38%</td>
</tr>
</tbody>
</table>

In Table 22, one can see that the number of matches for the attribute "price" was largest in the textual condition (60%) and smallest in the three-dimensional condition (26%). At first sight, it is striking to note that if no or little (e.g. textual) information is supplied on the attribute "form", the number of matches for the attribute "price" appears to be somewhat higher than that if a lot of information on this attribute is presented (schematic three-dimensional or realistic three-dimensional). These differences are statistically significant (the corresponding z-values varying between 1.99 and 2.68). The reason for this effect, which is in the direction opposite to what we expected, may be that if one has a clear idea of how the product will appear (in the case of a lot of information), a low price is preferred to a high price. To buy a filter coffee-maker on the market with the desired form, however, one has to pay a moderate to high price.

When concentrating on the schematic three-dimensional condition and the realistic three-dimensional condition, one can see that in both conditions the number of matches for the attribute "price" was low. This finding is in contrast to what we expected. In Table 22, it can be seen that about as
many respondents who preferred a moderate or a high price in the schematic pictorial condition as those in the realistic three-dimensional condition actually chose a filter coffee-maker at this price level (38% and 30% respectively). This means that there is no clear relationship between the number of respondents who prefer a moderate or a high price to a low price and the predictive validity of the corresponding part-worth utilities. Therefore, we have to reject Hypothesis 13. To this conclusion, we add a second conclusion which states that if the information on the attribute "form" becomes more realistic, the predictive validity of the part-worth utilities of the attribute "price" decreases.

9.3.2 The effects of presentation format

Due to the greater degree of realism of the information in the pictorial presentation format (as opposed to the textual presentation format), it was expected that the results from the pictorial stimuli would have a higher predictive validity than those from the textual stimuli, following the principle of compatibility of Fishbein and Ajzen (1977).

\[ H14: \text{If the pictorial presentation format is used, results with higher predictive validity will be obtained than if the textual presentation format is used.} \]

Hypothesis 14 was tested using the actual buying test and the magazine test. In these tests, we only concentrated on the predictive validity of the part-worth utilities. To provide some additional insight, however, we will also present some findings on the predictive validity of the importance.

Results for part-worths in the actual buying test

Table 23: Numbers of matches for the attribute "form" as a function of presentation format (N=12)

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Textual (n=7)</th>
<th>Pictorial (n=5)</th>
<th>df</th>
<th>Chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (14%)</td>
<td>4 (80%)</td>
<td>1</td>
<td>5.18*</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

In Table 23, one can see that in the pictorial condition the number of matches for the attribute "form" is substantially larger than that in the textual condition. Because of the small number of respondents who actually bought a new filter coffee-maker after participating in Experiment 3, however, we have to be very careful in drawing conclusions on the basis of these findings.
Results for part-worths in the magazine test

Table 24: Numbers of matches for the attribute “form” as a function of presentation format (N=41)

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Textual (n=20)</th>
<th>Pictorial (n=21)</th>
<th>df</th>
<th>Chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9 (45%)</td>
<td>9 (43%)</td>
<td>1</td>
<td>.02</td>
</tr>
</tbody>
</table>

In Table 24, the numbers of matches are identical in both conditions. The percentages are also almost the same. In both conditions, almost half of the respondents chose a filter coffee-maker from the magazines, which had a form corresponding with the level of that attribute with the largest positive part-worth utility.

When considering the findings presented in Tables 23 and 24, one has to be careful in drawing conclusions about the predictive validity of part-worth utilities generated with different presentation formats. On the one hand, however, the results of the actual buying test are more important than those of the magazine test. On the other hand, the number of respondents in the actual buying test was very small. Because of the latter, and because of the fact that the results of the magazine test did not support the findings of the actual buying test, we conclude that there is not enough reason to confirm that the pictorial presentation format generates part-worth utilities with higher predictive validity than those from using the textual presentation format.

Therefore, we reject Hypothesis 14.

Results for the importances of the magazine test

Table 25: Means of importances for the attribute “form” as a function of matches and mismatches (N=41)

<table>
<thead>
<tr>
<th>Matches and mismatches</th>
<th>Match (n=18)</th>
<th>Mismatch (n=23)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>.87</td>
<td>.69</td>
</tr>
</tbody>
</table>

The means of the importances for the attribute “form” are only slightly higher for those respondents who showed matches on that attribute than they are for those who did not. This means that those respondents who strongly favoured one of the attribute-levels did not chose a filter coffee-maker having this form more often than respondents who only weakly favoured one of the levels of form.

9.3.3 The effects of product-category knowledge

In Chapter 7, the expectation was expressed that respondents with a lot of product-category knowledge would generate more reliable results than
respondents with little product category knowledge, (and which also have greater concurrent validity). As a result, these respondents were expected to generate results which have greater predictive validity than those from respondents who are less well informed.

**H15: Respondents with a lot of product-category knowledge will produce results which have greater predictive validity than those from respondents with little product-category knowledge**

Since only seven respondents in Experiment 2 actually bought a new filter coffee-maker after participating in the experiment, the test of H15 was only tested using the results of the magazine test. To provide some additional insight, we also present some findings on the predictive validity of the importances.

**Results for the part-worths in the magazine test**

In the magazine, test we made use of the attributes "form", "price" and "thermos flask". The numbers of matches for these attributes are shown in Table 26.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Little (n=33)</th>
<th>Moderate (n=31)</th>
<th>Large (n=32)</th>
<th>df</th>
<th>Chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>form</td>
<td>10 (30%)</td>
<td>7 (23%)</td>
<td>14 (44%)</td>
<td>2</td>
<td>3.32</td>
</tr>
<tr>
<td>price</td>
<td>8 (24%)</td>
<td>11 (34%)</td>
<td>8 (25%)</td>
<td>2</td>
<td>1.02</td>
</tr>
<tr>
<td>thermos flask</td>
<td>18 (55%)</td>
<td>13 (41%)</td>
<td>10 (31%)</td>
<td>2</td>
<td>3.67</td>
</tr>
</tbody>
</table>

Table 26 shows that there were no statistically significant differences between the numbers of matches at the different levels of product-category knowledge. With respect to the attribute "form", the respondents in the large-knowledge group matched somewhat more than both other groups. Contrasting outcomes are found with respect to the attribute "thermos flask".

**Results for importances in the magazine test**

Below in Table 27 it is shown that with respect to the attributes "form" and "thermos flask", the respondents with matches for these attributes found these attributes significantly more important than did the respondents with mismatches for these attributes. These findings indicate that the importances had predictive validity.

On the basis of the findings presented in Table 26, we conclude that Hypothesis 15 should be rejected. Respondents with a large product-category knowledge do not produce results which have greater predictive validity than those from respondents with little product-category knowledge.
Table 27: Means of importances of the attributes form, price, and thermosflask as a function of matches and mismatches (N=97)

<table>
<thead>
<tr>
<th>Matches and mismatches</th>
<th>Match (n=32)</th>
<th>Mismatch (n=65)</th>
</tr>
</thead>
<tbody>
<tr>
<td>form</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>1.30 (.84)</td>
<td>.93 (.63)</td>
</tr>
<tr>
<td>price</td>
<td>(n=27)</td>
<td>(n=70)</td>
</tr>
<tr>
<td></td>
<td>.82 (.56)</td>
<td>.63 (.39)</td>
</tr>
<tr>
<td>thermos flask</td>
<td>(n=41)</td>
<td>(n=56)</td>
</tr>
<tr>
<td></td>
<td>.85 (.49)</td>
<td>.62 (.40)</td>
</tr>
</tbody>
</table>

\* \(p < .05\)

9.3.4 The effects of data-collection method

In Chapter 7, the expectation was expressed that ACA, because of being interactive and therefore more meaningful and less demanding for the respondents, would generate more reliable results than traditional data-collection methods, (and which would also have greater concurrent validity). As a result, ACA is also expected to generate results which have higher predictive validity than those using traditional data-collection methods.

\textit{H16: If the ACA data-collection method is used, this will yield results with higher predictive validity than those if a traditional conjoint-analysis data-collection method is used.}

Since, in our opinion, enough respondents actually bought a filter coffee-maker after participating in Experiment 4 (full-concept ranking) and in Experiment 5 (ACA), Hypothesis 16 was tested by using the actual buying test. The numbers of respondents were 15 and 10, respectively. To provide some additional insight, we also present some findings about the predictive validity of the importances.

Results for the part-worths of the actual buying test

Table 28: Numbers of matches for the attributes "price", "dripstop", "capacity", and "thermos flask" as a function of data-collection method (N=25)

<table>
<thead>
<tr>
<th>Data-collection method</th>
<th>Attribute</th>
<th>ACA (n=10)</th>
<th>Full-profile ranking (n=15)</th>
<th>df</th>
<th>Chi-square$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>price</td>
<td>1 (10%)</td>
<td>2 (14%)$^3$</td>
<td>1</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>dripstop</td>
<td>6 (60%)</td>
<td>8 (53%)</td>
<td>1</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>capacity</td>
<td>6 (60%)</td>
<td>11 (73%)</td>
<td>1</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>thermos flask</td>
<td>7 (70%)</td>
<td>7 (47%)</td>
<td>1</td>
<td>1.33</td>
</tr>
</tbody>
</table>

$^2$ All cells have an expected frequency smaller than 5.

$^3$ Only fourteen respondents preferred a particular level of price.
Table 28 shows that there are no statistically significant differences between the two data-collection methods with respect to the numbers of matches for any of the attributes. Table 28 also shows that the numbers of matches for the attribute "price" were much lower than those for the other attributes.

Results for the importances in the actual buying test

Table 29: Means of importances for the attributes "dripstop", "capacity", and "thermos flask" as a function of matches and mismatches (N=25)

<table>
<thead>
<tr>
<th>Matches and mismatches</th>
<th>Match (n=14)</th>
<th>Mismatch (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>dripstop</td>
<td>1.68</td>
<td>1.29</td>
</tr>
<tr>
<td>capacity</td>
<td>1.61</td>
<td>1.14</td>
</tr>
<tr>
<td>thermosflask</td>
<td>1.26</td>
<td>.76</td>
</tr>
</tbody>
</table>

Table 29 shows that for each attribute there were no statistically significant differences between the respondents who bought products with the attribute levels corresponding to those most preferred in the concept test (matches) and those who did not. This means that strong attribute-level preferences, assessed by using either ACA or full-concept ranking, are not reflected more often in purchases of new products than weak attribute-level preferences.

Regarding the findings in Table 28, we conclude concerning Hypothesis 16, that this hypothesis should be rejected. ACA does not yield results with higher predictive validity than those from full-concept ranking.

9.4 Summary and conclusions

In this chapter, we presented the results of the tests of hypotheses about the effects on the reliability and the validity of conjoint-analysis results of the following variables: "amount of information", "presentation format", "product-category knowledge", and "data-collection method". We will first summarize the findings and then discuss the conclusions.

Reliability

With respect to the reliability of conjoint-analysis results, four hypotheses (H7 to H10) were tested. Regarding amount of information, we observed that large numbers of changes in the most preferred level of price occurred if the information on the attribute "form" was absent ("no information" condition: 42% of changes) and if the information on the attribute "form" was presented by means of schematic three-dimensional stimuli (in the schematic three-dimensional condition: 50% of changes). If the information on the attribute "form" was presented textually or pictorially, most respondents preferred the same attribute-level in both the test and the retest (25% of changes).
On the basis of these findings, we concluded that the variable "ambiguity with respect to a relevant attribute" did not cause the variation in the number of changes in the most preferred level of price. As a result, Hypothesis 7 was rejected.

With respect to presentation format, we observed that the test-retest correlations of the evaluations (raw data) were significantly higher when the textual presentation format was used. Presentation format did not have an effect on the part-worth vectors or on the importance vectors. This means that the textual presentation format did not generate more reliable conjoint-analysis results than those from using the pictorial presentation format. Consequently, Hypothesis 8 was rejected.

Respondents with a lot of product-category knowledge and with moderate product-category knowledge produced results with higher correlations at the level of the raw data than respondents with little product-category knowledge. Also, respondents with moderate and large product-category knowledge produced somewhat higher correlations at the level of the part-worth vectors, and at the level of the importance vectors, than respondents with little product-category knowledge. Although, the last two differences were not found to be statistically significant at the .05 level, the effect was clearly in the expected direction. Moreover, when concentrating on the little-knowledge group and the large-knowledge group, statistically significant differences were found in the test-retest correlations of the importance vectors. We, therefore, accepted Hypothesis 9. Respondents with large and moderate product-category knowledge produce more reliable results than respondents with little product-category knowledge.

The means of the test-retest correlations of neither the part-worth vectors nor the importance vectors differed much between the two data-collection methods (ACA and full-concept ranking). Thus, Hypothesis 10 was rejected. ACA does not yield more reliable results than those from using a traditional data-collection method.

Concurrent validity

With respect to the concurrent validity of conjoint-analysis results, two hypotheses (H11 and H12) were formulated. The effect of presentation format on the goodness-of-fit was found to be very small since in the textual and in the pictorial conditions, similar values were obtained. This implied that the textual presentation format does not yield results with higher concurrent validity than those from using the pictorial presentation format. Therefore, we rejected H11.

With respect to the variable "product-category knowledge", however, differences were found. The goodness-of-fit was found to be high for high and moderate levels of product-category knowledge and low in the case of low levels of product-category knowledge. Consequently, Hypothesis 12 was accepted. Respondents with large, and moderate, product-category knowledge produce results with higher concurrent validity than those from respondents with little product-category knowledge.

Predictive validity

With respect to the predictive validity of conjoint-analysis results, four hypotheses (H13 to H16) were tested. Regarding the effect of amount of information, we observed that if no or little (e.g. textual) information was supplied for the attribute "form", the number of matches for the attribute "price" were somewhat higher than that if a lot of information on this attribute
was presented (either schematic three-dimensional or realistic three-dimensional information). Nevertheless, Hypothesis 13 was rejected because of the minimal differences between the numbers of matches in the schematic three-dimensional condition (with ambiguous information on the attribute "form" and in which many respondents preferred a moderate or a high price to a low one), and those in the realistic three-dimensional condition (in which all information on the attribute "form" was presented and in which few respondents preferred a moderate or a high level of price to a low one). If the moderate or the high levels of the attribute "price" are preferred, the predictive validity of the attribute "price" is as high as that when a low level is preferred. Related to this, we found that if little or no information was presented about the attribute "form", the predictive validity of the attribute "price" was higher than that if a lot of information was given about this attribute.

Presentation format proved to have an effect on the predictive validity of the part-worths of the attribute "form" in the actual buying test. In the pictorial condition, the number of matches for the attribute "form" was substantially larger than that in the textual condition. In this test, however, the number of respondents was small. In addition, these findings were not confirmed in the magazine test for the part-worth utilities, nor for the importances. We, therefore, rejected Hypothesis 14. The pictorial presentation format does not generate results with higher predictive validity than those from using the textual presentation format.

To test the effects of product-category knowledge on the predictive validity of conjoint-analysis results, the magazine test was applied. In this test, no statistically significant differences were found between the numbers of matches at the three different levels of product-category knowledge. Consequently, we concluded that Hypothesis 15 should be rejected. Respondents with a lot of product-category knowledge do not produce results with higher predictive validity than those from respondents with little product-category knowledge.

With respect to data-collection method, no statistically significant differences were found between the numbers of matches for any of the attributes. Thus, Hypothesis 16 was also rejected. ACA does not generate results with higher predictive validity than those from using a traditional data-collection method.

Conclusions

In Chapter 3, the second research question was formulated. The second research question concerned the reliability and the validity of evaluations based on early product-concepts. With respect to this second research question some conclusions can be drawn:

1. In this study, the most important factor influencing reliability appears to be product-category knowledge. Respondents with moderate and high levels of product-category knowledge generate more reliable conjoint-analysis results than do respondents with low levels of product-category knowledge. The effect of product category is very strong with respect to direct evaluations. For these, respondents with moderate and large product-category knowledges generate much higher test-retest correlations than do respondents with little product-category knowledge. Similar effects of product-category knowledge, although not as strong as those on direct evaluations, can be found on the reliability of part-wort utilities and importances. On the basis of these aforementioned findings,
we draw the conclusion that in order to obtain reliable evaluations of early-product concepts, respondents should have at least a moderate level of knowledge of the product category. The variables “amount of information”, “presentation format”, and “data-collection method” do not seem to have a direct effect on the reliability of conjoint-analysis results. With respect to amount of information, nevertheless, we found that the number of attributes in the product-concepts may be an important factor in explaining the variance in the reliability of the part-worths of the attribute “price”. In the case where there were few attributes in the profiles (in the textual and the pictorial conditions), the number of changes in the most preferred level of the attribute “price” was smaller than in the case where there were many attributes in the profiles (no information condition and schematic three-dimensional condition).

With respect to presentation format, we found that the effects which were present at the level of the direct evaluations disappeared after performing a conjoint analysis on these raw data. This means that presentation format is much more important when interpreting direct evaluations (e.g. in the case of concept screening) than when interpreting conjoint-analysis results. On the basis of these findings, no advice can be given as to which presentation format is best used for conjoint analysis. Data-collection method does not affect the reliability of the results and in this sense there appears to be little reason for preferring one method over another.

2. Concurrent validity is affected by product-category knowledge but not by presentation format. Respondents with moderate and high levels of product-category knowledge generate the most concurrently valid data. From this, we can conclude that respondents with large and moderate product-category knowledges are more internally consistent with their evaluations of early product-concepts than are respondents with little product-category knowledge. Consequently, in order to obtain results with high concurrent validity, it is necessary to invite respondents with at least a moderate level of product-category knowledge. The fact that no differences in concurrent validity were found between the textual and the pictorial presentation formats implies that, on these grounds, no clear preference for one format over the other can be specified.

3. None of the variables that were selected proved to have a strong influence on the predictive validity of conjoint-analysis results. The amount of information about the attribute “form” did affect the evaluations of the attribute “price”, but not affect the predictive validity of the part-worth utilities of this attribute. From this, we conclude that the finding that “price” is sometimes used to reduce uncertainty in a difficult evaluation task does not necessarily have implications for the predictive validity of evaluations of that attribute. Related to this, we found that if little or no information is presented about the attribute “form”, the predictive validity of the attribute “price” is higher than that if a lot of information is given about that attribute. It is possible that if a lot of information is given about a relevant attribute, respondents concentrate on that attribute and this may lead to the neglect of another attribute. If little information is given about a relevant attribute, the respondents concentrate on other attributes such as “price”.

The variable “product-category knowledge”, in spite of its effects on the reliability and on the concurrent validity of conjoint-analysis results, did
not have an effect on the predictive validity of those results. This is remarkable since reliability and concurrent validity are preconditions for obtaining results with predictive validity. Explanations for this finding may include the large number of attributes in the profiles (6), and the fact that the attribute "form" was presented in isolation from the cards with the textual information about the attributes. Possibly, the respondents with a lot of product-category knowledge focused on this attribute most, as is indicated by a relatively large number of matches for this attribute. We add to this that the respondents in the large-knowledge group were younger than the respondents in the little-knowledge group, and were therefore, on average, more strongly interested in the form of the product than the older respondents. The (older) respondents with little product-category knowledge, by contrast, may have concentrated more on the textual information.

ACA, the promising new data-collection method, does not appear to outperform full-concept ranking in terms of the predictive validity of the results. Similar findings were found concerning the reliability of ACA-generated results. Nevertheless, on the basis the practical advantages of ACA, we prefer ACA when no pictorial or three-dimensional information on the form of a new product-concept can be given. In all other instances, one could rely on traditional data-collection methods.

No clear results were found with respect to the predictive validity of importances. In one instance, the importances appeared to score high on predictive validity (Experiment 2), whereas in two other instances they did not (Experiment 3, and the combination of Experiment 4 and 5). We conclude from this, that attribute levels with large part-worth utilities are not represented in future-selected products more often than attribute-levels with small part-worth utilities.

When considering the findings on the predictive validity of part-worth utilities and importances it can be stated that the predictive validity of conjoint-analysis results is rather low. In our opinion, this can for the largest part be explained by the low correspondence between different aspects of the conjoint task and actual buying behavior. In the next chapter we will discuss this matter in more detail.
10. General Discussion

10.1 Introduction

As we stated in the introduction to this thesis, the crucial decision in new-product development is that of deciding which characteristics the new product should have. Consumer research can supply useful information to support this decision. The application of much of the consumer research in the early stages of the new-product development process, however, is based on the assumption that consumers are capable of evaluating early product-concepts. The question as to whether or not this assumption holds has been the starting point for this investigation. Two research questions have been formulated. The first research question (i.e. in what respects are early product-concepts evaluated differently from actual products?) is answered in the Chapters 3 to 6; the second research question (how reliable and how valid are evaluations of early product-concepts) is answered, with the use of conjoint analysis, in the Chapters 7 to 9.

The findings with respect to both research questions have both theoretical and practical implications. These implications will be discussed in the following sections.

10.2 Theoretical Implications

A distinction is made between implications regarding the proposed Model of Product-Concept Evaluation and implications of the consumer-research technique of "conjoint-analysis". Both types of implications will be discussed.

10.2.1 Theoretical implications of the Model of Product-Concept Evaluation

The model of Product-Concept Evaluation is introduced in Chapter 4. The elements of this model are: product-information, processing, integrating, consumer characteristics, and evaluative response. For the sake of clarity, we will depict this model here once again.

```
  ┌────────────────────────┐
  │ Consumer Characteristics │
  │                        │
  ├───────────────┬───────────────┤
  │Product        │ Processing   │ Integrating │ Evaluative Response │
  │Information    │             │             │                    │
  └───────────────┴───────────────┘
```

Figure 30: Model of Product-Concept Evaluation

The attention for the processing of product information is largely based on Holbrook (1981), and that for consumer characteristics on Finn (1985). The processing of product information covers the level of processing and the type of processing. Consumer characteristics is comprised by product-category knowledge.

In this study, we made use of experiments to study the elements of the model. Experiments offer the advantage of experimental control. As a result, causal relations can be assessed very carefully. To guarantee the external
validity of the findings, a number of precautions were taken. The respondents were given the impression that the goal of the investigation was to construct a better filter coffee-maker. Also, "ordinary" consumers instead of students were invited as respondents.

Nevertheless, with respect to our study, some points of criticism should also be mentioned. The conclusions of this study are only based on one product, the filter coffee-maker. Because we have not investigated other products we have to be careful in generalizing our conclusions to other products. Moreover, the numbers of attributes varied in the various experiments. Because of this, it is difficult to make comparisons between the findings of these experiments.

*Level of processing*

From the results, it can be concluded that the level of processing of product information is affected by the amount of information (amount of detail on the attribute "form"). Early product-concepts (which only have little information) are processed at a more shallow level than are actual products. As a consequence, the product information in early product-concepts is often not comprehended and not elaborated upon. Nevertheless, because the product-concepts were given to the respondents, it is expected that the respondents at least have reached focal attention in processing the information in the concepts. In this respect, there is a major difference with advertising (the field from which the communication model of Greenwald and Leavitt originated). In product-concept testing, the task guarantee that respondents take notice of information in the concepts. In advertising, by contrast, a lot of effort (money) has to be allocated to receiving initial attention from potential buyers.

Related to this, it was observed that different evaluative responses are obtained if small amounts of information are available on a given level of the attribute "form", than if a lot of information is available on this attribute level.

If no, or ambiguous, information is present on a relevant attribute (e.g. form), even respondents with a lot of product-category knowledge may feel uncertain about rating the concepts. In those instances, price is used as an indicator of the level of this attribute. This finding supports earlier findings such as those of Gabor and Granger (1966).

Due to pragmatic considerations, the operationalization of amount of information was only put through in the pictorial presentation format. It would have been interesting to have studied this variable also within the textual and the three-dimensional presentation formats.

*Type of processing*

The type of processing of the information in early product-concepts is not affected by presentation format. Both pictorial concepts (which include pictorial and textual information) and textual concepts (which include textual information only) are processed analytically. Comparing these findings to those of Holbrook and Moore (1981), it can be deduced that the textual information in our pictorial concepts has given rise to an analytical type of processing. In terms of Chaiken (1980) a systematic processing strategy, and in terms of Sujan (1985) a piecemeal processing, is followed. As a consequence, one can conclude that possible differences in evaluative responses between both presentation formats, do not result from a different type of processing. In addition, an important implication of the finding that the processing of the information in early product-concepts is analytical and
not holistic, is that our choice for the linear additive model, (and thereby for the use of fractional-factorial designs instead of full-factorial designs), in the Experiments 2, 4 and 5 is empirically supported.

At first sight, the finding that early product-concepts are processed at a shallow level conflicts with the finding that this information is processed analytically. This contradiction can be explained by the fact that in the test on the level of processing, strictly pictorial stimuli were used, whereas in the test on the type of processing, the pictorial stimuli also contained textual information. Possibly, in the test on the type of processing, those respondents who were incapable of comprehending the pictorial information elaborated upon the textually presented attributes. This may explain why, as can be seen in Figure 26, the part-worth utilities of the attribute "form" were less pronounced in the pictorial condition than in the textual condition.

Another explanation is that in the test on the level of processing, the consumers were invited to verbalize what they saw, whereas the test on the type of processing consisted of an evaluation task. In an evaluation task in which the concepts consist of small numbers of attributes respondents probably are more inclined to systematically process and evaluate the information in the concepts than in a verbalization task. This suggestion is supported by Johnson & Russo (1981; 1984) who find that respondents in an evaluation task systematically process the information in each of the attributes, whereas in (the beginning of) a choice task respondents only concentrate on one or two attributes. Consequently, an evaluation task, therefore, provides the researcher with an insight in the desirability of all attributes included. In a choice task, by contrast, only particular attributes receive full attention. In this respect there are major differences between the actions in the conjoint task and actual buying behavior (see Ajzen & Fishbein, 1977).

**Consumer characteristics**

The finding that amount of information on an attribute affects the evaluative responses does not hold if consumers have a moderate or a high level of product-category knowledge. In those cases, respondents do comprehend the information in the early product-concepts and are capable of elaborating upon this information. Correspondingly, identical evaluative responses are observed of the same attribute level presented with different amounts of information. In Finn's (1985) terminology, respondents with a lot of product-category knowledge are capable to estimate the amount of utility which is supplied by a particular attribute level or concept.

In contrast to what has been found in the literature, our measure of product-category knowledge does not correlate with recency and frequency of buying, nor with frequency of use. The fact that we did find statistically significant correlations with age and with level of education, possibly implies that, to a certain extent, the trained capacity to deal with school-like tasks determines both the completion of the questionnaire (measure of product-category knowledge) and the evaluative responses of the early product-concepts. On the other hand, it can be argued that younger respondents (who are higher educated because of improvements in the school system in the last decades) interact more intensively with new products than older respondents, and therefore, in spite of having little buying experience and a low frequency of use, have much product-category knowledge.

In conclusion, it is shown that the processing of product information and consumer characteristics are important variables in the evaluation process
of early product-concepts. By broadening the Two-Stage Model presented by Holbrook (1981) with more general notions on information processing in the Model of Product-Concept Evaluation, and by adding consumer characteristics a psychological explanation is given for the rule of thumb which states that consumers co-operating in the development of new products need to have an orientation with respect to the particular product (e.g. Von Hippel, 1978; Wind, 1982; Mantel & Meredith, 1986; Holt, 1989). So the edition of consumer characteristics to the Two-Stage Model signifies an improvement to the original model.

10.2.2 Theoretical implications for conjoint analysis

In this investigation, conjoint analysis was chosen to measure the evaluative responses. It can be argued that conjoint analysis not only measures responses, but that its strength lies in linking the evaluative responses to particular attributes and attribute-levels. By determining part-worth utilities and importances, an insight is obtained into the way in which the attribute-levels presented in the product-concepts are integrated into the evaluative responses.

Conjoint analysis was selected both for practical and theoretical reasons. As was shown in academic literature on this technique, many studies are presented which focus on the reliability and/or the validity of its results. In reviewing these studies, at least two aspects appear to be striking. Firstly, several points of criticism can be mentioned regarding methodology: (i) students were invited as respondents, (ii) in assessing the reliability of the conjoint-analysis results, unrealistically short interval periods were taken into account, (iii) attribute-level preferences were not related to real-life buying behavior. As a result, often positive, but not very generalizable, conclusions were drawn about the reliability and the validity of conjoint-analysis results.

Secondly, most attention was given to the model of preference, to the data-collection method, and to the estimation method. Stimulus presentation and individual differences, by contrast, hardly received any attention until now.

In our study, after testing the hypotheses about the evaluation of early product-concepts, however, it was concluded that the effects of these factors (amount of information, presentation format, and consumer characteristics) on the reliability and the validity of conjoint-analysis results had to be tested. In assessing the reliability of the conjoint-analysis results, "ordinary" consumers were invited and realistic interval periods were taken into account. In assessing the predictive validity, the experimental results about the attribute-level preferences were related to real-life buying behavior. By taking these precautions, a serious attempt was made to close the gap between academic research and the practice of new-product development. The findings of our study, compared to those of most other academic studies, can therefore be more easily generalized to new-product development.

On the basis of the findings on the reliability and the concurrent validity of conjoint-analysis results, we concluded that respondents with moderate and large product-category knowledges are capable of evaluating early product-concepts, for those respondents generate reliable results which are also high on concurrent validity. We draw these conclusions with some reservation. Again, in these tests, the trained capacity to deal with school-like tasks (as an antecedent of product-category knowledge) may have
influenced the stability and the internal consistency of the evaluations. In addition, it can be stated that reliability is a difficult concept in the area of new-product development. During the considerable lead times, consumer preferences may change. When, as a result, other evaluations are obtained in the retest than in the test, it would not be justified to conclude that the results of the test necessarily contained a substantial amount of error. What can be concluded from the finding that respondents with a lot of product-category knowledge generate more reliable conjoint-analysis results than respondents with little product-category knowledge, is that respondents with a lot of product-category knowledge have stable preferences and that the evaluative responses of this group contain little error.

When considering the results on the predictive validity of part-worth utilities and of importances, one may be inclined to conclude that the insights obtained by conjoint analysis are of little value. The fact, that even in the case of respondents with a lot of product-category knowledge, the findings with respect to the predictive validity are not so positive, can, among others, be explained by the following reasons:

1. The period for testing the hypotheses on the predictive validity was too short. Only few respondents had bought a new filter-coffee maker after participating in one of the experiments. Because of this, we only got little data to test our hypotheses on the predictive validity. If the lead time of a doctoral-research project had been longer, more respondents would have bought a new drip coffee maker after having participated in one of the experiments.

2. The respondents which we invited were just potential buyers of drip coffee makers. Only a small group of these respondents proved to be prospective buyers.

3. The correspondence on the different aspects between the conjoint task and actual buying behavior is very low (see Ajzen & Fishbein, 1977). Major differences exist with respect to action, target, context, and time.

4. A large number of external factors influence the actual buying behavior. The variables in the Model of Product-Concept Evaluation are only part of the factors which play a role (e.g. situational factors, advertising; see Wilkie, 1990).

When taking a closer look at the third and the fourth reason, in our opinion, it can be stated that our ambitions to relate evaluations of early product concepts to actual buying behavior were rather optimistic. To be able to at least find an effect of one of the main variables on the predictive validity a much larger sample and a larger testing period should be used. This, however, would have major practical and financial implications. One way to solve this dilemma is to develop alternative dependent variables (or criteria) such as the magazine test.

In conclusion, the findings of our study indicate that the effort in studies on the reliability and the validity of conjoint-analysis results is not effectively allocated. The implication is that more effort should be invested in gaining an insight into how individuals with different characteristics process information in early product-concepts, relative to gaining an insight in how the processed information is integrated into an evaluative response. In this respect, it can be stated that the Model of Product-Concept Evaluation offers a useful starting point. In this model, it is shown that, when investigating the integration processed information, the reliability or the validity of evaluative responses, attention should be paid to the processing
of product information and consumer characteristics. It goes without saying that much attention should be given to the methodological aspects we have mentioned.

10.3 Suggestions for further research

The Model of Product-Concept Evaluation needs more empirical testing. In particular, research is needed on the effects of consumer characteristics on evaluations. Many characteristics can be thought of apart from product-category knowledge. In De Bont et al. (1992) the relevance of tolerance of ambiguity and categorization width to evaluations of product design was demonstrated. Moreover, in the article by Snelders et al. (1992) the feel/think dimension is related to the evaluation of product-concepts. The introduction of this dimension, which poses an interaction between consumer characteristics and a particular product, is helpful in gaining insight into the evaluations of product-concepts with intangible attributes. A possible consequence of this type of research is that the conviction of many researchers that only tangible attributes can be used in conjoint-analysis (e.g. Stokmans 1991) needs to be revised.

Attention should be given to the link between processing and integrating. In particular, when pictorial or three-dimensional information is supplied to respondents, attempts should be made to determine which percepts have been dominant in forming the evaluation. By gaining a better insight into this process, market researchers can contribute more directly to the new-product development process.

In applications of conjoint analysis, until now, very few studies have been conducted which included the attribute form. Considering the importance which is given to "design" at this moment, in our opinion this is an omission. Even more so because, in the area of computer graphics, enormous advances have been made in the visualization of early product-concepts. Especially, within the pictorial presentation format, representations at many degrees of realism can be generated. In this respect, more research is needed on the reliability and the validity of evaluations which are generated at the different degrees of realism within the pictorial presentation format. In this respect it is important to note that the visualisations which can be constructed with user-friendly software packages at this moment are much more realistic than the ones presented in this thesis.

In order to gain a better insight into the predictive validity of conjoint-analysis results many more studies are needed. In those studies, it is necessary to allow a considerable lead time, or to concentrate on products which are purchased more frequently than filter coffee-makers. In addition, attention will be needed to develop criteria which can be used to circumvent measuring the predictive validity. In these criteria as much as possible aspects of actual buying behavior should be reflected. Before focussing on the predictive validity, however, it may be wise to pay attention to the disregarded construct validity of the results.

10.4 Practical implications

The findings of our study have practical implications for researchers (both academic researchers and market researchers) and for product designers. The implications will be discussed for both categories of stakeholders. In doing so, a distinction is made between the implications from the Model of Product-Concept Evaluation and those from the tests of the reliability and the validity of conjoint-analysis results.
10.4.1 Practical implications for researchers

On the basis of the findings with respect to the Model of Product-Concept Evaluation, a number of implications can be mentioned for the application of consumer research in the early stages of the new-product development process. The most important one is based on the finding that respondents with at least a moderate level of product-category knowledge are capable of evaluating early product concepts. This finding implies with respect to the application of consumer research in the early stages of the new-product development process, that consumer evaluations of early product-concepts can be integrated into the development process if, (and only if), the respondents who are invited at least have a moderate amount of knowledge of the product category. Under that restriction, the assumption that consumers are capable of evaluating early product-concepts, is valid.

It is important to note, that although this implication may seem rather straightforward, the essential finding of this dissertation is that when compared to the other variables of the Model of Product-Concept Evaluation such as amount of information and presentation format, the consumer characteristic of product-category knowledge is much more relevant. In contrast to the other variables, a moderate (or high) level of product-category knowledge is a necessary condition for obtaining reliable and valid conjoint-analysis results.

A close parallel can be drawn to authors such as Holt (1989) and Von Hippel (1978) who state that the applicability of consumer-based techniques in new-product development is only beneficial in situations where the would-be consumer is overtly aware of his product need. Consequently, when conducting consumer research in this early stage of the product-development process, much attention should be paid to the selection of the respondents.

The findings with respect to the reliability and the validity of conjoint-analysis results have the following implications:

1. The respondents who are invited to perform a conjoint-analysis task on a given product should at least have a moderate level of knowledge on that product (and/or on its category).

   The fact that we have chosen a filter-coffee maker, which is a product almost everyone has some knowledge about in the Netherlands, has had the effect that not only respondents with a lot of product-category knowledge, but also respondents with moderate product-category knowledge, generated reliable and concurrently valid results. In the case of products which are less widely used, such as food processors, this will probably be restricted to respondents with much product-category knowledge only.

2. When performing a conjoint-analysis study one has to be careful in interpreting the part-worth utilities of the attribute price.

   Especially when much (unambiguous) information is supplied on the relevant attributes, the predictive validity of the part-worth utilities of price will probably be low. An explanation given for this finding is that, in those instances, respondents give the most positive evaluations to the alternatives which have the most preferred level of the relevant attribute at the lowest price. As a consequence, price appears to be a sensitive attribute. In an actual buying situation, this combination is seldomly offered. In such a situation, consumers have to decide whether or not they will pay an extra amount of money to obtain the most preferred alternative. In that case, the predictive validity of the part-worth utilities of
the attribute "price" appears to be very low. If, by contrast, no, or ambiguous, information is given on the relevant attribute, high or moderate prices are often preferred to low prices. It is not surprising that, in these instances, the part-worth utilities of the attribute "price" often appear to be predictively valid.

A possible way to avoid these problems of the attribute "price", is to exclude this attribute from a first concept-optimization test. In this first test, by means of conjoint-analysis, optimal combinations of attribute levels can be determined, which can be tested in a second concept-screening test in which price is included. Another solutions to this problem is suggested by Huisman (1981). He proposes to use conjoint analysis in such a way that a particular price is added to each attribute-level. This solution, however, should only be applied when there are few attributes.

3. Since presentation format and data-collection method do not affect the reliability or the validity of conjoint-analysis results, it is not possible to supply clearcut advice as to which procedure should be followed.

With respect to the conclusion which was drawn about the effects of presentation format on the predictive validity of the conjoint-analysis results, it can be said that this conclusion is influenced by the small numbers of respondents in the actual buying test. The findings in this small group, nevertheless show that pictorial stimuli generate more predictively valid part-worth utilities than do textual stimuli. Since the degree of realism of pictorial stimuli is higher than that of textual stimuli (if the same amount of detail is presented), the general rule of thumb which states that new-product concepts should be presented as realistically as possible (e.g. Wind, 1982) is in that sense weakly supported. Moreover, since this rule of thumb is not contradicted by the findings either, it may be wise to accept this rule as much as possible.

With respect to data-collection method, it can be stated that ACA is more practical than traditional data-collection techniques. ACA, nevertheless, has the major disadvantage that only textual attribute information can be used. Therefore, ACA needs to be developed further (e.g. by including pictorial information) to become a more balanced data-collection technique for research on new-product development.

We conclude that, if the selection of the respondents is conducted with great care, conjoint-analysis can be a useful technique in the early stages of the new-product development process. The use of this technique can support major product decisions early in the development process at a relatively low cost. Considering that the redesign of existing products is a frequent design activity, conjoint analysis can be applied in many situations.

10.4.2 Practical implications for product designers

A more intensive use of conjoint analysis in new-product development will have implications for product designers, for the use of conjoint analysis requires them to concentrate on the desirability of particular attributes and attribute levels, when re-designing products. Consumers' preference models are to be used as input for further development.

In our opinion, product designers should make more use of these preference models to get a better understanding of consumer preferences. Intuition should not be used for this purpose, but should be employed in generating new-product ideas.
The finding that consumer research can be usefully applied in the early stages of the new-product development process, also requires a different organisation of new-product-development projects. In those projects, consumer researchers and product designers should closely interact. A good example is given by Griffin and Hauser (1991), who demonstrate how "the voice of the customer" can be utilized to modify existing products and services or to create new products and services.

For this moment it is up to the consumer researchers (or up to general management) to convince product designers that it makes sense to use input from consumer research already in the early stages of the development process. Knowledge of part-worth utilities and importances should not be seen as a threat to the creative freedom of the product designer, but as a challenge to his creativity.
Summary

In the introduction to this dissertation, the aim of the study was described. In brief, the aim was to find out whether consumers are capable of evaluating early product-concepts. Only if they are, it makes sense to integrate consumer evaluations into the new-product development process from its early stages. In investigating this matter, two research questions were formulated:

1. In what respects are early product-concepts evaluated differently from actual products?
2. How reliable and how valid are evaluations of early product-concepts?

An attempt was made to define the main differences between early product-concepts and actual products. It was suggested that there are two dimensions on which differences exist: amount of information and presentation format. Early product-concepts consist of less information than actual products, and are often presented in a different presentation format (i.e. textually or pictorially instead of three-dimensionally).

In order to understand how early product-concepts are evaluated, various models of product evaluation are briefly discussed, such as the Model of Information Integration (Anderson, 1981), the Two-Stage Model of Evaluative Judgment (Holbrook, 1981), the Model of Hierarchical Information Integration (Louvriére, 1984), and the Model of the Consumer Evaluation Process for New Products and New-Product Concepts (Finn, 1985). Essential elements from these models, such as the attention for the processing of the product information, and the relevance of consumer characteristics, were combined into a new model, which we refer to as the Model of Product-Concept Evaluations. From this model, a number of hypotheses were constructed about the evaluation of early product-concepts. Testing these hypotheses resulted in the following conclusions:

1. Early product-concepts are processed at a more shallow level than are actual products. This means that the information in early product-concepts is processed at a broad, total-product level, whereas actual products are processed more deeply, at the level of attributes.

2. The small amount of information available on the attribute "form" in early product-concepts does not affect the evaluations of that attribute when respondents have a moderate or a high level of product-category knowledge, but does have an effect on evaluations when respondents only have a low level of product-category knowledge.

3. If information on a relevant attribute is presented ambiguously (as often happens in the case of early product-concepts), respondents are often inclined to prefer moderate or high prices to low prices.

4. Including pictorial information on the attribute "form" in textual concepts does not lead to an holistic type of processing. These concepts are processed in an analytical manner, consisting of a systematic evaluation of each of the single attributes.

5. Presenting information on the attribute "form" pictorially leads to different evaluations from those which are obtained by presenting the same attribute information textually. This means that levels of the attribute "form" may be positively evaluated when they are presented in the
pictorial presentation format, and negatively when they are presented in the textual presentation format (and vice versa).

These findings form the answer to the first research question. Due to the smaller amount of information in early product-concepts, the processing of the product information is affected. In addition, because of the variables "amount of information" and "presentation format", the evaluations are different from those of actual products. The variable "product-category knowledge" seems to have a mediating effect on the evaluations.

From the answer to the first research question, it becomes clear that all variables in the Model of Product-Concept Evaluation are relevant for answering the second research question. Because of this, these variables were included as factors which may influence the reliability and the validity of evaluations of early product-concepts. The first step in answering the second research question was the choice for a technique to measure evaluative responses. Both on theoretical and on practical grounds we decided to use conjoint analysis. Subsequently, relevant literature was reviewed to gain an insight into the reliability and the validity of conjoint-analysis results. Due to the fact, however, that the review of literature provided very little insight into the effects of the relevant variables, empirical research was carried out. The findings from this research are:

1. The reliability of conjoint-analysis results based on early product-concepts is acceptable (but not high) when respondents have a moderate or a high level of product-category knowledge.
   The variables amount of information, presentation format, and data-collection method did not appear to have an effect on the reliability of conjoint-analysis results.

2. The concurrent validity of conjoint-analysis results is acceptable in the case respondents are invited with a moderate or a high level of product category knowledge. In that case, the averages of the goodness of fit are .68 and .70 respectively.
   The variable presentation format did not appear to have an effect on the concurrent validity of conjoint-analysis results.

3. The variables of the Model of Product-Concept Evaluation proved to have hardly any effect on the predictive validity of conjoint-analysis results.
   The percentage of correct predictions (matches) on the part-worth utilities of the attribute form was 45 in the textual and 43 in the pictorial condition. The percentages of matches of the part-worth utilities varied from 24 (little knowledge group, attribute price) to 55 (little knowledge group, attribute thermos flask).

   The predictive validity of the part-worth utilities of the attribute price appeared to be substantially higher if little or no information was given on the attribute form (60% and 50% correct predictions) than if much information was given on that attribute (30% and 26% correct predictions).

   The predictive validity of importances is rather low, since in only one out of three tests, the mean importance of the respondents with matches was higher than that of respondents with mismatches. From this we conclude that attribute-levels with large part-worth utilities are not reflected in the products which are bought more often than attribute-levels with small part-worth utilities.

   Since evaluative responses are operationalized by conjoint-analysis results, the answer to the second research question is that evaluations
based on early product-concepts are reliable, and demonstrate concurrent validity if respondents with moderate or a lot of product category-knowledge are invited. The predictive validity of the evaluations, however, is low. This can be partly explained by the facts that (apart from the product itself) many external factors determine actual buying behavior, only a small number of all respondents had bought a new filter coffee-maker after participating in one of the experiments, and the group of respondents consisted both of prospective and potential buyers.

The implications of the findings are manifold. Both theoretical and practical implications can be mentioned. The most important theoretical implication is that the Model of Product-Concept Evaluation appears to be a valid framework for obtaining insight into the evaluation process of early product-concepts. In this respect, it is shown that the processing of product information and consumer characteristics are important variables. Those variables should receive much more attention in studies on the reliability and/or the validity of consumer-research techniques which are applied to integrate consumer evaluations into the new-product development process.

The most important practical implication of the findings is that the respondents who are invited to evaluate early product-concepts should at least have a moderate knowledge, but preferably a lot of knowledge, about the product category. Only these respondents are capable to evaluate early product-concepts. Therefore, only the evaluations of respondents with a moderate or a lot of product-category knowledge should be used to guide the new-product development process. So, if these respondents are selected, it makes sense to use the outcomes of consumer research from the early stages of the new-product development process. To profit as much as possible from these outcomes, a proper communication between consumer researchers and product designers is required.
Samenvatting

Het doel van dit onderzoek is beschreven in de introductie van dit proefschrift. Kort samengevat bestaat het doel uit het bepalen of consumenten in staat zijn vroege produktconcepten te evalueren. Immers, slechts wanneer consumenten hiertoe in staat zijn, heeft het zin om reeds vanaf het begin van produktontwikkelingstrajecten gebruik te maken van evaluaties van consumenten. Om dit te onderzoeken zijn twee onderzoeksvragen geformuleerd:

1. In welke opzichten worden vroege produktconcepten anders geëvalueerd dan gerede produkten?
2. Hoe betrouwbaar en hoe valide zijn evaluaties die gebaseerd zijn op vroege produktconcepten?

In de uitwerking van deze vragen, is getracht vast te stellen wat nu de belangrijkste verschillen zijn tussen vroege produktconcepten en gerede produkten. Hieruit kwam naar voren dat er twee hoofddimensiones zijn waarin verschillen bestaan: de hoeveelheid informatie en de aanbiedingsvorm. Vroege produktconcepten bevatten minder informatie dan gerede produkten en worden vaak tekstueel of pictoriel aangeboden (in plaats van drie-dimensioneel).


1. Vroege produktconcepten worden op een minder diep niveau verwerkt dan gerede produkten. Dit betekent dat de informatie in de vroege produktconcepten globaal, op het niveau van het totale produkt verwerkt wordt. Gerede produkten, daarentegen, worden wel op een diep niveau verwerkt, hetgeen blijkt uit de aandacht voor de informatie in de attributen.
2. De kleine hoeveelheid informatie over het attribuut vorm die in vroege produktconcepten aanwezig is heeft geen effect op de evaluaties van dat attribuut, mits de respondenten een redelijke hoeveelheid, of veel, kennis van het produktveld hebben. Het effect is wel aanwezig indien de respondenten weinig kennis hebben van het produktveld.
3. Als de informatie over een belangrijk attribuut ambig is, (hetgeen vaak voorkomt bij vroege produktconcepten), zijn respondenten vaak geneigd een hoge prijs te prefereren boven een lage prijs.
4. Indien behalve tekstuele attribuutinformatie, ook pictoriele informatie over het attribuut vorm verschafft wordt, leidt dit niet tot een holistische
verwerking. Dergelijke concepten worden, net als andere concepten met tekstuele informatie, op analytische wijze verwerkt. Dit wil zeggen dat de afzonderlijke attributen systematisch geëvalueerd worden.

5. Indien informatie over het attribuut vorm pictoriële wordt weergegeven leidt dit tot andere evaluaties dan wanneer dezelfde informatie tekstueel wordt weergegeven. Een tekstueel weergegeven niveau van het attribuut vorm kan negatief geëvalueerd worden, terwijl een pictoriële weergave van datzelfde attribuut positief geëvalueerd wordt, en omgekeerd.

Deze bevindingen zijn het antwoord op de eerste onderzoeks vraag. Het voorleggen van vroege produktconcepten (bestaande uit minder informatie dan gerede produkten), heeft gevolgen voor de verwerking van de produktinformatie. Bovendien zijn, zowel door weinig (en/of ambiguë) informatie en de aanbiedingsvorm, de evaluaties anders dan bij gerede produkten. Hierop blijkt de variabele "kennis van het produktveld" een mediërende invloed te hebben.

Uit het antwoord op de eerste onderzoeks vraag, wordt het duidelijk dat alle variabelen van het "Model of Product-Concept Evaluation" van belang zijn voor de beantwoording van de tweede onderzoeks vraag. Om die reden zijn deze variabelen opgenomen als factoren die de betrouwbaarheid en/of de validiteit van de evaluaties van vroege produktconcepten kunnen beïnvloeden. De eerste stap om de tweede onderzoeks vraag te beantwoorden bestond uit de keuze voor een techniek om evaluatieve responsen te kunnen meten. Op theoretische en praktische gronden viel hierbij de keuze op conjuncte analyse. Vervolgens heeft een literatuurstudie plaatsgevonden naar de betrouwbaarheid en de validiteit van met behulp van conjuncte analyse verkregen resultaten. Deze studie, echter, verschafte weinig inzicht in de invloed van de factoren die wij van belang achten.

Omdat wij dit inzicht noodzakelijk vonden, is additioneel onderzoek opgezet en uitgevoerd om dit te verkrijgen. De bevindingen van dit onderzoek zijn:

1. De betrouwbaarheid van resultaten verkregen door middel van conjuncte analyse is acceptabel (maar niet hoog) indien respondenten een redelijke hoeveelheid, of veel, kennis hebben van het betreffende produktveld. De variabelen aanbiedingsvorm en dataverzamelingsmethode bleken geen invloed te hebben op de betrouwbaarheid.

2. De "concurrent validity" van resultaten verkregen door middel van conjunct analyse is acceptabel, indien respondenten een redelijke hoeveelheid, of veel, kennis hebben van het betreffende produktveld. In die gevallen bleek de "goodness of fit" respectievelijk .68 en .70 te zijn. De variabelen aanbiedingsvorm bleek geen invloed te hebben op de "concurrent validity".

3. De variabelen van het "Model of Product-Concept Evaluation" bleken nauwlijks van invloed te zijn op de predictieve validiteit van de door middel van conjunct analyse verkregen resultaten. Het percentage juiste voorspellingen (matches) ten aanzien van de deeluiteilten van het attribuut vorm was 45 in de tekstuele conditie en 43 in de pictoriële conditie. Indien een respondent in een concept test veel waarde hecht aan een druppelstop, hetgeen blijkt uit een uitgesproken positieve deeluitilititeit ten aanzien van dit attribuutniveau, dan is er sprake van een match wanneer deze respondent een koffiezet-apparaat koopt waarin een druppelstop aanwezig is. De percentages matches varieerden tussen 24 (weinig kennis van produktveld, attribuut prijs) en 55 (weinig kennis van produktveld, attribuut thermostan).
De predictieve validiteit van de deelutiliteiten van het attrubtuur prijs bleek aanzienlijk hoger te zijn als weinig of geen informatie werd verschaft over het belangrijke attribuut vorm (60% en 50% correcte voorspellingen), dan wanneer hierover veel informatie werd verschaft, (30% en 26% correcte voorspellingen).

De predictieve validiteit van sensitiviteiten is nogal laag. In slechts één van drie tests lag de gemiddelde sensitiviteit van de respondenten met matches hoger dan van respondenten met mismatches. Hieruit kunnen we concluderen dat attribuutniveaus met geprononceerde deelutiliteiten niet vaker gereflecteerd worden in de gekoorte produkten dan attribuutniveaus met weinig uitgesproken deelutiliteiten.

Omdat evaluatieve responsen in dit onderzoek gemeten zijn door middel van conjuncte analyse, is het antwoord op de tweede onderzoeks vraag dat resultaten verkregen door middel van conjuncte analyse betrouwbaar zijn en "concurrently valid", mits de respondenten een redelijke hoeveelheid, of veel, kennis hebben van het betreffende produktveld. De predictieve validiteit van de resultaten, daarentegen, is laag. Deze laatste bevinding kan ten dele verklaard worden door de volgende redenen: ten opzichte wordt (behalve door het aanbod) mede bepaald door een groot aantal externe factoren die in de praktijk een rol spelen en in een laboratorium uitgeschakeld zijn, slechts een klein aantal respondenten had een nieuwe koffiezetter gekocht na geïnacteerd te hebben in een van de experimenten, en de respondenten waren zowel "prospective buyers" als "potential buyers".

De bevindingen van dit onderzoek hebben veel implicaties, zowel op theoretisch als op praktisch gebied. De belangrijkste theoretische implicatie is dat het voorgestelde "Model of Product-Concept Evaluation" een bruikbaar kader vormt voor de bestudering van het evaluatieproces van vroege produktconcepten. In dit verband is aan te tonen dat de verwerking van produktinformatie en kenmerken van consumenten belangrijke variabelen zijn. Deze variabelen verdienen dan ook meer aandacht in onderzoek naar de betrouwbaarheid en/of de validiteit van technieken van consumentenonderzoek die erop gericht zijn evaluaties van consumenten te integreren in het produktontwikkelingsproces.

De belangrijkste praktische implicatie van de bevindingen is dat respondenten die worden uitgenodigd om vroege produktconcepten te evalueren tenminste een redelijke hoeveelheid, maar bij voorkeur veel kennis, van het betreffende produktveld dienen te hebben. Alleen deze respondenten zijn in staat vroege produktconcepten te evalueren. Bijgevolg, geldt dat alleen de evaluaties van respondenten met tenminste een redelijke hoeveelheid kennis van het produktveld dienen gebruikt te worden om het produktontwikkelingsproces bij te sturen.

Als deze respondenten geselecteerd worden heeft het zin om de uitkomsten van konsumentenonderzoek reeds vanaf de vroege stadia van het ontwikkelingsproces te gebruiken. Om nu optimaal gebruik te kunnen maken van deze uitkomsten is een goede afstemming tussen produktontwikkelaar en consumentenonderzoeker vereist.
References

Acito, F., (1977);

Acito, F. & T.P. Hustad (1981);

Acito, F. & A.K. Jain (1980);

Ajzen, I. (1988);

Ajzen I. & M. Fishbein (1977);
Attitude-behavior relations: a theoretical analysis and review of empirical research. Psychological Bulletin 84, 888-918.

Akaah, I.P. & P.K. Korgaonkar (1983);

Alba, J.W. (1983);
The effects of product knowledge on the comprehension, retention, and evaluation of product information. A.C.R. 10, 577-580.

Alba, J.W. & J.W. Hutchinson (1987);

Anderson, J.C. (1987);

Anderson, J.C. & N. Donthu (1988);

Anderson, N.H. (1981);

Anderson, N.H. (1982);

Argarwal, M.K. & P.E. Green (1991);

Baker, K.G. & G.S. Albaum (1986);
Batsell, R.R. & Y. Wind (1980);  
Product testing: current methods and needed developments.  

Beckwith, N.E. & D.R. Lehmann (1975);  
The importance of halo effects in multi-attribute attitude models.  
Journal of Marketing Research 12, 265-275.

Belk, R. (1975);  
Situational variables and consumer behavior. Journal of Consumer  
Research 2, 157-164.

Bettman, J.R. (1979);  
An information processing theory of consumer choice. Reading,  
Mass.: Addison-Wesley.

Boecker, F. & H. Schweikl (1988);  
Better preference prediction with individualized sets of relevant  

De Bont, C.J.P.M. (1988);  
Achtergronden bij het gebruik van koffiezetapparaten. Delft:  
Technische Universiteit Delft, Faculteit van het Industrieel  
Ontwerpen, intern rapport.

De Bont, C.J.P.M. (1990);  
Using computer-aided design in concept testing: computer  
techniques and the collection of market research data. Athens:  
proceedings of the EMAC/ESOMAR symposium, 19-27.

De Bont, C.J.P.M. & G. Loosschilder (1992);  

De Bont, C.J.P.M. & L.C.M.J. de Graaf (1989);  
Schematic pictorial stimuli in product-concept testing. Innsbruck:  
proceedings of the European Marketing Academy, 1287-1298.

De Bont, C.J.P.M. & M.A. de Lind van Wijngaarden (1991);  
De toepasbaarheid van het CAD-pakket Microsolid voor  
konsumentenonderzoek. Delft: Technische Universiteit Delft,  
Faculteit van het Industrieel Ontwerpen, Intern rapport.

De Bont, C.J.P.M., J.P.L. Schoormans, & M.T.T. Wessel (1992);  
Consumer personality and the acceptance of product design.  
Design Studies 13, 200-208.

Booz, Allen & Hamilton (1968);  
Inc.

Booz, Allen & Hamilton (1982);  
Hamilton Inc.

Box, J.M.F. & G.H.A. van Eyk (1983);  
Industriële produktontwikkeling en marktonderzoek, Deel II De  

Bretton-Clark (1986);  

Bruccs, M. (1985);  
The effects of product class knowledge on consumer search  

Brunswik, E. (1943);  
Organismic achievement and environmental probability.  
Psychological Review 50, 255-72.

Brunswik, E. (1956); Perception and the representative design of psychological experiments. Berkeley: University of California Press.


Cattin, P., & M. Weinberger (1980); Some validity and reliability issues in the measurement of attribute utilities. A.C.R. 7, 780-783.


Cooper, R.G. (1979);

Cooper, R.G. (1982);

Cooper, R.G. (1986);

Crawford, C.M. (1985):

Crawford, C.M. (1987):

Cronbach, L.J. (1984);

Cronbach, L.J., Gleser, G.C., Harinder Nanda, C. & N. Rajaratnam (1972);

Currim, I.S., C.B. Weinberg & D.R. Wittenk (1981);

Darmon, R.Y. & D. Rouzies (1989);

Darmon, R.Y. & D. Rouzielès-Ségallia (1990);
Assessing the internal validity of various conjoint analysis procedures for estimating continuous utility functions. Proceedings of the EMAC, Innsbruck, 1127-1143.

Day, G.S. & R. Wensley (1988);

Domzal, T.J. & L.S. Unger (1985);
Judgments of verbal versus pictorial presentations of a product with functional and aesthetic features. A.C.R., 12, 268-72.

Drenth, P.J.D. (1975);
Inleiding in de testtheorie. Deventer: Van Loghum Slaterus

Edell, J.A. & R. Staelin (1983);

Edwards, A.L. (1969);

Ericson, K.A. & H.A. Simon (1980);

Fazio, R.H. & M.P. Zanna (1981);
Finn, A. (1985);  

Fishbein, M. & I. Ajzen (1975);  

Gabor, A. & C.W.J. Granger (1966);  

Garner, W.R. (1978);  

Green, P.E., F.J. Carmone & Y. Wind (1972);  
Subjective evaluation models and conjoint measurement (1972), Behavioral Science 17, 288-299.

Green, P.E., J.D. Carrol & S.M. Goldberg (1981a);  

Green, P.E., S.M. Goldberg, & M. Montemayor (1981b);  

Green, P.E., S.M. Goldberg, & J.B. Willey (1982);  
A cross-validation test of hybrid conjoint models. A.C.R., 147-150.

Green, P.E. & K. Helsen (1989);  

Green, P.E., K. Helsen and B. Shandler (1988a);  

Green, P.E., A.M. Krieger, & P. Bansal (1988b);  

Green, P.E. & V. Srinivasan (1978);  
Conjoint analysis in consumer research: issues and outlook. Journal of Consumer Research 5, 103-123.

Green, P.E. & V. Srinivasan (1990);  

Green, P.E. & V.R. Rao (1971);  

Green, P.E., V.R. Rao, & W.S. Desarbo (1978);  

Greenwald, A.G. & C. Leavitt (1984);  

Griffin, A. & J.R. Hauser (1991);  
The voice of the customer. Cambridge: working paper Sloan School of Management.


Holbrook, M.B. (1981); Integrating compositional and decompositional analysis to represent the intervening role of perceptions in evaluative judgments. Journal of Marketing Research 18, 13-28.


Holt, K. (1989); Does the engineer forget the user? Design Studies 10, 163-168.


Huber, J. (1975); Predicting preferences on experimental bundles of attributes: a comparison of models. Journal of Marketing Research, 12, 290-297.


References


Iuso, B. (1975); Concept testing: an appropriate approach. Journal of Marketing Research 12, 228-231.


Klein, N.M. (1986); Assessing unacceptable attribute levels in conjoint analysis. A.C.R., 154-158.


Kuehn, A.A. & R.L. Day (1962);
Lancaster, K.J. (1966);
Lancaster, K.J. (1971);
Laric, M. (1979);
Are consumers able to understand concepts in the early stages of development; cost effective approaches. Amsterdam: Esomar conference papers, Dubrovnik, 103-106.
Leeflang, P.S.H. & P.A. Beukenkamp (1987);
Probleemgebied marketing, een management benadering. Antwerpen: Stenfert Kroese.
Leigh, T.W., MacKay, D.B. & J.O. Summers (1981);
Leigh, T.W., MacKay, D.B. & J.O. Summers (1984);
Locander, W.B. & R.W. Scamell (1976);
Louviere, J.J. (1984);
Louviere, J.J. (1988);
Analyzing decision making: metric conjoint analysis. Quantitative applications in the social science, A Sage University paper.
Louviere, J.J. & G.J. Gaeth (1987);
Louviere, J.J. & G. Woodworth (1983);
Design and Analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. Journal of Marketing Research 20, 350-367.
Luce, R.D. & J.W. Tukey (1964);
Simultaneous conjoint measurement: a new type of fundamental measurement. Journal of Mathematical Psychology 1, 1-27.
Maheswaran, D. & B. Sternthal (1990);
The effects of knowledge, motivation, and type of message on ad processing and product judgments. Journal of Consumer Research 17, 66-73.
Malhotra, N.K. (1986);

Mantel, S.J. & J.R. Meredith (1986);
The role of customer cooperation in the development, marketing, and implementation of innovations. In: The art and science of innovation management. Ed. by H. Huebner. Amsterdam: Elsevier.

McCullough, J. & R. Best (1979);

McGuire, W.J. (1969);

Meyer, R.J. (1987);

Mishra, S., U.N. Umesh & D.E. Stem (1989);
Attribute importance weights in conjoint analysis: bias and precision. A.C.R. 16, 605-611.

Moore, W.L. (1980);

Moore, W.L. & M.B. Holbrook (1982);

Moore, W.L. & M.B. Holbrook (1990);
Conjoint analysis on objects with environmentally correlated attributes: the questionable importance of representative design. Journal of Consumer Research 16, 490-497.

Muller, W. (1990);
Vormgeven, ordening en betekenisgeving. Utrecht: Lemma B.V.

Nisbett, R.E. & T. Wilson (1977);
Telling more than we can know: verbal reports on mental processes. Psychological Review 84, 231-259.

Nunnally, J.C. (1967);

Oppedijk van Veen, W.M. & D. Beazley (1977);

Oppewal, H. (1992);

Ortt, R.J., J.P.L. Schoormans & W.M. Oppedijk van Veen (1992);
Incorporating consumer research in the development process of major innovations: two cases in telematics. Brussels: paper presented at the EIASM.

Page, A.L. & H.F. Rosenbaum (1987);
Parker, B.R. & V. Srinivasan (1976);
A consumer preference approach to the planning of rural primary

Pavila, T.M. (1991);
The early stages of new-product development in entrepreneurial

Petty, R.E., Cacioppo, J.T. & D. Schuman (1983);
Central and peripheral routes to advertising effectiveness: the
moderating role of involvement. Journal of Consumer Research 10,
135-146.

Pieters, R.G.M. (1989);
Attitudes and behavior in a source-separation program: a garbology

Ratneshwar, S., & S. Chaiken (1991);
Comprehension's role in persuasion: the case of its moderating
effect on the persuasive impact of source cues. Journal of
Consumer Research 18, 52-62.

Van Raaij, W.F. (1977);
Consumer choice behavior: an information processing approach.
Dissertation, Tilburg University.

Rao, A.R. & K.B. Monroe, (1988);
The moderating effect of prior knowledge on cue utilization in

Reibstein, D., Bateson, J.E.G., & W. Boulding (1988);
Conjoint analysis reliability: empirical findings. Marketing Science
7, 271-286.

Reynolds, T.J. & J. Gutman (1984);
Advertising is image management. Journal of Advertising Research
24, 27-37.

Roozenburg, N. en J. Eekels (1991);

Rosenberg, M.J. (1956);
Cognitive structure and attitudinal affect. Journal of Abnormal and

Rothwell, R., C. Freeman, A. Horsley, V.I.P. Jervis, A.B. Robertson,
& J. Townsend (1974);
'SAPPHO' updated - project SAPPHO, phase 2. Research Policy 3,
30-38.

Sawtooth Software (1986);
ACA system for adaptive conjoint analysis. Ketchum: Sawtooth
Software.

Schwager, W. (1988);
Theories of measurement in the social sciences, a critical review.
Rotterdam: Erasmus University.

Scott, J.E. & P. Wright (1976);
Modeling an organizational buyer's product evaluation strategy:
validity and procedural considerations. Journal of Marketing
Research 13, p.211-224.

Scott, J.E. & S.K. Keiser (1984);
Forecasting the acceptance of new industrial products with
Segal, M.N. (1982);
Reliability of conjoint analysis: contrasting data collection procedures. Journal of Marketing Research 19, 139-143.
Selles, F. & K. Grönhaug (1986);
Sherak, B. (1966);
Testing new product ideas. Amsterdam: Esomar congress papers.
Simmons, C.J. & J.G. Lynch (1991);
Slovic, P. & S. Lichtenstein (1971);
Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance 6, 649-744.
Smed, R.J., J.B. Wilcox & R.E. Wilkes (1981);
How valid are product descriptions and protocols in choice experiments. Journal of Consumer Research 8, 37-42.
Snelders, H.M.J.J., J.P.L. Schoormans & C.J.P.M. de Bont (1992);
Spearman, C. (1910);
Srinivasan, V. (1988);
Srinivasan, V., P.G. Flachsbart, J.S. Dajani & R.G. Hartley (1981);
Forecasting the effectiveness of work-trip gasoline conservation policies through conjoint analysis. Journal of Marketing Research 45, 157-172.
Srinivasan, V., A.K. Jain & N.K. Malhotra (1983);
Srinivasan, V. & A.D. Shocker (1973);
Srull, T.K. (1983);
The role of prior knowledge in the acquisition, retention, and use of new information. A.C.R. 10, 572-576.
Steenkamp, J.B.E.M. (1989);
Steenkamp, J.B.E.M., B. Wierenga & M.T.G. Meulenberg (1986);
Stevens, S.S. (1975);

Sujan, M. (1985);

Tan, A.H.L. (1989);

Tauber, E.M. (1972);
What is measured by concept testing. Journal of Advertising Research 12, 35-7.

Timmermans, H.J.P. (1984);

Triesscheijn, E.W.M. (1982);

Troutman, C.M. & J. Shanteau (1976);

Unnava, H.R. & R.E. Burnkrant (1991);

Urban, G.L., J.R. Hauser & N. Dholakia (1987);

Verhallen, Th.M.M. & H. Vogel (1982);
Technieken van kwalitatief onderzoek. Tijdschrift voor Marketing, November 28-32.

Virzy, R.A. (1989);

Vriens, M. & D.R. Wittink (1990);

Walker, B., R. Celsi, & J. Olson (1987);

Wilding, D. (1986);
The contribution of market research to the testing of new concepts. Marketing Intelligence and Planning 3, 4-12.

Wilkie, W.L. (1990);

Wind, Y.J. (1982);
Product policy: concepts, methods, and strategy. Reading (Mass.): Addison-Wesley.

Wittink, D.R. (1989);
Wittink, D.R. & P. Cattin (1981);

Wittink, D.R. & P. Cattin (1989);

Wittink, D.R., L. Krishnamurthi, & J.B. Nutter (1982);

Wittink, D.R., L. Krishnamurthi & D.J. Reibstein (1989a);
The effect of differences in the number of attribute levels on conjoint results. Marketing Letters 1, 113-123.

Wittink, D.R. & D.B. Montgomery (1979);
Predictive validity of trade-off analysis for alternative segmentation schemes.

Wittink, D.R., D.J. Reibstein, W. Building, J.E.G. Bateson & J.W. Walsh (1989b);

Woodside, A.G. & W.G. Pearce (1989);

Wright, P.L. (1975);

Wright, P.L. (1980);

Wright, P. & M.A. Kriewall (1980);
State-of-mind effects on the accuracy with which utility functions predict marketplace choice. Journal of Marketing Research 17, 277-293.

Yuspeh, S. (1975);
Diagnosis - the handmaiden for prediction. Journal of Marketing 39, 87-89.

Zaltman, G., C.R.A. Pinson & R. Angelmar (1973);

Zufryden, F.S. (1981);
Appendix 1: An illustration of conjoint analysis: The Sunbeam case

From the literature on conjoint analysis, different examples of conjoint analysis (e.g. Acito & Hustad, 1981; Currim et al., 1981; Srinivasan et al. 1981) can be chosen for illustration. We have chosen the Sunbeam case presented by Page & Rosenbaum (1987) for three reasons:

1. It is positioned in the context of new-product development.
2. It is an application in a commercial setting.
3. The methodology is commented by other investigators.

The Sunbeam case refers to a demonstration of conjoint analysis by Page & Rosenbaum (1987) applied in the commercial setting of Sunbeam Appliance Company (S.A.C.). S.A.C. produces small kitchen appliances. This illustration concerns the redesign of a product line of a food processor. The starting point for developing a new-product line was the recognition of the fact that the existing product line already was mature by the new management team. Management's goal for the redesign of this product line was to optimize their market share of each product category. Central questions to the problem of redesign were:

1. What models should be in the line?
2. What should their physical appearance be?
3. What should their performance characteristics be?

The consumer-research technique applied in answering these questions is conjoint analysis. The use of this technique, however, is embedded in other steps. The first step is a consumer usage and attitude survey. The main purpose of this step is collecting background information on: how and for what purposes products in the product category are used, frequency of use, etc. The second step is a consumer attribute and benefit survey. In this step, importance ratings of product attributes and benefits are collected along with their perceived presence of absence in each of the products in each of the key competitive brands including S.A.C. The purpose of this second step is identifying 'need gaps', where the perceived benefits are not being supplied by the existing brands. In the third step conjoint analysis is applied. First, insights from exploratory group discussions are collected to determine the attributes and attribute levels for the conjoint-analysis study. Second, the structure of consumer preferences for different product attributes is determined by conjoint analysis. The fourth, and last, step consists of product line simulations. In this step, the market shares of different versions of product lines are predicted from the structure of consumers' preferences.

The conjoint-analysis study of Page & Rosenbaum (1987) can be illustrated along the steps involved in conjoint analysis distinguished by Green & Srinivasan (1978):
1. Selection of a model of preference: the part-worth model. Both qualitative (bowl shape) and quantitative (number of speeds) attributes are included.
2. Data-collection method: a preference ranking of full profiles. The profiles consisted of twelve attributes each.
3. Stimulus set construction: a fractional factorial design was chosen, leading to twenty-seven orthogonal arrays (profiles) of the twelve attributes.
4. Stimulus presentation: the profiles consisted of combinations sketches (wireframe models with hidden lines) produced by industrial designers, and printed labels about the features.
5. Measurement scale for the dependent variable: preference rankings were collected, yielding data at the ordinal level.
6. Estimation method: MONANOVA

The results of the study of Page & Rosenbaum (1987) consisted of twelve utility functions (one for each attribute) for each of the (more than 500) respondents. These data were entered in a hierarchical-cluster analysis program to find groups of respondents who have similar patterns of utility functions. The clustering results revealed four market segments of food processor buyers who vary according to the importance they attach to specific attributes and who desire different combinations of attributes when purchasing a food processor. An example of a segment is the 'cheap and large segment'. Respondents in this segment require a large processor with only a minimum number of features at a low price.

After the market simulations, S.A.C. ended up with three basic models of food processors. S.A.C. felt that those products would cover the market segments they wanted to target. The introduction of the redesigned product line, finally consisting of four products, yielded a substantial increase in market share.

The application of conjoint analysis of Page & Rosenbaum (1987) is commented upon by Wittink (1989). According to Wittink (1989) a nice illustration is provided by the authors, with both excellent features and features which deserve reconsideration. The excellent features, among others, are: the use of pictorial presentations, the collection of data in four different geographic locations, and the use of simulations to predict market shares. The features which deserve reconsideration are (among others): the number of attributes, comparing across attributes with differing numbers of attribute-levels, and the use of MONANOVA. According to Wittink (1989), consumers are unlikely to keep track of twelve attributes. Green & Srinivasan (1978, p. 108), mention that "...the full-profile procedure is generally confined to, at most, five or six factors". With respect to comparing across attributes with differing numbers of attribute-levels, different studies (e.g. Wittink et al., 1982) showed that this difference affected both derived importances and market share predictions. According to Wittink (1989), the least-squares regression is more quickly and less costly than MONANOVA.
Appendix 2: Results of the control questionnaire

Experiment 1
Table A: Socio-demographic background and drip coffee maker related aspects of the respondents in the realistic pictorial condition (r.p.c.) and the schematic pictorial condition (s.p.c.)

<table>
<thead>
<tr>
<th></th>
<th>r.p.c.</th>
<th>s.p.c.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 26)</td>
<td>(n = 25)</td>
</tr>
<tr>
<td>Sex</td>
<td>abs.</td>
<td>%</td>
</tr>
<tr>
<td>Male</td>
<td>12</td>
<td>46,2</td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>53,8</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>5</td>
<td>19,2</td>
</tr>
<tr>
<td>31-40</td>
<td>9</td>
<td>34,6</td>
</tr>
<tr>
<td>41-50</td>
<td>3</td>
<td>11,5</td>
</tr>
<tr>
<td>51-60</td>
<td>6</td>
<td>23,1</td>
</tr>
<tr>
<td>older than 60</td>
<td>3</td>
<td>11,5</td>
</tr>
<tr>
<td>Family size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>11,5</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>38,5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>7,7</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>19,2</td>
</tr>
<tr>
<td>more than 4</td>
<td>6</td>
<td>23,1</td>
</tr>
<tr>
<td>No. coffee drinkers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>—</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>15,4</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>69,2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3,8</td>
</tr>
<tr>
<td>more than 3</td>
<td>3</td>
<td>11,5</td>
</tr>
<tr>
<td>Ownership</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>23</td>
<td>88,5</td>
</tr>
<tr>
<td>no</td>
<td>3</td>
<td>11,5</td>
</tr>
</tbody>
</table>
## Experiment 2

Table B: Socio-demographic background and drip coffee maker related aspects of the respondents in the schematic three-dimensional condition (s.t.c.) and the realistic three-dimensional condition (r.t.c.)

<table>
<thead>
<tr>
<th></th>
<th>s.t.c.</th>
<th>r.t.c.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 46)</td>
<td>(n = 50)</td>
</tr>
<tr>
<td></td>
<td>abs. %</td>
<td>abs. %</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>19 41,3</td>
<td>26 52,0</td>
</tr>
<tr>
<td>Female</td>
<td>27 58,7</td>
<td>24 48,0</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>8 17,4</td>
<td>4  8,0</td>
</tr>
<tr>
<td>31-40</td>
<td>15 32,6</td>
<td>13 26,0</td>
</tr>
<tr>
<td>41-50</td>
<td>11 23,9</td>
<td>14 28,0</td>
</tr>
<tr>
<td>51-60</td>
<td>2  4,3</td>
<td>4  8,0</td>
</tr>
<tr>
<td>older than 60</td>
<td>10 21,7</td>
<td>15 30,0</td>
</tr>
<tr>
<td><strong>Family size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4  8,7</td>
<td>6 12,0</td>
</tr>
<tr>
<td>2</td>
<td>17 37,0</td>
<td>22 44,0</td>
</tr>
<tr>
<td>3</td>
<td>7  15,2</td>
<td>6  12,0</td>
</tr>
<tr>
<td>4</td>
<td>13 28,3</td>
<td>13 26,0</td>
</tr>
<tr>
<td>more than 4</td>
<td>5  10,8</td>
<td>2  4,0</td>
</tr>
<tr>
<td><strong>No. coffee drinkers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1  2,2</td>
<td>1  2,0</td>
</tr>
<tr>
<td>1</td>
<td>10 21,7</td>
<td>12 24,0</td>
</tr>
<tr>
<td>2</td>
<td>28 60,9</td>
<td>28 56,0</td>
</tr>
<tr>
<td>3</td>
<td>2  4,3</td>
<td>6  12,0</td>
</tr>
<tr>
<td>more than 3</td>
<td>5  10,9</td>
<td>3  6,0</td>
</tr>
<tr>
<td><strong>Ownership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>39 84,8</td>
<td>36 72,0</td>
</tr>
<tr>
<td>no</td>
<td>7 15,2</td>
<td>14 28,0</td>
</tr>
</tbody>
</table>
Table C: Socio-demographic background and drip coffee maker related aspects of the little knowledge group (n = 33) in the realistic three-dimensional condition (r.t.c.) and the schematic three-dimensional condition (s.t.c.)

<table>
<thead>
<tr>
<th></th>
<th>r.t.c. (n = 22)</th>
<th>s.t.c. (n = 11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>abs. %</td>
<td>abs. %</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>11 50,0</td>
<td>65 4,4</td>
</tr>
<tr>
<td>Female</td>
<td>11 50,0</td>
<td>54 5,5</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>1 4,5</td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td>3 13,6</td>
<td>2 18,2</td>
</tr>
<tr>
<td>41-50</td>
<td>6 27,3</td>
<td>3 27,3</td>
</tr>
<tr>
<td>51-60</td>
<td>2 9,1</td>
<td></td>
</tr>
<tr>
<td>older than 60</td>
<td>10 45,5</td>
<td>6 54,5</td>
</tr>
<tr>
<td>Family size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 9,1</td>
<td>2 18,2</td>
</tr>
<tr>
<td>2</td>
<td>13 59,1</td>
<td>6 54,5</td>
</tr>
<tr>
<td>3</td>
<td>1 4,5</td>
<td>1 9,1</td>
</tr>
<tr>
<td>4</td>
<td>5 22,7</td>
<td>1 9,1</td>
</tr>
<tr>
<td>more than 4</td>
<td>1 4,5</td>
<td>1 9,1</td>
</tr>
<tr>
<td>No. coffee drinkers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1 4,5</td>
<td>1 9,1</td>
</tr>
<tr>
<td>1</td>
<td>6 27,3</td>
<td>1 9,1</td>
</tr>
<tr>
<td>2</td>
<td>10 45,5</td>
<td>6 54,5</td>
</tr>
<tr>
<td>3</td>
<td>3 13,6</td>
<td>1 9,1</td>
</tr>
<tr>
<td>more than 3</td>
<td>2 9,1</td>
<td>2 18,2</td>
</tr>
<tr>
<td>Ownership</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>14 63,6</td>
<td>8 72,7</td>
</tr>
<tr>
<td>no</td>
<td>8 36,4</td>
<td>3 27,3</td>
</tr>
</tbody>
</table>
Table D: Socio-demographic background and drip coffee maker related aspects of the moderate knowledge group (n=32) in the realistic three-dimensional condition (r.t.c.) and the schematic three-dimensional condition (s.t.c.)

<table>
<thead>
<tr>
<th></th>
<th>r.t.c.</th>
<th>s.t.c.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 14)</td>
<td>(n = 18)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>64</td>
<td>2,9</td>
</tr>
<tr>
<td>Female</td>
<td>8</td>
<td>57,1</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>2</td>
<td>14,3</td>
</tr>
<tr>
<td>31-40</td>
<td>5</td>
<td>35,7</td>
</tr>
<tr>
<td>41-50</td>
<td>3</td>
<td>21,4</td>
</tr>
<tr>
<td>51-60</td>
<td>2</td>
<td>14,3</td>
</tr>
<tr>
<td>older than 60</td>
<td>2</td>
<td>14,3</td>
</tr>
<tr>
<td><strong>Family size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>14,3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>35,7</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>28,6</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>21,4</td>
</tr>
<tr>
<td>more than 4</td>
<td>-</td>
<td>---</td>
</tr>
<tr>
<td><strong>No. coffee drinkers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>21,4</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>64,3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>14,3</td>
</tr>
<tr>
<td>more than 3</td>
<td>-</td>
<td>---</td>
</tr>
<tr>
<td><strong>Ownership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>10</td>
<td>71,4</td>
</tr>
<tr>
<td>no</td>
<td>4</td>
<td>28,6</td>
</tr>
</tbody>
</table>
Table E: Socio-demographic background and drip coffee maker related aspects of the much knowledge group (n = 31)\(^1\) and in the realistic three-dimensional condition (r.f.c.) and the schematic three-dimensional condition (s.f.c.)

<table>
<thead>
<tr>
<th></th>
<th>r.f.c.</th>
<th>s.f.c.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 14)</td>
<td>(n = 17)</td>
</tr>
<tr>
<td></td>
<td>abs.</td>
<td>%</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9</td>
<td>64,3</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>35,7</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>1</td>
<td>7,1</td>
</tr>
<tr>
<td>31-40</td>
<td>5</td>
<td>35,7</td>
</tr>
<tr>
<td>41-50</td>
<td>5</td>
<td>35,7</td>
</tr>
<tr>
<td>51-60</td>
<td>-</td>
<td>——</td>
</tr>
<tr>
<td>older than 60</td>
<td>3</td>
<td>21,4</td>
</tr>
<tr>
<td><strong>Family size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>14,3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>28,6</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>7,1</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>35,7</td>
</tr>
<tr>
<td>more than 4</td>
<td>2</td>
<td>14,3</td>
</tr>
<tr>
<td><strong>No. coffee drinkers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>21,4</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>64,3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>7,1</td>
</tr>
<tr>
<td>more than 3</td>
<td>1</td>
<td>7,1</td>
</tr>
<tr>
<td><strong>Ownership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>12</td>
<td>85,7</td>
</tr>
<tr>
<td>no</td>
<td>2</td>
<td>14,3</td>
</tr>
</tbody>
</table>
### Experiment 3

Table F: Socio-demographic background and drip coffee maker related aspects of the respondents in the pictorial condition and the textual condition

<table>
<thead>
<tr>
<th></th>
<th>Pictorial condition</th>
<th>Textual condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td>(n = 21)</td>
<td>(n = 20)</td>
</tr>
<tr>
<td></td>
<td>abs.    %</td>
<td>abs.    %</td>
</tr>
<tr>
<td>Male</td>
<td>8 38,1</td>
<td>8 40,0</td>
</tr>
<tr>
<td>Female</td>
<td>13 61,9</td>
<td>12 60,0</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>1 4,8</td>
<td>6 30,0</td>
</tr>
<tr>
<td>31-40</td>
<td>5 23,8</td>
<td>4 20,0</td>
</tr>
<tr>
<td>41-50</td>
<td>2 9,5</td>
<td>2 10,0</td>
</tr>
<tr>
<td>51-60</td>
<td>6 28,6</td>
<td>7 35,0</td>
</tr>
<tr>
<td>older than 60</td>
<td>7 33,3</td>
<td>1 5,0</td>
</tr>
<tr>
<td><strong>Family size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4 19,0</td>
<td>4 20,0</td>
</tr>
<tr>
<td>2</td>
<td>10 47,6</td>
<td>8 40,0</td>
</tr>
<tr>
<td>3</td>
<td>2 9,5</td>
<td>3 15,0</td>
</tr>
<tr>
<td>4</td>
<td>3 14,3</td>
<td>3 15,0</td>
</tr>
<tr>
<td>more than 4</td>
<td>2 9,5</td>
<td>2 10,0</td>
</tr>
<tr>
<td><strong>No. coffee drinkers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5 23,8</td>
<td>6 30,0</td>
</tr>
<tr>
<td>2</td>
<td>13 61,9</td>
<td>11 55,0</td>
</tr>
<tr>
<td>3</td>
<td>2 9,5</td>
<td>2 10,0</td>
</tr>
<tr>
<td>more than 3</td>
<td>1 4,8</td>
<td>1 5,0</td>
</tr>
<tr>
<td><strong>Ownership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>20 95,2</td>
<td>19 95,0</td>
</tr>
<tr>
<td>no</td>
<td>1 4,8</td>
<td>1 5,0</td>
</tr>
</tbody>
</table>
Experiment 4

Table G: Socio-demographic background and drip coffee maker related aspects of the respondents in the "no information" condition (N=41)²

<table>
<thead>
<tr>
<th><em>no information</em> condition</th>
<th>(n = 41)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>abs.</td>
</tr>
<tr>
<td>Male</td>
<td>19</td>
</tr>
<tr>
<td>Female</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>21-30</td>
<td>7</td>
<td>17,1</td>
</tr>
<tr>
<td>31-40</td>
<td>14</td>
<td>34,1</td>
</tr>
<tr>
<td>41-50</td>
<td>8</td>
<td>19,5</td>
</tr>
<tr>
<td>51-60</td>
<td>8</td>
<td>19,5</td>
</tr>
<tr>
<td>older than 60</td>
<td>4</td>
<td>9,8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family size</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>19,5</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>43,9</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>12,2</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>14,6</td>
</tr>
<tr>
<td>more than 4</td>
<td>4</td>
<td>9,8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. coffee drinkers</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>26,8</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>58,5</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>9,8</td>
</tr>
<tr>
<td>more than 3</td>
<td>2</td>
<td>4,9</td>
</tr>
</tbody>
</table>
Experiment 5
Table H: Socio-demographic background and drip coffee maker related aspects of the respondents in the ACA-condition (N=40)\(^3\)

<table>
<thead>
<tr>
<th></th>
<th>ACA condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 40)</td>
<td>abs. %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>21 52,5</td>
</tr>
<tr>
<td>Female</td>
<td>19 47,5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>21-30</td>
<td>8 20,0</td>
</tr>
<tr>
<td>31-40</td>
<td>9 22,5</td>
</tr>
<tr>
<td>41-50</td>
<td>6 15,0</td>
</tr>
<tr>
<td>51-60</td>
<td>8 20,0</td>
</tr>
<tr>
<td>older than 60</td>
<td>9 22,5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family size</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 12,5</td>
</tr>
<tr>
<td>2</td>
<td>16 40,0</td>
</tr>
<tr>
<td>3</td>
<td>7 17,5</td>
</tr>
<tr>
<td>4</td>
<td>8 20,0</td>
</tr>
<tr>
<td>more than 4</td>
<td>4 10,0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. coffee drinkers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8 20,0</td>
</tr>
<tr>
<td>2</td>
<td>26 65,0</td>
</tr>
<tr>
<td>3</td>
<td>4 10,0</td>
</tr>
<tr>
<td>more than 3</td>
<td>2 5,0</td>
</tr>
</tbody>
</table>

\(^3\) The results of the full-profile ranking condition are already shown in Table G under the heading of the "without form" condition.
### Appendix 3: The structure of studies on the reliability and the validity of conjoint-analysis results

<table>
<thead>
<tr>
<th>Model of preference</th>
<th>Data collection method</th>
<th>Stimulus set construction</th>
<th>Stimulus presentation</th>
<th>Measurement scale for dependent variable</th>
<th>Estimation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green et al. 1972</td>
<td>part-worth, full-profile, 27 profiles of 3 attributes</td>
<td>full factorial</td>
<td>textual</td>
<td>preference ranking</td>
<td>MONANOVA, MDPREF, PREFMAP</td>
</tr>
<tr>
<td>Johnson 1974</td>
<td>part-worth, trade-off matrices</td>
<td>full factorial</td>
<td>textual</td>
<td>preference ranking</td>
<td>Non-metric scaling</td>
</tr>
<tr>
<td>Huber 1975</td>
<td>part-worth, ideal-point two-stage, full-profile, 23 profiles of 5 attributes</td>
<td>fractional factorial</td>
<td>textual</td>
<td>graded pair-comparisons</td>
<td>PREFMAP</td>
</tr>
<tr>
<td>Parker &amp; Srinivasan 1976</td>
<td>part-worth, full profile 25 profiles of 5 attributes</td>
<td>fractional factorial</td>
<td>textual</td>
<td>preference ranking</td>
<td>LINMAP</td>
</tr>
<tr>
<td>Scott &amp; Wright 1976</td>
<td>part-worth, full profile, 32 profiles of 2, 3 and 6 attributes</td>
<td>full factorial and fractional factorial</td>
<td>textual</td>
<td>preference rating</td>
<td>O.L.S.</td>
</tr>
<tr>
<td>Acito 1977</td>
<td>part-worth, full profile, 27 profiles of 3 attributes</td>
<td>full factorial</td>
<td>textual</td>
<td>preference ranking</td>
<td>MONANOVA</td>
</tr>
<tr>
<td>Oppedijk, van Veen &amp; Beazley 1977</td>
<td>part-worth, full profile, 27 profiles of 3 attributes, trade-off matrices</td>
<td>full factorial</td>
<td>textual and pictorial</td>
<td>preference rating</td>
<td>MONANOVA, PERMUT</td>
</tr>
<tr>
<td>Carmone et al. 1978</td>
<td>part-worth, full profile, 243, 54, 27, 18 profiles of 5 attributes</td>
<td>full factorial and fractional factorial</td>
<td>?</td>
<td>preference rating</td>
<td>MONANOVA, ANOVA</td>
</tr>
<tr>
<td>Green et al. 1978</td>
<td>part-worth, full profile, 18 profiles of 6 attributes</td>
<td>fractional factorial</td>
<td>textual</td>
<td>preference rating</td>
<td>O.L.S.</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Methodology</td>
<td>Design Description</td>
<td>Factorial Design</td>
<td>Scale</td>
<td>Model Type</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Heeler et al.</td>
<td>part-worth</td>
<td>full profile, 16 profiles of 10 attributes, single attributes</td>
<td>fractional</td>
<td>textual</td>
<td>preference ranking</td>
</tr>
<tr>
<td>1979</td>
<td></td>
<td></td>
<td>factorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huber &amp; Moore</td>
<td>part-worth</td>
<td>full profile, 16 profiles of 8 attributes</td>
<td>fractional</td>
<td>textual</td>
<td>preference rating</td>
</tr>
<tr>
<td>1979</td>
<td></td>
<td></td>
<td>factorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jain et al.</td>
<td>part-worth</td>
<td>full profile, 27 profiles of 5 attributes, trade-off matrices</td>
<td>fractional</td>
<td>textual</td>
<td>preference ranking</td>
</tr>
<tr>
<td>1979</td>
<td></td>
<td></td>
<td>factorial</td>
<td></td>
<td>JOHNSON</td>
</tr>
<tr>
<td>McCullough &amp;</td>
<td>part-worth</td>
<td>full profile, 27 and 18 profiles of 3 attributes</td>
<td>fractional</td>
<td>textual</td>
<td>preference ranking</td>
</tr>
<tr>
<td>Best 1979</td>
<td></td>
<td></td>
<td>factorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wittink &amp;</td>
<td>trade-off</td>
<td>full profile, 27 profiles of 4 attributes, 16 profiles of 6 attributes, 16</td>
<td>fractional</td>
<td>textual</td>
<td>preference ranking</td>
</tr>
<tr>
<td>Montgomery</td>
<td>matrices</td>
<td>profiles of 5 attributes</td>
<td>factorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1979</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acito &amp; Jain</td>
<td>part-worth</td>
<td>full profile, 16 or 32 profiles of 6 or 9 attributes, trade-off matrices, direct estimates</td>
<td>fractional</td>
<td>textual</td>
<td>preference ranking</td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td>factorial</td>
<td></td>
<td>O.L.S.</td>
</tr>
<tr>
<td>Cattin &amp;</td>
<td>full profile,</td>
<td></td>
<td>fractional</td>
<td>textual</td>
<td>preference ratings, preference ranking, constant sum scale</td>
</tr>
<tr>
<td>Weinberger</td>
<td>16 or 32</td>
<td></td>
<td>factorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>profiles of 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>or 9 attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>trade-off</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>matrices</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moore</td>
<td>part-worth</td>
<td>full profile, 24 profiles of 5 attributes</td>
<td>fractional</td>
<td>textual</td>
<td>preference ratings</td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td>factorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wright &amp;</td>
<td>full profile,</td>
<td></td>
<td>fractional</td>
<td>textual</td>
<td>preference ratings</td>
</tr>
<tr>
<td>Kriewal</td>
<td>32 profiles of 5 attributes</td>
<td>factorial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carmone &amp;</td>
<td>full profile,</td>
<td></td>
<td>fractional</td>
<td>?</td>
<td>preference rating</td>
</tr>
<tr>
<td>Green 1981</td>
<td>18 or 27</td>
<td></td>
<td>factorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>profiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green et al.</td>
<td>full profile,</td>
<td></td>
<td>fractional</td>
<td>textual</td>
<td>preference rating</td>
</tr>
<tr>
<td>1981</td>
<td>single</td>
<td></td>
<td>factorial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>attributes,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>self-expli-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cated, hybrid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Methodology</td>
<td>Profiles/Attributes</td>
<td>Design/Estimation</td>
<td>Measure/Analysis</td>
<td>Technique(s)</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Leigh et al. 1981</td>
<td>part-worth</td>
<td>full profile, 2 or 4 profiles of 2 or 4 attributes</td>
<td>fractional factorial</td>
<td>textual rank order, pair-comparison, graded pair-comparison, direct estimates, ratings</td>
<td>MONANOVA O.L.S.</td>
</tr>
<tr>
<td>Srinivasan et al. 1981</td>
<td>part-worth</td>
<td>full profile, 2 or 4 profiles of 2 or 4 attributes</td>
<td>full factorial and fractional factorial</td>
<td>textual preference ranking</td>
<td>LINMAP III</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ANOVA LINMAP LOGIT MONANOVA</td>
</tr>
<tr>
<td>Green et al. 1982</td>
<td>part-worth, self-explained, hybrid</td>
<td>full profile, 32 profiles</td>
<td>fractional factorial</td>
<td>textual preference rating (calibration), preference ranking (validation)</td>
<td>O.L.S.</td>
</tr>
<tr>
<td>Moore &amp; Holbrook 1982</td>
<td>joint-space</td>
<td>full profile, 24 profiles of 5 attributes</td>
<td>fractional factorial</td>
<td>textual preference ratings</td>
<td>O.L.S.</td>
</tr>
<tr>
<td>Segall 1982</td>
<td>part-worth</td>
<td>full profile, 27 profiles of 3 attributes, trade-off matrices</td>
<td>full factorial</td>
<td>textual preference ranking</td>
<td>MONANOVA, Johnsons non-metric regression</td>
</tr>
<tr>
<td>Wittink et al. 1982</td>
<td>part-worth</td>
<td>full profile, 16 profiles of 4 attributes, trade-off matrices</td>
<td>fractional factorial</td>
<td>textual preference ranking</td>
<td>LINMAP</td>
</tr>
<tr>
<td>Akaah &amp; Korgaonkar 1983</td>
<td>self-explained, Huber-hybrid, part-worth</td>
<td>full profile, 27 profiles of 6 attributes</td>
<td>fractional factorial</td>
<td>textual preference ratings</td>
<td>O.L.S.</td>
</tr>
<tr>
<td>Srinivasan et al. 1983</td>
<td>part-worth</td>
<td>full profile, 27 profiles of 5 attributes, trade-off matrices</td>
<td>fractional factorial</td>
<td>textual preference ranking</td>
<td>LINMAP</td>
</tr>
<tr>
<td>Leigh et al. 1984</td>
<td>part-worth</td>
<td>full profile, 32, 16 or 8 profiles of 5 attributes</td>
<td>fractional factorial, and full factorial</td>
<td>textual rank order, pair-comparison, graded pair-comparison, direct estimates, ratings</td>
<td>MONANOVA ANOVA</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Methodology</td>
<td>Design</td>
<td>Estimation</td>
<td>Model</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------</td>
<td>-------------------------------</td>
<td>------------</td>
<td>------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Scott &amp; Keiser 1984</td>
<td>part-worth, discrete choice</td>
<td>full profile, 24 profiles of 5 or 4 attributes</td>
<td>fractional</td>
<td>market share estimate, discrete choice, preference ranking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>textual</td>
<td>O.L.S., PROBIT MONANOVA</td>
<td></td>
</tr>
<tr>
<td>Jaccard et al 1986</td>
<td>part-worth, self-explained</td>
<td>full profile, 16 profiles of 4 attributes</td>
<td>full</td>
<td>preference ranking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>MONANOVA</td>
<td></td>
</tr>
<tr>
<td>Klein 1986</td>
<td>part-worth, elimination</td>
<td>trade-off matrices</td>
<td>fractional</td>
<td>preference ranking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>JOHNSON</td>
<td></td>
</tr>
<tr>
<td>Malhotra 1986</td>
<td>part-worth</td>
<td>full profile, 25 profiles of 5 attributes</td>
<td>fractional</td>
<td>preference rating</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>O.L.S. Maximum Likelihood</td>
<td></td>
</tr>
<tr>
<td>Anderson 1987</td>
<td>part-worth</td>
<td>full profile, 18 profiles of 3 attributes</td>
<td>full</td>
<td>preference rating</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>O.L.S.</td>
<td></td>
</tr>
<tr>
<td>Page &amp; Rosenbaum 1987</td>
<td>part-worth</td>
<td>full profile, 27 profiles of 12 attributes</td>
<td>fractional</td>
<td>preference ranking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>MONANOVA</td>
<td></td>
</tr>
<tr>
<td>Anderson &amp; Donthu 1988</td>
<td>part-worth</td>
<td>full profile, 12 profiles of 3 attributes</td>
<td>full</td>
<td>preference rating</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>O.L.S.</td>
<td></td>
</tr>
<tr>
<td>Boecker &amp; Schweikl 1988</td>
<td>part-worth</td>
<td>full profile, individualized attribute sets, single attributes</td>
<td>fractional</td>
<td>graded pair-comparison, preference ranking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>O.L.S.</td>
<td></td>
</tr>
<tr>
<td>Green et al. 1988a</td>
<td>part-worth</td>
<td>full profile, 32 profiles of 7 attributes</td>
<td>fractional</td>
<td>preference rating</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>O.L.S.</td>
<td></td>
</tr>
<tr>
<td>Green et al. 1988b</td>
<td>part-worth, self-explained</td>
<td>full profile, 32 profiles, single attributes</td>
<td>fractional</td>
<td>preference rating</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>O.L.S.</td>
<td></td>
</tr>
<tr>
<td>Krishnamurti 1988</td>
<td>part-worth</td>
<td>full profile, 28 profiles of 6 attributes</td>
<td>fractional</td>
<td>preference ranking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>LINMAP IV</td>
<td></td>
</tr>
<tr>
<td>Reibstein et al. 1988</td>
<td>part-worth</td>
<td>full profile, 25 profiles of 6 attributes, trade-off matrices</td>
<td>fractional</td>
<td>preference ranking, choose preferred profile, ranking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>factorial</td>
<td>O.L.S.</td>
<td></td>
</tr>
<tr>
<td>First Author and Year</td>
<td>Type 1</td>
<td>Type 2</td>
<td>Model Type</td>
<td>Measurement</td>
<td>Methodology</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Srinivasan 1988</td>
<td>part-worth</td>
<td>self-explained</td>
<td>trade-off &amp; attributes</td>
<td>fractional factorial</td>
<td>textual</td>
</tr>
<tr>
<td>Darmon &amp; Rouziès 1989</td>
<td>part-worth, trade-off</td>
<td>full profile</td>
<td>full profile</td>
<td>full factorial, fractional factorial</td>
<td>no</td>
</tr>
<tr>
<td>Green &amp; Heisen 1989</td>
<td>part-worth</td>
<td>full profile, 18 profiles of 6 attributes, single attributes</td>
<td>fractional factorial</td>
<td>textual</td>
<td>preference rating</td>
</tr>
<tr>
<td>Mishra et al. 1989</td>
<td>part-worth</td>
<td>full profile, 16 or 32 profiles of 5 or 8 attributes</td>
<td>fractional factorial</td>
<td>no</td>
<td>LINMAP MONANOVA O.L.S.</td>
</tr>
<tr>
<td>Wittink et al. 1989</td>
<td>part-worth</td>
<td>full profile, 25 profiles of 6 attributes, trade-off matrices</td>
<td>fractional factorial</td>
<td>textual</td>
<td>preference ranking, choose preferred profile</td>
</tr>
<tr>
<td>Woodside &amp; Pearce 1989</td>
<td>part-worth</td>
<td>full profile, 9 profiles of 4 attributes</td>
<td>fractional factorial</td>
<td>textual</td>
<td>constant sum scale</td>
</tr>
<tr>
<td>Darmon &amp; Rouziès-Ségala 1990</td>
<td>part-worth</td>
<td>trade-off</td>
<td>full profile</td>
<td>full factorial and fractional factorial</td>
<td>no</td>
</tr>
<tr>
<td>Moore &amp; Holbrook 1990</td>
<td>part-worth</td>
<td>full profile, 18 profiles of 5 attributes</td>
<td>fractional factorial</td>
<td>textual</td>
<td>preference rating</td>
</tr>
<tr>
<td>Argawal &amp; Green 1991</td>
<td>part-worth</td>
<td>full profile, 18 profiles of 6 attributes, single attributes</td>
<td>fractional factorial</td>
<td>textual</td>
<td>preference ratings, graded pair comparisons, preference ranking</td>
</tr>
<tr>
<td>Huber et al. 1991</td>
<td>part-worth</td>
<td>full profile, ? profiles of 5 or 9 attributes</td>
<td>fractional factorial</td>
<td>textual</td>
<td>preference ratings, graded pair-comparisons</td>
</tr>
</tbody>
</table>
### Appendix 4: Studies on the reliability of conjoint-analysis results

#### Empirical studies on the stability over time

<table>
<thead>
<tr>
<th>Respondents</th>
<th>measures of reliability</th>
<th>interval period</th>
<th>product</th>
<th>outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parker &amp; Srinivasan 1976</td>
<td>8 from set of 77 households new set of 25 profiles from master design, Pearson correlations on partworths</td>
<td>2 months</td>
<td>rural primary health care system</td>
<td>all correlations stat. sign.</td>
</tr>
<tr>
<td>Acito 1977</td>
<td>19 undergraduate students Spearman rankorder correlations on rankings, Kendalls coefficient of concordance</td>
<td>6 subsequent days</td>
<td>instant picture cameras</td>
<td>median .89 to 1.0, mean .86 to .97, W stat sign for 19 respondents</td>
</tr>
<tr>
<td>McCullough &amp; Best 1979</td>
<td>100 students same profiles Spearman rank order correlations on rankings, mean absolute difference of attribute levels</td>
<td>2 days</td>
<td>apartments soft drinks</td>
<td>median .925 and .94, mean .78 and .88, 1 part-worth stat. sign.</td>
</tr>
<tr>
<td>Cattin and Weinberger 1980</td>
<td>41 male college juniors and seminars new set of 16 profiles from master design Pearson correlations on importances</td>
<td>2 days</td>
<td>men’s clothing store</td>
<td>averages from .69 (9,16) to .83 (6,16)</td>
</tr>
<tr>
<td>Leigh et al. 1981</td>
<td>52 undergraduate students same tasks and stimuli, correlations of partworths and importances</td>
<td>1 month</td>
<td>pocket cameras</td>
<td>average adj. correlations .00 to .88 (p-w) and .79 to .99 (imp.)</td>
</tr>
<tr>
<td>Segall 1982</td>
<td>46 students same tasks and stimuli, Spearman correl. of rankings, and importances, paired t-test for mean diff. in importances</td>
<td>7 - 10 days</td>
<td>apartments</td>
<td>mean (Sp.) correlation .84, mean (P.) corr. from .345 (trade-off) to .87 (full-prof.) not stat. sign.</td>
</tr>
<tr>
<td>Study / Authors</td>
<td>Sample Size</td>
<td>Tasks and Stimuli</td>
<td>Time</td>
<td>Method</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Leigh et. al. 1984</td>
<td>122 undergraduate students</td>
<td>same tasks and stimuli, correlations of raw data, estimated stimulus utilities, partworths, mean squared diff. in importances</td>
<td>2 weeks</td>
<td>hand-held calculators</td>
</tr>
<tr>
<td>Scott &amp; Keiser 1984</td>
<td>9 plant managers and 9 corporate officers</td>
<td>test-retest relations between judgments of replicated profiles</td>
<td>one shot</td>
<td>poultry processing</td>
</tr>
<tr>
<td>Reibstein et al. 1988</td>
<td>600 inhabitants of major U.S. city</td>
<td>different stimulus set or attribute set, F-test of communality of partworth vectors</td>
<td>one session</td>
<td>color television typewriter yoghurt banking service long distance service</td>
</tr>
</tbody>
</table>
### Empirical studies on the stability over method

<table>
<thead>
<tr>
<th>Respondents</th>
<th>measures of reliability</th>
<th>product</th>
<th>outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green et al. 1972</td>
<td>43 housewives comparisons with PREFMAP, MDPREF, self-explicated model</td>
<td>discounts</td>
<td>part-worth functions are linear, all 3 variables in m.d. space, Average Spearman .85</td>
</tr>
<tr>
<td>Scott &amp; Wright 1976</td>
<td>26 purchasing agents and 27 consulting engineers self-explicated model and derived part-worths, correlations</td>
<td>electrical resistors</td>
<td>correlations .19 - .80</td>
</tr>
<tr>
<td>Oppedijk van Veen &amp; Beazley 1977</td>
<td>600 respondents comparison of partworths, importances (Pearson) corr., relations between observed and predicted ranks</td>
<td>?</td>
<td>p.w: same pattern, Pearson .79 - .99, Spearman .85 - .99</td>
</tr>
<tr>
<td>Heeler et al. 1979</td>
<td>98 students of business comparison of conjoint analysis with self-explicated model and information display board (pattern of attribute importance ranks and intermethod correlations)</td>
<td>electric blenders</td>
<td>different pattern, corr .32 - .58</td>
</tr>
<tr>
<td>Jain et al. 1979</td>
<td>212 consumers comparison of attribute importance ranks</td>
<td>checking account</td>
<td>different patterns</td>
</tr>
<tr>
<td>Leigh et al. 1981</td>
<td>52 undergraduate students comparison with dollar-metric estimates of part-worths</td>
<td>pocket cameras</td>
<td>correlations .11 - .77</td>
</tr>
<tr>
<td>Wittink et al. 1982</td>
<td>161 MBA students regression analysis to explain variance between attribute importances by different methods</td>
<td>summer jobs</td>
<td>stimulus set and data collection method sign betas</td>
</tr>
<tr>
<td>Akaah &amp; Korgaonkar 1983</td>
<td>80 residents of metropolitan area rankorder correlations between importances by different methods</td>
<td>health maintenance organization</td>
<td>Spearman .075 - .94</td>
</tr>
<tr>
<td>Scott &amp; Keiser 1984</td>
<td>10 office builders comparing results of conjoint procedure and discrete choice analysis</td>
<td>heating system</td>
<td>correlation of .97</td>
</tr>
<tr>
<td>Jaccard et al. 1986</td>
<td>110 female undergraduates of psychology comparison of conjoint analysis with: elicitation, information search, importance ratings subj. prob. measures, Thurstone</td>
<td>car birth control method</td>
<td>correlations birth control .08 - .82, cars .08 - .68</td>
</tr>
<tr>
<td>Study</td>
<td>Measures</td>
<td>Results/Findings</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Wittink et al. 1989</td>
<td>comparison of attribute importance ranks of attributes with varying levels</td>
<td>5 products differences because of: data collection method, and no. of attribute levels</td>
<td></td>
</tr>
<tr>
<td>Synthetical studies on the accuracy: Monté-Carlo studies</td>
<td>measure of reliability outcomes</td>
<td>Tau correlations from .64 - .82</td>
<td></td>
</tr>
<tr>
<td>Carmone et al. 1978</td>
<td>average Tau correlations between original and recovered part-worths</td>
<td>mean absolute error 1.44 - 10.05</td>
<td></td>
</tr>
<tr>
<td>Darmon &amp; Rouziès 1989</td>
<td>difference between true and recovered attribute importance weights</td>
<td>significant in ANOVA: algorithm, evaluation strategy, number of attributes, and 2 interaction terms</td>
<td></td>
</tr>
<tr>
<td>Mishra et al. 1989</td>
<td>ANOVA on bias of the estimated attribute importance weights</td>
<td>in full factorial designs: OLS and MONANOVA provide good recoveries,</td>
<td></td>
</tr>
<tr>
<td>Darmon &amp; Rouziès- Ségalla 1990</td>
<td>ANOVA on bias of the estimated attribute importances in fract. fact. des.: problems with all algorithms</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix 5: Studies on the concurrent validity of conjoint-analysis results

### Empirical studies on the concurrent validity

<table>
<thead>
<tr>
<th>Respondents</th>
<th>measures of validity</th>
<th>product</th>
<th>outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson 1974</td>
<td>hold-out matrix (correlations between observed and predicted rankings), hold-out</td>
<td>?</td>
<td>average value of Tau:.78 (matrix), median of</td>
</tr>
<tr>
<td></td>
<td>profiles, predicted first choice votes (hits)</td>
<td></td>
<td>correlations .80 (profiles), 45% hits</td>
</tr>
<tr>
<td>Huber 1975</td>
<td>Hold-out profiles, correlations between actual and predicted preference scales</td>
<td>iced tea</td>
<td>root mean square of correlations:.79 in</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>metric additive model</td>
</tr>
<tr>
<td>Scott &amp; Wright 1976</td>
<td>goodness of fit</td>
<td>electrical</td>
<td>R² between .5 and .9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>resistors</td>
<td></td>
</tr>
<tr>
<td>Oppedijk van Veen &amp;</td>
<td>correlations between observed and predicted ranks (goodness of fit)</td>
<td>?</td>
<td>Spearman:.85 - .99</td>
</tr>
<tr>
<td>Beazley 1977</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huber &amp; Moore 1979</td>
<td>correlations between observed and predicted highest evaluation (choice)</td>
<td>snowmobiles</td>
<td>Pearson .71 - .73</td>
</tr>
<tr>
<td>Jain et al. 1979</td>
<td>Hold-out profiles correlations between observed and predicted ranks, correct</td>
<td>checking</td>
<td>Spearman corr..71 to .78, 16% to 96% (lo-west</td>
</tr>
<tr>
<td></td>
<td>predictions of holdout ranks (under which first choice), correlations between</td>
<td>account</td>
<td>rank), between 39% and 51%, Spearman:.71 -.78</td>
</tr>
<tr>
<td></td>
<td>observed and predicted ranks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acito &amp; Jain 1980</td>
<td>badness of fit, prediction of hold-out importance ranks (absolute values of</td>
<td>health</td>
<td>stress .135 - .678, hold-</td>
</tr>
<tr>
<td></td>
<td>predicted minus actual rank</td>
<td>maintenance</td>
<td>predictions better than</td>
</tr>
<tr>
<td></td>
<td></td>
<td>organizations</td>
<td>null model</td>
</tr>
<tr>
<td>Cattin &amp; Weinberger 1980</td>
<td>goodness of fit, prediction of ratings in hold-out set.</td>
<td>men's clothing</td>
<td>adj. R²from .76 (6,32) to .92 (6,16) average</td>
</tr>
<tr>
<td></td>
<td></td>
<td>store</td>
<td>Pearson:.66 (9,16) to .80 (6,16)</td>
</tr>
<tr>
<td>Source</td>
<td>Sample Size</td>
<td>Measure</td>
<td>Methodology</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Moore 1980</td>
<td>87 graduate students</td>
<td>Hold-out profiles</td>
<td>Cars</td>
</tr>
<tr>
<td>Wright &amp; Kriewel 1980</td>
<td>161 families with high school juniors</td>
<td>Mean predicted ranks determined on differences between observed and predicted ranks</td>
<td>Colleges</td>
</tr>
<tr>
<td>Green et al. 1981</td>
<td>404 physicians</td>
<td>Goodness of fit, correct first choice predictions, average Tau and average Pearson correlations of actual and predicted evaluations</td>
<td>Antibiotic drug</td>
</tr>
<tr>
<td>Srinivasan et al. 1981</td>
<td>106 automobile commuters</td>
<td>Goodness of fit</td>
<td>Work-trip transport</td>
</tr>
<tr>
<td>Green et al. 1982</td>
<td>476 respondents</td>
<td>Hold-out concepts, Tau correlations</td>
<td>Household appliance</td>
</tr>
<tr>
<td>Moore &amp; Holbrook 1982</td>
<td>87 respondents of two MBA classes</td>
<td>Goodness of fit, hold-out concepts</td>
<td>Dogs</td>
</tr>
<tr>
<td>Akaash &amp; Korgaonkar 1983</td>
<td>80 residents large midwestern metropolitan area</td>
<td>Hold-out concepts, number of correct predicted rankings, correlations between observed and predicted evaluations in calibration set</td>
<td>Health maintenance organization</td>
</tr>
<tr>
<td>Srinivasan et al. 1983</td>
<td>106 full profile, 115 trade-off</td>
<td>Hold-out profiles (Spearman corr.), percentage first choice correct predicted</td>
<td>Checking accounts</td>
</tr>
<tr>
<td>Scott &amp; Keiser 1984</td>
<td>10 office builders, 9 plant managers, 9 corporate officers</td>
<td>Goodness of fit</td>
<td>Heating system</td>
</tr>
<tr>
<td>Klein 1986</td>
<td>120 university students</td>
<td>Choice from hold-out set (proportion of correctly predicted choices)</td>
<td>Pens</td>
</tr>
<tr>
<td>Malthotra 1986</td>
<td>118 respondents</td>
<td>Hold-out concepts (Spearman correlations between predicted and observed rankings of hold-outs)</td>
<td>In-home energy audit</td>
</tr>
<tr>
<td>Anderson 1987</td>
<td>25 members of marketing management, 11 suppliers, 14 customers</td>
<td>Goodness of fit</td>
<td>Clothing product</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Measure of Fit</td>
<td>Product</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Anderson &amp; Donthu 1988</td>
<td>307 individuals of 6 metropolitan areas</td>
<td>goodness of fit</td>
<td>clothing</td>
</tr>
<tr>
<td>Boecker &amp; Schweikl 1988</td>
<td>200 respondents (convenience)</td>
<td>goodness of fit, hold-out concepts, first choice predictions</td>
<td>videorecorders</td>
</tr>
<tr>
<td>Green et al. 1988a</td>
<td>120 business students</td>
<td>goodness of fit, hold-out concepts: Pearson correlations, root mean square error, Spearman correlations, first choice hits</td>
<td>apartments</td>
</tr>
<tr>
<td>Green et al. 1988b</td>
<td>160 business students</td>
<td>goodness of fit, hold-out concepts (Pearson correlations)</td>
<td>apartments</td>
</tr>
<tr>
<td>Krishnamurti 1988</td>
<td>MBA's and spouses</td>
<td>goodness of fit</td>
<td>jobs</td>
</tr>
<tr>
<td>Green &amp; Heise 1989</td>
<td>99</td>
<td>hold-out concepts: Pearson corr., mean square error, first choice hits</td>
<td>apartments</td>
</tr>
<tr>
<td>Moore &amp; Holbrook 1990</td>
<td>62 MBA students more than 10 years driving experience</td>
<td>goodness of fit, hold-out concepts (root mean squared error, correlations)</td>
<td>cars</td>
</tr>
<tr>
<td>Argawal &amp; Green 1991</td>
<td>170 undergraduate business students</td>
<td>hold-out concepts (correlations, first choice hits, correct hits for all ranks)</td>
<td>apartments</td>
</tr>
<tr>
<td>Huber et al. 1991</td>
<td>393 respondents from 11 cities</td>
<td>choose from hold-out concepts, hits</td>
<td>refrigerators</td>
</tr>
</tbody>
</table>

**Synthetic studies on the concurrent validity: Monte-Carlo**

<table>
<thead>
<tr>
<th>Study</th>
<th>Measure of Concurrent validity</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carmone &amp; Green 1981</td>
<td>Hold-out set (correlations and first choice predictions)</td>
<td>Correlations: .60 - 1.0, hits: 13% - 100%</td>
</tr>
</tbody>
</table>
## Appendix 6: Studies on the predictive validity of conjoint-analysis results

### Empirical studies on the predictive validity

<table>
<thead>
<tr>
<th>Respondents</th>
<th>measures of validity</th>
<th>product</th>
<th>outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parker &amp; Srinivasan 1976</td>
<td>177 households in rural U.S. community</td>
<td>prediction of present physician</td>
<td>physicians</td>
</tr>
<tr>
<td>Green et al. 1978</td>
<td>54 MBA students</td>
<td>prediction of preferred vacation site (Tau correlation between predicted and actual vacation site rankings)</td>
<td>vacation sites</td>
</tr>
<tr>
<td>Wittink &amp; Montgomery 1979</td>
<td>153 MBA students (48 in final set)</td>
<td>prediction of choice from job offers, Spearman between predicted and actual rankings</td>
<td>jobs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wright &amp; Kriewal 1980</td>
<td>161 families with school juniors</td>
<td>hit rates, predictions of actual choice</td>
<td>colleges</td>
</tr>
<tr>
<td>Srinivasan et al. 1981</td>
<td>106 automobile commuters</td>
<td>prediction of actual behavior transit or auto</td>
<td>work trip transport</td>
</tr>
<tr>
<td>Moore &amp; Holbrook 1982</td>
<td>67 respondents of 2 MBA classes</td>
<td>preference rating of 15 real breads</td>
<td>dogs</td>
</tr>
<tr>
<td>Leigh et al. 1984</td>
<td>122 undergraduate students</td>
<td>raffle choices: percentage correct predictions, corrected for chance, mean squared differences between estimated utilities and choices</td>
<td>hand-held calculators</td>
</tr>
<tr>
<td>Page &amp; Rosenbaum 1987</td>
<td>more than 500</td>
<td>introduction of new product ideas on the market</td>
<td>food processor</td>
</tr>
<tr>
<td>Anderson &amp; Donthu 1988</td>
<td>307 individuals of 6 metropolitan</td>
<td>coupon choice task (confusion matrix of predicted and actual coupon choice)</td>
<td>clothing product percentage hits 50.4%</td>
</tr>
<tr>
<td>Study</td>
<td>Group Description</td>
<td>Type of Prediction</td>
<td>Actual Job</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Krishnamurti 1988</td>
<td>MBA’s and spouses</td>
<td>Prediction of actual job chosen</td>
<td>jobs</td>
</tr>
<tr>
<td>Srinivasan 1988</td>
<td>85 MBA study and spouses</td>
<td>Prediction of choice from job offers</td>
<td>jobs</td>
</tr>
<tr>
<td>Woodside &amp; Pearce 1989</td>
<td>48 operations support and technical support personnel</td>
<td>New company position in different market segments after introduction</td>
<td>furnace cleaning, modest improvement of service, company position in 3 out of 4 segments</td>
</tr>
</tbody>
</table>