Consumer Search for Information on the Internet

Ph.D. dissertation

Jepsen, Anna Lund

Publication date:
2005

Document version:
Final published version

Citation for pulished version (APA):
Jepsen, A. L. (2005). *Consumer Search for Information on the Internet: Ph.D. dissertation*. Syddansk Universitetsforlag.

Consumer Search for Information on the Internet

by

Anna Lund Jepsen

Ph.D. thesis in Marketing
Faculty of Social Sciences
Institute for Environmental and Business Economics
University of Southern Denmark
2003

Contents

1. Introduction 7

1.1 Methodological and Theoretical Approach 9
1.2 Definitions and Delimitations 19
1.3 Outline of the Thesis 21

2. Consumer Search for Information 24

2.1 Search for Information as a Part of the Buying Process 25
2.2 Determinants of Level of Consumer Information Search 29
2.2.1 Search in Relation to an Upcoming Purchase 29
2.2.2 Explanations for Lack of Search before Purchase 37
2.2.3 Conclusions on Consumer Search Behaviour 40
2.3 The Internet as One Single Source of Information 41
2.3.1 A Place for Exchange of Consumer Information 41
2.3.2 Research into Consumer Behaviour on the Internet 47
2.4 The Internet as Several Types of Sources of Information 53
2.4.1 Individualized, Non-marketer Dominated Sources 57
2.4.2 Non-individualized, Non-marketer Sources 68
2.4.3 Individualized, Marketer-Dominated Sources 70
2.4.4 Non-individualized, Marketer Dominated Sources 73
2.4.5 Expected Use of the Sources on the Internet 74
2.5 Determinants of Choice of Information Source 76
2.5.1 The Decision and Tie to the Source 77
2.5.2 Differences in Costs of Information Search 81
2.5.3 Type of Product 84
2.5.4 Facilitating Factors 86
2.5.5 Expectations About Use of Internet Sources 91
2.6 Hypotheses about Consumer Use of the Internet 93

3. Research Design 105

3.1 Preliminary Plan for Data Analysis 108
3.2 Issues in the Development of a Measurement Model 110
3.3 Variables Related to Internet Use on the General Level 112
3.3.1 Setting the Context for the Respondent 113
3.3.2 Use of the Internet as an Information Source 115

3.3.3 Knowledge about and Interest in the Product 117
3.3.4 Perceived Cost of Search Using the Internet 118
3.3.5 The Shopping Aspect 119
3.3.6 Expected Usability of Information on the Internet 121
3.3.7 Importance of the Internet 122
3.3.8 Facilitating Factors 123
3.3.9 Background Variables 125
3.3.10 Summing Up Model Operationalization 126
3.4 Variables Relating to Use of Newsgroups 132
3.4.1 Setting the context 132
3.4.2 Determination of Segment Membership 132
3.4.3 Motivation factors 135
3.4.4 Communication and Perceived Influence 135
3.4.5 Opinion-leadership and -seekership 136
3.4.6 Substitution of Traditional Information Sources 138
3.4.7 Summarizing Questions for Newsgroup Study 139
3.5 Survey Method and Questionnaire Design 143
3.6 Sampling and Contacting Respondents 150
3.6.1 The Study on Internet use on the general level 151
3.6.2 The study in Newsgroups on the Internet 157

4. Results Concerning Internet Users in general 163
4.1 Validation and Initial Description of the Sample 163
4.1.1 Sampling Results 164
4.1.2 Socio-economic Profile and Representativeness 169
4.1.3 Products Bought 172
4.1.4 Product Search, Interest and Knowledge 173
4.1.5 Internet Use on the General Level 177
4.1.6 Use of the Internet for Information Search 182
4.1.7 Results on Facilitating Factors 191
4.1.8 Attitudes towards Shopping 194
4.1.9 Summing Up Initial Description and Validation 196
4.2 Validation of the Measurement Model 198
4.2.1 Initial Validation of the Measurement Scales 199
4.2.2 Validation of the Measurement Model in Total 204
4.3 Analysis of Structural Equations 217
4.4 Determinants of Use of the Internet for Information Search 230

5. Results from the Newsgroup Study 234
5.1 Size of and Communication within the Newsgroups 237
5.2 Evaluation of the Sample 240
5.3 Description of the Newsgroup Users 246
5.3.1 Description of the Sample on Demographic Variables 247
5.3.2 Interest in and Knowledge about the Product 248
5.3.3 Importance of the Internet in the Consumer's Life 249
5.3.4 Relation to the Selected Newsgroups 250
5.3.5 Membership Segments in the Sample 252
5.3.6 Summary of the Description of the Newsgroups 253
5.4 Motivation for Entering the Newsgroup 254
5.4.1 Model of Motivation for Entering the newsgroup 255
5.4.2 Explaining Differences in Motivation for Entering 258
5.4.3 Different Motives for Different Segments? 262
5.5 Opinion Leadership and Communication 263
5.5.1 Opinion Leaders in the Study 264
5.5.2 Test of the Measurement Model 267
5.5.3 Explaining Interaction in the Newsgroup 270
5.5.4 Conclusions on Communication Flow in Newsgroups 278
5.6 Replacement of Other Sources of Consumer Information 279
5.7 Conclusion on Findings on Consumer Search in Newsgroups 287

6. Limitations and Issues for Further Research 290

7. Conclusion and Implications for Marketing 299

8. Dansk Sammenfatning 307

References 321

List of Figures 337

List of Tables 338

Appendices can be downloaded from the Internet at:
http://www.sam.sdu.dk/~alj/appendices_consumer_behaviour_internet.pdf

1. Introduction

There is much roar about the Internet in the media. Some propose that the Internet will take over all trade in a few years while, at the same time, others argue that the Internet contains nothing really new when it comes to marketing, and that most of the belief in its glorious future is founded in thin air.

Regardless who owns the truth, the diffusion of access to the Internet across society has given consumers new possibilities for finding the information they need in order to make a good choice between alternative (consumption) solutions to their problems. On the Internet, consumers can easily get in personal contact with other consumers living far away whom they might not even know in a traditional sense. Some of these meetings take place in new types of groups, the so-called virtual communities which have emerged on the Internet. In these communities, consumers meet virtually – that is, not physically. In addition to easier personal contact, consumers can also more easily visit companies (at their websites) than they are able to in the physical marketplace as they are free from replacing themselves physically to do so. On account of the same reason, the Internet has given more consumers access to alternative information sources like consumer groups and state institutions. In addition, these organizations can publish their knowledge on the Internet at a smaller cost than the cost of publishing on print using traditional distribution methods. This gives these, often not wealthy, organizations ability to publish more material, thereby enlarging the supply of this kind of information, available to the consumer.

Research has shown that the possibility of information search for information does exist, not only in theory, but also in reality for the 'ordinary' consumer. One study revealed that 76% of American households on-line seek information related to a hobby, and 16% have looked for information concerning a product before purchase (Pew Internet & American Life Project, 2000). Another study has shown that 13% of the on-line population in Denmark has looked for information on the Internet prior to making a purchase off-line (Taylor Nelson Soffres, 2000). Data from Georgia University of Technology (1998) shows that 71% of

the respondents search for information on the Internet on products in general and prior to making a purchase. This number is largest for younger people and for experienced Internet users. Data from the same source say that 35.2% of Internet users make buying decisions that are based primarily on information gathered on the Internet once or twice a month, and also that for half of the respondents, this was the case for most of the buying decisions made.

The above shows that consumers do use the Internet for information search. However, the data does not tell why consumers use the Internet for information search or which sources they use, when searching on the Internet. There has been done some research into this area. Yet, most of the empirical findings originate from commercial sources and are descriptive. If the empirical findings are based on non-commercial research, the results have mostly been based on samples drawn by self-selection among users of specific websites. Moreover, the majority of the data are on American consumers and not on Danish, or for that matter, European consumers. Finally, the results tell only very little about use of the various sources of information on the Internet. Especially, there is only a limited amount of research into consumers' use of virtual communities as a place for exchanging consumer information as most research into virtual communities does not focus on consumer behaviour.

From a marketing perspective, it is important to understand the determinants of consumers to take advantage of the possibilities for information search using the Internet in order to discern the importance of presence on the Internet and what this presence should consist of. Of special interest is whether consumers use the new types of communication which have occurred on the Internet when searching for consumer information, and if they do, to what extent these new possibilities substitute sources, formerly used. If some of the new types of communities have actually replaced traditional sources like family and friends, presence in or access to these communities is interesting to companies wanting to know and communicate with their customers.

From a research perspective, it is important to discover whether generally accepted theories about determinants of consumer search behaviour are applicable to consumers using the Internet for information search or if the theories need to be modified as a result of the new possibilities available to consumers. Especially, virtual communities are interesting as a special type of information source from a research perspective. These communities may act as a reference group for consumers seeking information on the Internet. If this is the case, the list of reference groups needs adjustment. Of special interest is to know what kind of reference group the virtual community is perceived as by the consumer.

The research question thus is twofold:

1: What are potential determinants of consumer use of the Internet for in formation search in general and

2: What is the significance of virtual communities in consumer search for information?

To answer these questions, this thesis presents and discusses theories on consumer behaviour in general and consumer search behaviour in detail. The discussion is aimed at resulting in a set of hypotheses about consumers' use of the Internet as an information source. In order to investigate this topic, it is necessary first to clarify how, and under what circumstances, consumers can be expected to search for information and secondly, what type of source he uses, when searching for information. However, before starting this process, it is necessary to clarify the rules under which this research project is carried out. Therefore, the following chapters are devoted to choice of research paradigm and approach to consumer behaviour followed by important definitions and delimitations of the research area. Finally, this introductory chapter concludes by an outline of the thesis in chapter 1.3.

1.1 Methodological and Theoretical Approach

The research paradigm governing this thesis is neopositivism. The goal is, therefore, to find law like answers to the research questions. These

answers are in the form of hypotheses built on deductions from generally accepted theory. The methodological standards of neopositivism are to a large extent the same as in positivism. This means that in the empirical study the aim is to measure as precisely and as objectively as possible even though the objects measured are attitudes and reported behaviour and not physical phenomena, and the measuring instrument, therefore, has to rely on communication with persons about their opinions and what they are able to remember about their behaviour instead of refined technical devices measuring objectively observable phenomena. The remedy to the frail measurement situation is to design the empirical study in such a way that all responses (measurements) are obtained in the same way using an, as precise as possible, measurement instrument in a setting chosen to balance wishes for internal and external validity of the study. Because the research paradigm is neopositivism, and not positivism, external validity is considered quite important (Guba, 1990).

As a result of the chosen paradigm, the task in this thesis is to build hypotheses which are testable and to design an empirical study which uses valid measurement scales, measuring in a representative sample from the population, to challenge the stated hypotheses. In everyday terms this normally implies that the empirical research preferably is carried out as an experiment or a survey using a questionnaire with mainly closed-end answering formats. As the hypotheses have to be testable by the methodology used, it is appropriate to consider possible methodological approaches at this stage although it is early in the process.

The survey is considered to be the best option in this case because, in the survey, we are able to ask questions about attitudes and behaviour to discover reasons behind using or not using the Internet for information search. An alternative to the survey is the experiment or a combination of experiment and survey. However, I wish to cover the entire information search process. This process can evolve over quite a long time span. Therefore, it would be difficult, if not impossible, to design an experiment aimed at testing hypotheses concerning this process. To carry out such an experiment would imply to find consumers, starting their buying process and logging all behaviour on their computer in the entire

buying process. Such a study would have high internal validity in terms of measuring exact behaviour instead of relying on reported or intended behaviour. However, the study would probably have low external validity for at least two reasons. First, the consumers in the study would be aware that they were being watched while searching for information and their behaviour might be influenced by this. This is a problem in generalizing from this behaviour to everyday behaviour. Second, the sample would probably be quite small and far from representative and, therefore, there would be problems in generalizing from the experiment to the average Internet user. The conclusion to the above considerations is that a survey seems to be the best option in this case even though the survey entails problems in terms of internal validity because the measurements are of past or intended behaviour and not of actual behaviour.

There are, of course, other ways of defining 'good' research than the standards of neopositivism. Some may argue that it is naïve to think that it is possible to find generally applicable truths and also to believe that it is possible to objectively test generally applicable hypotheses about human beings as there is no 'true reality' but only a perception of reality and, therefore, multiple realities which are, in fact, social constructions (Guba, 1990).

However, it is my conviction that it is possible to find truths which are generally applicable (to a certain extent), and also to measure abstract constructs by asking predefined questions to human beings about their attitudes and behaviour on a general level. I also believe that it is a sound goal for a researcher to try and discover these general truths and describe them in models and theories which are testable and communicable to others. At the same time, I recognise that this rather firm way of doing research is not well suited for studies wanting to dig deep into the psychological reasons for behaviour. I also agree with Guba (1990), who points to the imbalance between discovery and verification. This imbalance is the result of the demand of positivism that research is to challenge what is in some way already known and, therefore, does not bring anything new (discovery). Neopositivism, however, allows discoveries along the way. Therefore, I have been open to new findings and ideas in my work on this thesis. An important part in this is to consider

findings from different schools of the study of consumer behaviour, even though I base my work mostly on the findings from one school.

The result of the above is that, in the aim to become wiser on consumer search behaviour using the Internet, the work in this thesis follows the rules for research governed by neopositivism with an open mind towards other ways of thinking.

After having clarified choice of scientific paradigm on a general lever, we now turn to the scientific perspective on consumer behaviour. In the 80'ies and the 90'ies a scientific crisis has arisen resulting in several competing paradigms to choose from. This crisis in research in consumer behaviour means that theoretical developments on consumer behaviour in general and, as a part of this, also research into consumer behaviour in relation to the Internet evolve along separate lines within their different paradigms. Therefore, the competition between different theories about consumer behaviour is not merely a competition between models but also forms part of a more fundamental competition among the different paradigms. Because of this, it is necessary to shortly present, evaluate, and choose from the competing paradigms before turning to the study of literature and development of hypotheses.

There are three main approaches to research into understanding, explaining, and/or predicting consumer buying behaviour in general and also in the subfield consumer search behaviour (Jepsen, 2001). At two extremes, we find Radical Behaviourism which is strictly positivistic, and Postmodernism which is constructivistic in its approach to consumer behaviour as a science. Along the same line as Postmodernism is what some researchers have called the Dialectic Approach and others Critical Theory. Between the two extremes lies the Cognitive Approach which rests on the premise that consumer behaviour is a result of cognitive processes rather than involuntary reactions to impulses from the surroundings or chance.

The main characteristics of the three (four) paradigms are presented in figure 1.1 below. In the following, I shall elaborate a bit on the contents of the figure, leading to a choice of path to follow in the study of theory

on consumer search behaviour carried out later in this thesis as a basis for the development of hypotheses.

Figure 1.1: The competing schools: Differences in the view on consumer behaviour as a science

Paradigm / Concept	Radical Behaviourism	Cognitvism	Humanism: Dialectic Approach	Postmodernism
Underlying general scientific paradigm	Positivism	Neopositivism	Critical Theory	Constructivism
Understanding behaviour as	(Involuntary) reaction to stimuli	Cognitive reaction to stimuli	Interaction with surroundings	Erratic
Understanding consumption mostly as	Fulfilling needs according to stimuli	Fulfilling needs after a cognitive evaluation of possible solutions. Satisfaction.	Fulfilling needs according to pressure from material environment	A way, the consumer can express himself by the products and brands, he chooses.
Social context	Not an issue	Important because it limits behaviour. Emphasis on governing norms and resulting consumption	Important because it limits behaviour. Emphasis on contradictions and needed change	Not important because of fragmented life
Theoretical development mainly in the form of	Causal models	Descriptive models	Discussions	Discussions

To the left in figure 1.1, we find Radical Behaviourism. Radical Behaviourism is strictly positivistic. As a consequence of this, research along this path focuses on structured observation of consumer behaviour (what

the consumer does) (Foxall, 1980, 1986), and not consumers' accounts of their behaviour and reasons behind it. The reason for this is that consumption is seen as passive reactions to internal or external stimuli. The consumer, therefore, does not know why he acts as he does. Thus, it has no meaning to ask him questions concerning reasons behind behaviour, as this would only make him rationalize past behaviour.

Cognitive research into consumer behaviour is based on the basic premise that the consumer involves himself in cognitive processing on some level in order to choose the product/brand which will best solve a problem that he has recognized (Engel et al., 2001; Sheth, 1974). This means that consumer search behaviour occurs only after problem recognition. Because the consumer consciously solves problems, he can also account for his actions. Because consumers' choices are based on personality and adjustments to social contexts, their behaviour is expected to be predictable, depending on context. Therefore, in cognitive research, focus has been on finding the elements of, and explaining differences in, the decision process by characteristics in the consumer and his surroundings. Because many consumers live under the same conditions and have personalities that are alike, and behaviour is expected to be governed by (bounded) rationality, it is generally accepted in the cognitive school of marketing research that consumer behaviour can be represented by general models, often in the form of flow-charts.

Although a vast amount of theory and empirical evidence supports the cognitivist school of consumer behaviour there are theorists who believe that knowledge in this field has to be found along a different path, because the cognitive path overlooks important aspects of consumer behaviour by relying too much on consumers' rationality.

This most recent path of research in consumer behaviour is inspired by humanistic research. The researcher who works along this line does not completely reject the theories on cognitive information processing. Nonetheless, researchers along this line feel that cognitive theory on consumer behaviour is, in a sense, antiquated as it belongs to way of life that in many ways no longer exists (the industrial age and the modern

world). These researchers also emphasize that the focus on consumer behaviour as a problem solving activity is too limited in scope.

There are two directions based on humanistic research into consumer behaviour – Postmodernism and The Dialectic Approach. The two directions differ in the aspects deemed the most important although they both originate from mostly the same criticism of cognitivism.

The Dialectic Approach to research in consumer behaviour involves, among others, the work of Hirschman (1993), Holbrook and Hirschman (1982), and Miller (1998a), and is explained in Marsden and Littler (1999). There is a strong link between the Dialectic Approach and critical and Marxist theory (*e.g.* Hirschman, 1993). Dialectical theory considers consumption to be shaped by the *material environment (*Jackson 1999; Miller et al., 1998b). The second concept in dialectical theory is *change.* The consumer interacts with the material world (his environment) and, therefore, his behaviour changes as a result of changes in the environment. The third concept in dialectic theory is *totality.* Totality means that the elements of the world are connected to each other and form a whole different from the sum of the parts. The final concept is *contradiction.* In this context, contradiction is important because change is though to stem from contradictions within a system.

Researchers in consumer behaviour following the Dialectic Approach have mainly focused on shopping as entertainment, and shopping for other reasons than material needs, because they find these areas under-researched. They argue that only a small part of consumer behaviour is concerned with brand choice and, more broadly, with problem solving. Therefore, they find that in order to understand a larger portion of consumer behaviour, research has to engage in consumer behaviour, not oriented towards buying a product intended to solve a problem. Examples of this kind of research are shopping as leisure, gift giving (for one self and others), second hand shopping, and how culture shapes shopping behaviour (Bloch, Ridgeway and Nelson, 1991; Bloch and Richins, 1993). Research along this path also tends to focus on the opposite interests between the marketer and the consumer.

The second direction of research based on humanistic research traditions is Postmodernism. Researchers adhering to this paradigm (*e.g.* Brown, 1997; Firat, Dholakia and Venkatesh, 1995; Venkatesh, 1998, 1999) believe that reality is constructed by the humans who participate in it and, therefore, there is not one single truth about the world to understand and to find general laws about. To achieve a better understanding of consumer behaviour, the postmodern school of marketing research point to the necessity of accepting the following facts about life in postmodern society:

The postmodern researcher emphasises that the world, to the consumer, is *hyperreal.* That is, consisting of the things objectively there plus the perceived symbolic meaning of goods and actions. According to Postmodernists, postmodern life is *fragmented.* Another important theme for researchers belonging to the postmodern school is that production is no longer the most important factor in society. *Consumption* is far more important in life in postmodern society, as individuals spend a large part of their time and, therefore, have knowledge about consuming and spend only little time and have little knowledge about, and interest in, production. As there are no strict rules as to what products should be used together, or under what circumstances certain items are suitable or necessary, products are put together and used, not according to guidelines from others, but according to what the consumer wants to express at that particular moment. Postmodern researchers refer to this as *juxtaposition of opposites*. The last concept is *loss of commitment*.

According to researchers, belonging to the postmodern school, a consequence of the characteristics of consumer behaviour in postmodern society is that it is not possible to predict consumer behaviour on the basis of knowledge of for example personality, lifestyle, and demographics. This is because the consumer is constantly changing – over time but also over roles during the day (life is fragmented) and is not governed by rules about acceptable behaviour (he is not entirely committed to any group). Therefore, consumer behaviour is not predictable (anymore). Furthermore, the consumer does not buy products in order to fulfil needs, but in order to experience and express his self. Therefore, seeing buying solely as problem solving is too narrow. In addition to this, as

there are no rules, as to which experiences, it is acceptable to seek, or which experiences are achievable for any given consumer, it is very difficult to say anything in general about which products will be purchased by whom, and for what use. Also, the fact that the consumer uses products and their symbolic meaning to produce himself (Firat et al., 1995) makes the role of the consumer much more active than is stipulated in the traditional models of consumer behaviour.

All this means that modelling consumer behaviour and trying to predict consumer behaviour in general, explained as rational problem solving is not very interesting and in any case, too limited in scope. Researchers, who work along this line, therefore, are interested in investigating consumers' experiences and the symbolic and cultural meaning of products and behaviour *(the sign system*) rather developing models explaining the choice of product based on physical attributes.

The paradigms presented in this chapter have different views on the consumer and on how behaviour is shaped. Therefore, choice of paradigm has consequences as to how one would try and answer the research questions in this thesis and probably also which questions are asked in the first place, as should be clear from the above presentation of the paradigms. As a consequence of this, it is necessary to clarify, which consumer research paradigm mainly governs the development of hypotheses in the following. Still, the paradigms are seen as supplementary rather than contradictory, at least regarding the issue at hand and, therefore, the choice made in this chapter does not mean that research belonging to other schools of research, is not considered at all while developing the hypotheses.

As stated in the above, Radical Behaviourism is closely connected to positivism and the use of experiments as a way of learning about consumer behaviour. On the Internet, research under this paradigm would focus on click-stream data from websites or log-files from consumers' computers as described in the above during the discussion of using experiment as a methodological option. As put forward earlier, experiments are not considered a fruitful way to go, because the aim in this thesis is to learn about, among other things, attitudes behind the use of

the Internet for information search. As an example of the limitations of relying on observations only, research has shown that although consumers click on only ½-1% of all banners they are exposed to, there is a significant difference in recall if the person has seen a banner-ad, as compared to a person who has not seen a banner ad (Kargaard, 2001). This indicates that consumers actually see banner ads even if they do not react to them instantaneously. Concluding to the above, I do not consider Radical Behaviourism a fruitful way to go in this thesis.

Researchers, belonging to the cognitive school, would be interested in modelling determinants of the choice of the Internet as an information source, based on the premise that the consumer would choose this if it would seem rational to him. This is very much in line with the goal of this thesis which probably reflects the choice of research paradigm (neopositivism) and the theoretical upbringing of the author.

Dialectic researchers would focus on whether consumers use, or should use, the Internet because they feel a contradiction between the lives they lead and how life could be. Because of the changes in society, one might expect that consumers would feel that they benefited from a change in behaviour regarding information search toward using the new possibilities of finding information on the Internet. However, the Internet is not available to all consumers. Therefore, researchers belonging to this school focus on what is termed 'The Digital Divide'. In addition, the focus in this line of research is on contradicting interests between marketer and consumer, and between society and consumer. As the planned research in this thesis does not adhere to these contradictions as being important, it does not fit very well with the Dialectic Approach.

Finally, a Postmodern researcher looking into consumer use of the Internet would focus on Internet culture, the symbolic meaning of the virtual communities, and how the consumer can create his identity through consumption on and off the Internet. This is not the focus of the present study. Therefore, this line of research is also not the main line, which the hypotheses are built on. However, the notions about postmodern world, which are deemed important in this line of research, fit life

on the Internet well and, therefore, I find it important to also consider research done along this line in the development of the hypotheses.

As can be seen from the above, research paradigm and methodology chosen combined with the research question clearly point to the cognitive school of research in consumer behaviour as the paradigm which best fits the approach in this thesis. The choice of the cognitive school of research in consumer behaviour means that the consumer buying process is mainly seen as a problem solving process and that search is seen as a part of this process. It also means that it is a premise that the consumer tries to solve this problem by making the best decision using the minimum amount of resources. However, as is also clear, it is important to consider findings based on the other schools of research in the study of theory, on which the hypothesis are built

Before looking further into theoretical developments on consumer search in general and the use of the Internet for information search, it is necessary to state important delimitations and definitions used in the thesis. This is the topic of the following chapter.

1.2 Definitions and Delimitations

The Internet contains information put there by companies or individuals. The Internet is, therefore, strictly speaking, not an information source but an information medium or channel. However, once the consumer searches for information and he chooses to perform this search using information on the Internet, the Internet can be perceived as the information provider, and thus, in this sense, a source of information. Therefore, in this thesis, the term source of information is used about the Internet and the various types of sources on the Internet.

The scope of the thesis is limited to the information search process related to products for which the consumer engages in extensive – or at least limited problem solving. The aim is to find possible determinants of use of the Internet for information search and influence from newsgroups. The aim is not to discover anything about consumers actually buying on the Internet. Also, the thesis does not treat how the consumer

actually searches the Internet for information. That is, in what sequence he uses the various sources, how many websites, he visits, and for how long, he visits these websites.

The discussion of virtual communities, and especially newsgroups on the Internet, also poses a problem of terminology. Virtual communities can be defined as: 'social aggregations that emerge from the net when enough people carry on … public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace' (Rheingold, cited in Kozinets, 1999) and the question then is if consumers perceive the kind of communities treated in this thesis as qualifying as such. The discussion of this question is postponed until chapter 2.4.1 on individualized, non-marketer dominated sources on the Internet.

Another problem along this line is that, strictly speaking, the consumer subscribes to a newsgroup and then downloads communication from the newsgroup whenever it suits him. However, in papers and articles on the subject the term 'member' is used about the status in the groups subscribed to. As this term suits the subject of this thesis and is also a term commonly used by researchers in the field, this is also the term I am going to use.

As research into exchange of consumer information in virtual communities on the Internet is quite limited, the hypotheses about this part of consumer search will inevitably be built on a theoretical ground, much less firm than theories on search on the Internet on the general level. Therefore, the goal in this part of the research is mainly to find out whether these communities are interesting, from a marketing point of view. That is, if consumers use these groups when searching for information in relation to an upcoming purchase and if the groups substitute or replace traditional sources of information on a level worth noticing.

A problem, often brought up (Hoffman and Novak, 1997; Peterson et al., 1997) regarding consumer's use of Internet is lack of infrastructure and security in transmitting sensitive data. This means that customers face problems and risk in paying for products that they have bought. In

addition to this, they face the risk that the product, they have ordered, never arrives. As the purpose of this thesis is primarily search for information and not purchase as such, the security problems are not considered an important aspect in this thesis and are, therefore, not treated in this thesis. Of course, the consumer can fear that information given to somebody on the Internet – which happens merely by visiting a homepage – ends up in the wrong place. As concerns about this have been found to exist mostly among business users (Korgaonkar and Wolin, 1999), this aspect is also left out of the analysis.

1.3 Outline of the Thesis

As pointed to earlier in this chapter, the aim of this thesis is to develop and test hypotheses relating to the research questions which were posed at the beginning of this chapter. These hypotheses are to be built from deductions from generally accepted theory in the research area.

Chapter 2, therefore, deals with theoretical and empirical evidence on consumer behaviour in general and on the Internet and discusses the Internet as an information source and possible determinants of choice of information source. The evidence presented is mainly based on cognitivism. The result from this chapter is a model of the determinants of the extent, to which the consumer uses the Internet for information search purposes on a general level. The model is supplemented by hypotheses on the use of the virtual communities on the Internet.

In order to find out whether the propositions made in chapter 2 are valid, it is necessary to perform an empirical study. Chapter 3 thus discusses research design with special attention to development of the measurement scales used to measure the concepts in the hypotheses, choice of collection method, and sampling issues.

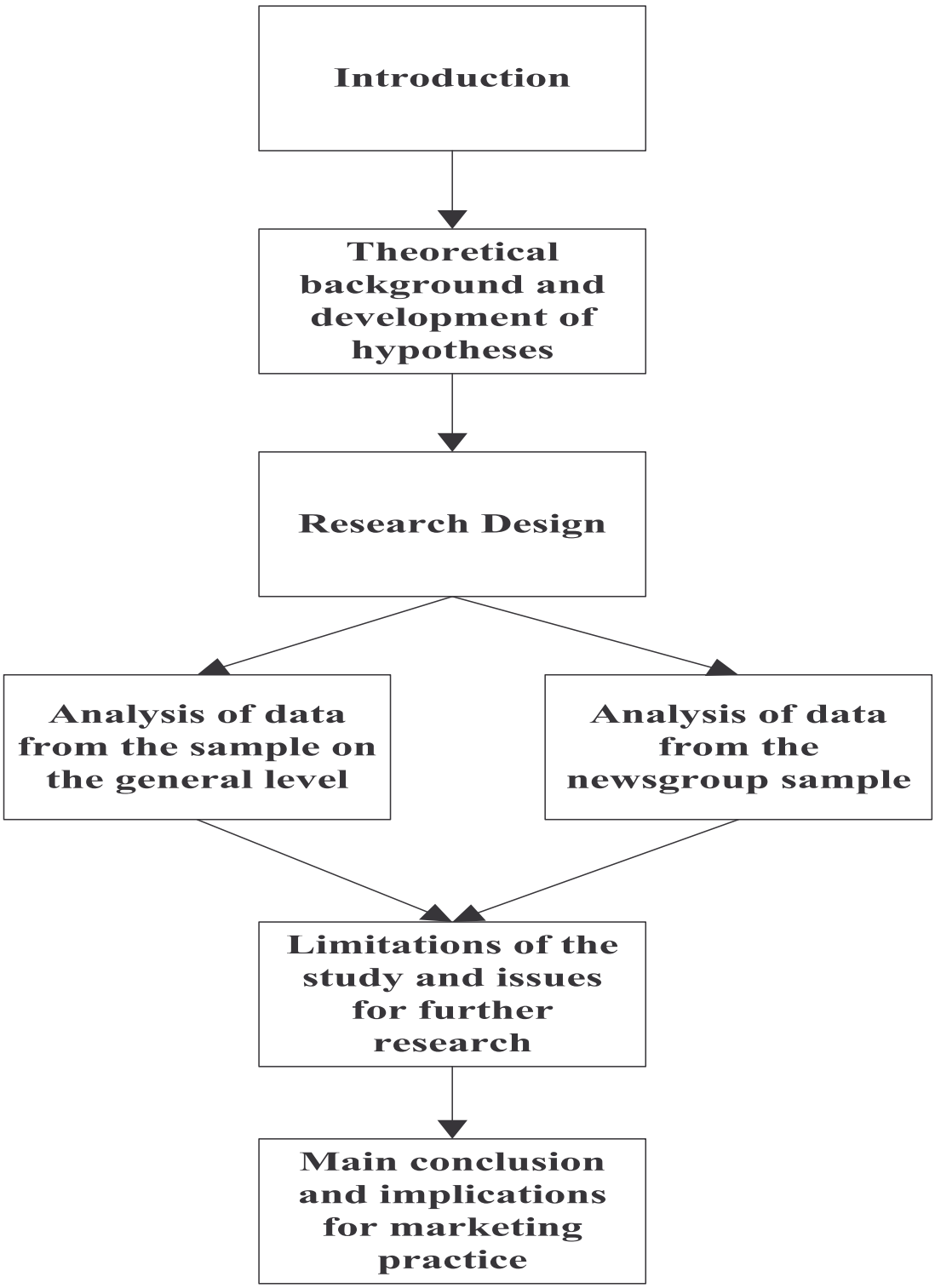
Chapter 4 deals with analysis of the data gathered to test the hypotheses on the use of the Internet for consumer information search on the general level, whereas chapter 5 is devoted to analysis of the data gathered in order to test the hypotheses about the impact of information search in virtual communities.

In chapter 6, the limitations of the theoretical and empirical studies done in this thesis are discussed, and issues for further research are pointed to. Finally, chapter 7 contains the main conclusion and implications for marketing practice.

Chapter 8 is in Danish and contains a summary of the theoretical and empirical research in this thesis.

Figure 1.2 below summarises the contents of this thesis.

Figure 1.2: Overview of the thesis



2. Consumer Search for Information

This chapter is devoted to an investigation into theory and empirical evidence which can help answering the two research questions posed in the introduction. The two research questions both focus on consumer search for information and, consequently, consumer search for information is the main theme in this chapter. The first part of the chapter treats search for information in its context: the buying decision process. After the context has thus been set, chapter 2.2 treats consumer search for information in further detail to establish a better knowledge of determinants of level of search as a part of the buying decision process.

As the topic of the thesis is not consumer search for information as such, but consumer search for information on the Internet, chapter 2.3 subsequently looks into consumer use of the Internet, seen as one single source of information. The chapter treats both why the consumer can be expected to use the Internet for information search and how the consumer experiences the Internet, while searching for information.

The Internet can be seen as one single source which contains information. However, the Internet, in fact, consists of many different types of sources of information. Chapter 2.4 is, therefore, devoted to a discussion of the different types of sources on the Internet compared to their off-line counterparts in order to acquire a better understanding of differences between off-line sources and sources on the Internet. Of special interest are, due to research question number two, virtual communities on the Internet.

Both research questions deal with the consumer using a new type of source of information: The Internet as such and virtual communities in particular. To be able to hypothesize on why the consumer would choose the Internet for information search and whether the information which can be gathered in virtual communities can be expected to be influential, it is necessary to investigate determinants of choice of source and facilitators to a change in the pattern of information sources used. This is the topic of chapter 2.5.

The final section in this chapter, 2.6, contains the conclusion to the theoretical discussion: Proposition of a model explaining the extent to which a consumer searches for information on the Internet as a part of the buying process, and hypotheses on the use of one of the new sources of consumer information: The virtual community.

2.1 Search for Information as a Part of the Buying Process

In the theoretical line of research based on the view that consumer behaviour can mainly be understood through investigation into the cognitive problem solving process leading to a purchase, consumer search for information is not a new area. In the process leading to a decision to buy, the stages are commonly assumed to be as follows below. The proposed process is presented in Assael (1984) and is based on Dewey (1910), as cited in Brim (1962). The stages are:

1. *Problem recognition* – a need arises which the consumer thinks can be satisfied by buying a product. The consumer is now motivated to enter the next stage:

2. *Search for information* regarding the different alternatives (regarding products, brands, quantity, channel, timing, and method of payment). The search can go on inside the consumer (memory) and among external sources. When the consumer feels that he or she has sufficient information, he or she will proceed to the next stage:

3. *Evaluation of alternatives.* The different brands, or technological solutions, are compared using a decision rule based on the things that the consumer finds are important regarding the product's ability to satisfy the need in question. This stage will lead to the next stage,

4. *Choice* which in the absence of obstacles results in a purchase of the chosen brand.

5. *Post purchase evaluation.* Did the chosen brand satisfy the need as expected? If it did not: What was the reason?

Focus in this thesis is on stage 2 in the model: Search for information. It is clear from the model that search is the result of a need for information which it is necessary to possess in order to be able to make a good – or satisfactory choice through evaluation of the alternatives which are available to the consumer.

The model is very simple and thus easy to understand. The model has, however, several weaknesses limiting its use as a tool in research and practical applications. In terms of search for information the largest problems are probably the following:

First of all, there are different kinds of decisions. A common way to deal with this problem in consumer behaviour is to work with different levels of decision-making (Howard and Sheth, 1969: 46-47): Extensive problem solving, limited problem solving, and routinized response behaviour. Depending on the level of problem solving, the consumer is assumed to run through more or fewer stages of the model in more or less detail. Especially, the level of search is expected to be dependent on the level of problem solving activity, needed to reach a decision.

Second, there is no possibility of loops during the process. Consider, for example, stages 2 and 3: If the consumer finds out that he does not posses enough information to evaluate the alternatives during the evaluation in stage 3, he will have to return to information gathering (stage 2) in order to get this information. As a result, the consumer often will make several loops as he goes through the process.

A *third* problem is that the model assumes that all consumers go through the process in the same way, and that each consumer goes through this process the same way, regardless of the circumstances. Individuals differ far too much from each other, and the behaviour of each person var-

ies too much, depending on the context, for it to be possible to describe their behaviour in a single, general model without considering personality factors and context (Kollat et al., 1972).

Finally, the stages as such are not explained in sufficient detail for use of the model as a tool in researching consumer behaviour. Because of its simplicity, the model cannot be tested. In relation to the current issue, the term information gathering is far too broad to use in measurements.

As a reaction to the shortcomings of the simple model, research into consumer behaviour in the late sixties and the early seventies tried to develop theories explaining all buyer behaviour in comprehensive models capturing every possible influence and outcome (Kollat et al., 1972; Sheth, 1972). The results from this research still form the reigning paradigm in commonly used textbooks on consumer behaviour (*e.g.* Engel et al., 2001; Schiffman and Kanuk, 2000).

The comprehensive models of consumer behaviour (Kollat et al., 1972; Sheth, 1972) understand still see buying as mainly a problem solving process, in which search for information is a tool employed by the consumer with the purpose of reaching the best possible solution to his problem. However, compared to the simple model, there are several theoretical developments incorporated in the comprehensive models: First, an important addition in the comprehensive models is that the mind of the consumer plays an important role in the buying process. The consequence of this is that the comprehensive models contain two, related systems working in the consumer's mind. These two systems are: The buying decision process and information processing. Information *search* is still a part of the buying process, whereas information *processing* covers the consumer's perception and understanding of the information found during the information search. Another difference in the comprehensive models compared to the simple model is that the comprehensive models allow feedback, both during and after the process in the form of general or product/brand-specific experience with the consequence that internal search in the consumer's memory in the future may be sufficient to reach a satisfactory decision. Finally, the comprehensive models incorporate influences from both within and around the

consumer in the form of, for instance, different perceptions of risk, influences from friends and family, and influences from the macroenvironment.

The comprehensive models were an improvement, seen as a framework for understanding the consumer decision process, but their use as a research tool in consumer (search) behaviour is, nonetheless, limited. The reason for this is that the comprehensive models are untestable (Lunn in: Sheth, 1974). The main reasons why they are untestable are that there are a large number of elements and that these elements are not clearly operationalized as the researchers, in their model development, were not specific on the content of the elements in the model. Along the same line of criticism are authors arguing that the researchers behind the models are not clearly stating the focus and assumptions of the research, that there is overlap among the variables, and that the statements of causal relationships among the variables in the model are unclear (Foxall, 1980; Rau and Samiee, 1981).

Criticism from another angle hold the opinion that the Engel-Kollat-Blackwell model was (and is), like the Howard-Sheth model, regarding the consumer as too rational (Foxall, 1986). Some believe that the consumer is not rational at all (Foxall, 1986) while others just think that the consumer's rationality is bounded (Bettman, 1979). Bettman's theory on information processing assumes that the consumer is limited in processing capacity and, therefore, not capable of making the objectively best choice in a given situation. This again means that the consumer's search for information depends on his information processing capability, former knowledge, and experience. Often, the consumer will not search for information even though the risk associated with buying the product is high. This may be the case, if the consumer has experience with the product class, and therefore, relies on a known brand. Bettman refers to the risk the consumer reduces though acquisition of information as 'Handled Risk', whereas the total amount of perceived risk is referred to as 'Inherent Risk' (Bettman, 1973). That the consumer is less rational in a traditional sense than originally assumed is also the focus of Bettman's recent research which assumes that the consumer does not have a clear set of preferences at the beginning of the buying process. The con-

sumer forms his preferences as a part of the buying process and, therefore, preferences will be highly dependent on the context (Bettman et. al., 1998). Bettman et al. (1993) argue that because the consumer is limited in processing capacity, he will be able to make better decisions if he receives aids during this decision (and information search) process. These aids can help him form the 'right' preferences and help him make the best choices according to these preferences. This again means that search for information does not really play a different role in this line of research than in the comprehensive models. Search for information is still seen as an aid in choosing between alternatives and takes place if the consumer does not feel that he or she possesses a sufficient amount of knowledge to make a decision. The main difference is that the assumption of bounded rationality implies that the consumer searches until he or she is able to make a satisfactory choice, as he or she sees the situation, and not until he or she has an objectively sufficient amount of information.

The research described in this chapter has clarified that search for information plays an important role in the decision process. It is also clear that the level of search and sources used depends on the consumer and context in question.

2.2 Determinants of Level of Consumer Information Search

According to the cognitive school and as discussed in the previous chapter, a general precondition for consumer search for information to occur is that the consumer has a need that he wishes to fulfil. In order to do this in the best way, he seeks for information using various sources. Before looking at the choice between available sources it is important to understand more about determinants of level of search as search actually taking place is a prerequisite for the consumer to search on the Internet for information. Therefore, looking into possible determinants of level of consumer search is the topic of the following.

2.2.1 Search in Relation to an Upcoming Purchase

The comprehensive models explaining consumer behaviour and also general theory on information processing hypothesize that the consumer

will engage in external search until he is able to make the best (or rather a satisfactory) choice. In order to understand more about this, several studies have been carried out to try and understand the process from problem recognition to external search in further detail. In that connection, there have also been attempts to operationalize the elements in the models explaining the consumer's motivation.

An example of research into determinants of consumer search for information is a model explaining level of external search for information by, among other things, perceived risk[1] (Srinivasan and Ratchford, 1991). The model is displayed in figure 2.1. An empirical test of the relations in the model confirmed that the effort invested in search is higher if the perceived benefit of search is high, if the amount of experience in buying the product is low, and/or the evoked set is large. A high perceived benefit is, in turn, the result of high perceived risk, much product knowledge, a large evoked set, and, finally, interest in the product. Regarding the latter, Srinivansan et al. (1991) assume that a high level of interest in the product renders it satisfying in itself to search for information on the product class. A high level of perceived risk is the result of lack of positive experience and lack of product knowledge whereas the size of evoked set depends on product knowledge, and lack of positive experience. The study was not able to confirm a hypothesis about a negative link between search cost and amount of search, although the possibility was in the model[2].

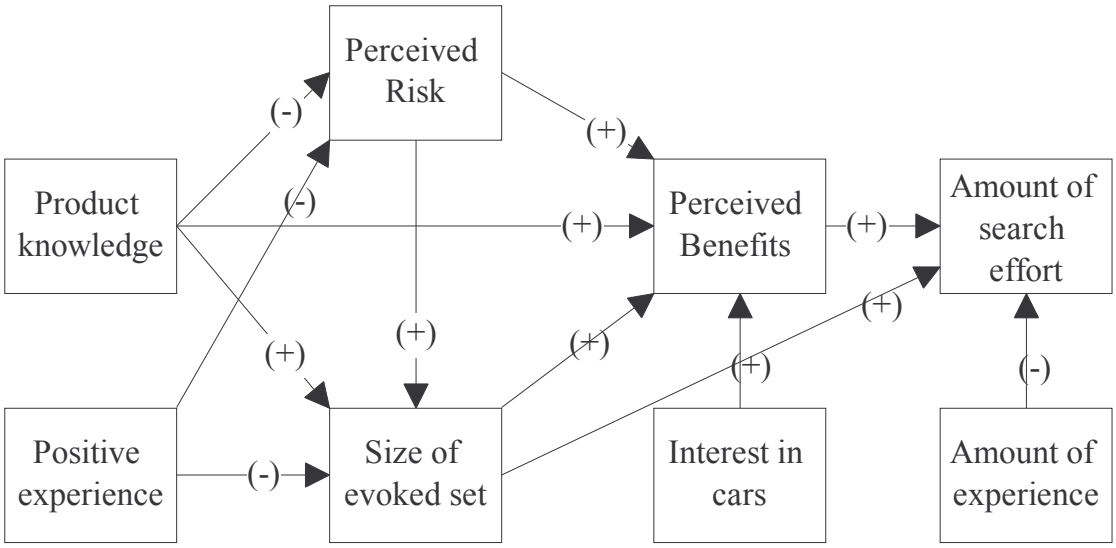
The model is displayed below in figure 2.1:

---

[1] The decision in question was choice of automobile

[2] The hypotheses were based on an extensive study of theoretical and empirical research in the field. The data were gathered as a mail survey among more than 3000 respondents and results were found using structural equation modelling.

Figure 2.1: A model of external search for automobiles



(Srinivasan and Ratchford, 1991)

The model shows that the consumer will search for information

- When he feels a need for it - because he has little relevant experience,
- Has many brands to choose from - a large evoked set, and
- Feels that he will gain from searching - a large expected benefit from the search.

He thinks he will benefit from searching because he either

- Knows what to search for - because he has a large evoked set combined with product knowledge or,
- He wants to reduce his perceived risk, which is due to lack of product knowledge and lack of positive experience,
- Spends his time learning about something, in which he is interested.

A positive experience reduces perceived risk and the size of the evoked set because the consumer then knows how to reach a positive experience again: By making a repeat purchase.

According to the Srinivasan and Ratchford model, perceived risk thus is an important determinant of external search to take place. Consequently it is relevant to look further into the contents of the term 'perceived risk'. Perceived risk can be defined as 'uncertainty about the consequences of buying a product' (Dholakia, 2001; Dowling and Staelin, 1994; Mitchell, 1999). On a general level, many researchers, in their research, apply the two-component proposed by Cunningham (Bettman, 1973; Mitchell, 1999; Srinivasan *et al.*, 1991). According to the two-component model, risk depends on dangerousness of (or less dramatic: probability of negative) consequences of making a purchase multiplied by the importance assigned to these consequences not becoming a reality.

However, the term, 'negative consequences' is rather broad. Therefore, models incorporating the term can differ even though they on the surface seem to be the same, if they use different ways of measuring 'negative consequences' and, thereby, meanings of 'negative consequences'. In the empirical test of the model displayed in figure 2.1, Srinivasan and Ratchford used the measurement scale for negative consequences developed by Peter and Tarpey in 1975 (Srinivasan *et al.*, 1991). This scale implies that the concept 'perceived risk' can be covered by perceived probability of four different kinds of loss which can be more or less important to the consumer: Financial, Performance, Physical, and Convenience loss. These types of risk all are what Dholakia (2001) classifies as 'cognitively evaluated types of risk' and risk is, therefore, not surprisingly, connected to product knowledge and not to interest in the product class in the Srinivasan and Ratchford model.

Several researchers believe that there are more types of risk than the ones covered by the Peter and Tarpey scale. A well-known and often referred-to (Mitchell, 1999; Schiffman *et al.*, 2000, Solomon *et al.*, 1999) example of such a typology is the scale developed by Jacoby and Kaplan (1972). Jacoby and Kaplan's scale includes five different types of risk. These are the three first types of risk in the Peter and Tarpey scale (risk of negative financial, performance, and physical loss) supplemented by two types of risk which are connected to the product's relation to the consumer's self-image, and which are more subjectively

evaluated: Negative psychological and sociological consequences. In addition to these five types of risk, a risk of wasting time if the product has to be repaired or replaced is later added by other researchers (Roselius, 1971; Schiffman *et al.*, 2000; Solomon *et al.*, 1999). In some respects, the 'time' risk is similar to the 'convenience' risk, and therefore, the main difference between this adjusted Jacoby and Kaplan scale and the Peter and Tarpey scale is still the addition of possible psychological and sociological 'losses' in the former.

Returning to the Srinivasan and Ratchford model, and the use of perceived risk herein, Chaudhuri (2000) argues that a shortcoming of the model is that it does not cover that there are both rational and emotional consequences of buying a product as also put forward by researchers as Holbrook and Hirschman (1982). Chaudhuri argues that different kinds of risk have to be associated with the two types of consequences. Chaudhuri thus finds that perceived risk should be seen as a two-dimensional construct with one dimension related to each kind of consequence of buying rather the one-dimensional scale, used by Ratchford and Srinivasan, or a 6-dimensional scale, as would be the case if the different types of risk were seen as uncorrelated. To test the two-dimensional model of risk based on the Jacoby and Kaplan scale, Chaudhuri carried out a study testing first a model measuring perceived risk as a combination of rational and emotional consequences, secondly a model for measurement of product involvement, and finally, relations between product involvement, risk, and search for information.

Chaudhuri (2000) found that the five types of risk, as defined by Jacoby and Kaplan (1972), could be reduced to two risk-factors. The first, Chaudhuri labelled 'Functional Risk' (containing financial, physical, and performance risk), and the second, he labelled 'Emotional Risk' (containing sociological and psychological risk). Approximately the same combination of the different types of risk, including the 'time' risk (Roselius, 1971) in 'Functional Risk', is later used in Dholakia (2001), also stressing the difference between cognitive and emotional outcomes of the product purchase. It is clear from the text that Chaudhuri refers to perceived risk and not objective risk, and, to avoid misunderstandings, I

shall in the following use the terms ‘Perceived Functional Risk’ and ‘Perceived Emotional Risk’ when referring to these constructs.

As mentioned above, Chaudhuri also studied product involvement. He found product involvement to be composed of three non-related factors: ‘Importance’, ‘Hedonic’, and ‘Risk’. ‘Importance’ was connected to degree that the product was perceived as being relevant, important, and meaning a lot to the consumer. ‘Hedonic’ were found to be related to whether the product is perceived as fun, interesting, exiting, appealing, and related to the consumer’s identity. An interpretation of the joint meaning of these words has led to that ‘Hedonic’, in the following, is labelled ‘Hedonic Value’. Finally, ‘Risk’ is related to difficulty in making the decision.

Finally, Chaudhuri tested the relation between the two types of perceived risk, type of product involvement, and level of external search. He found that ‘Importance’ augments perceived functional risk which, in turn, augments external search. This, along with the items attached to the variable as described above, indicates that this kind of importance of the product is related to importance of the product functioning correctly, and therefore, ‘Importance’ is renamed to the more explicit term ‘Functional Importance’ in the following. At the same time, Chadhuri’s study showed that ‘Hedonic Value’ augments perceived emotional risk and the amount of external search that the respondent would perform before buying the product. Chaudhuri’s argument for hypothesizing this relation is that high emotional value of the product has as a consequence that it matters to the consumer to make a bad choice resulting in him not experiencing the entertainment or excitement he is looking forward to acquire by purchasing the product.

Chaudhuri’s results indicate that the consumer is motivated for search for information prior to a purchase by two factors which differ in terms of the consumer’s relation to the product: One factor is the functional need for a solution to a problem and the risks incurred in choosing a product that does not function as expected. The second factor is connected to the product class’ relation to the consumer’s self-image and interest in the product class. A high value on the latter factor leads to a

high perceived risk of the product not fulfilling emotionally related expectations. The two factors are thus different aspects of the overall evaluation of the alternatives available to the consumer. Chaudhuri's model stresses that the two elements render expected benefit from search for information in different ways. A high level of functional importance and need for the product induces external search because of an expected benefit in terms of buying a product that functions as expected. If, in addition to this or, instead of this, the product has a high hedonic value to the consumer, he will expect to benefit from searching from the, often more subjective, knowledge related to this factor, because he can reduce his emotional risk by making sure that the product gives enjoyment either in itself or through appraisal by others. At the same, time, it may very well be that the consumer also expects a benefit from the search in itself because he finds the subject interesting.

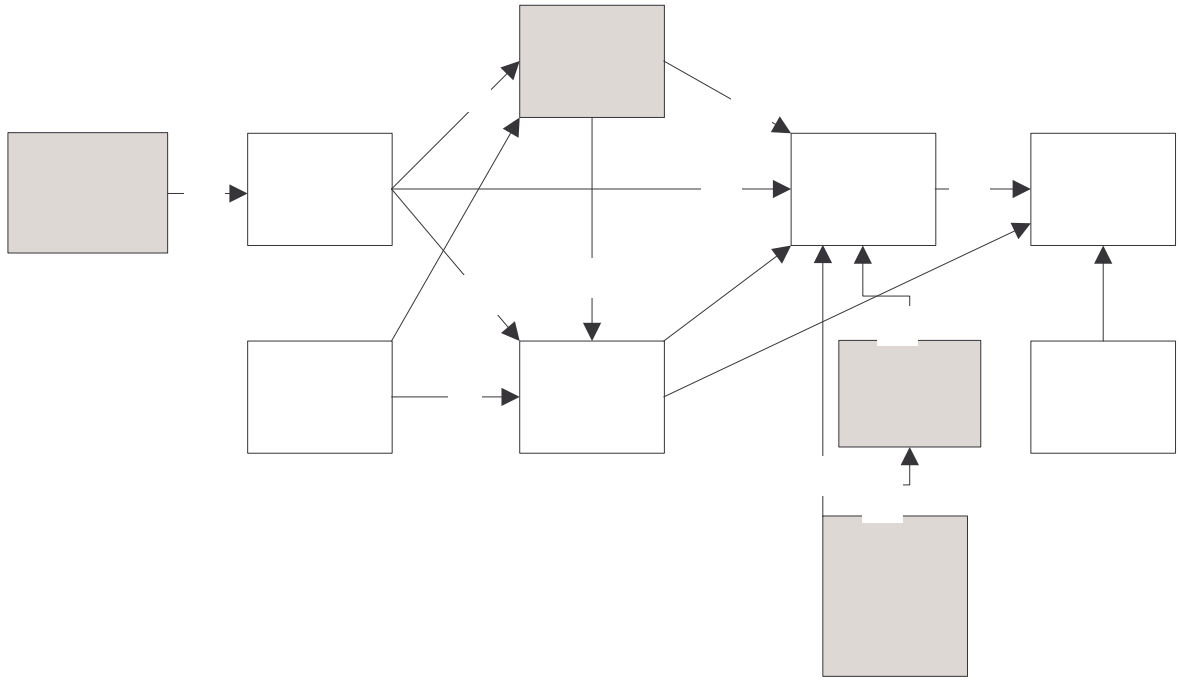
Even though Chaudhuri's results are, in a way, critical to Srinivasan and Ratchford's model, they are more supplementary than contradictory and can, therefore, be incorporated in the model in figure 2.1. Returning to the above discussion, it seems safe to assume that what Chaudhuri termed 'the importance of the product' is the cognitive need for the product because it is linked to functional risk. Therefore, 'perceived risk' in figure 2.1 could be replaced by 'perceived functional risk' and 'functional importance of product' is added as an antecedent of (functional) product knowledge.

Likewise, the product's hedonic value to the consumer is analogue to 'Interest in product' in figure 2.1, as 'Interest in product' was one of the items in Chaudhuri's scale which was connected to 'Hedonic Value'. As 'Hedonic Value' is broader, 'interest' is replaced by 'Hedonic Value'. Based on the above presented results, there should also be a path from 'Hedonic Value' to 'Perceived Emotional Risk' because a large hedonic value of the product to the consumer will heighten perceived emotional risk. Furthermore, a path directly from 'Hedonic Value' to 'Perceived Benefit' is added to the model, as it is assumed that hedonic value of the product makes the search as such beneficial – a sort of entertainment.

As functional importance and hedonic value are two sides of the consumer's need for the product, it can be discussed whether they should be related directly in the model. For the sake of simplicity, I have decided not to do so, and keep the elements apart, as did also Srinivasan and Ratchford (1991) and Chadhuri (2000).

The revised model is shown in figure 2.2:

Figure 2.2: A model of external search – revision 1



Alterations highlighted in grey.

As visualized in figure 2.2, the most important condition for external search to take place is that the consumer expects to gain a benefit from performing the search. The main difference between the model in figure 2.1 and the revised model in figure 2.2 is that the benefit in figure 2.2 is assumed to be the result of, not only a gain in knowledge about functional aspects of the product, but also an explicit benefit from gaining knowledge related to emotional risk, which is a result of the product's hedonic value to the consumer.

2.2.2 Explanations for Lack of Search before Purchase

Although, as shown in the previous chapter, research has been supportive to the notion that the consumer engages in information search before making a purchase decision other research has shown that the consumer does not engage in as much search as indicated by the comprehensive models of consumer behaviour (Formisano, Olshavsky and Tapp, 1982; Olshavsky and Granbois, 1979; Wilkie and Dickson, 1985). Several attempts have been made to try and explain this lack of search without having to reject the comprehensive models of consumer behaviour as such. This is the topic of the current chapter.

Relying on Others to Make the Decision

One angle, from which lack of search has been explained, is the one presented by Chhabra and Olshavsky (1986). Their theory is based on the premise that some consumers do not engage in external search either because they do not want to (maybe are too lazy), or are not able to make the decision themselves based on their own cognitive evaluation. Chhabra and Olshavsky tested 3 different choice strategies and 2 hybrid strategies (suggested by Formisano, Olshavsky and Tapp, 1982; Olshavsky, 1985). The different strategies were: 'Decision Making', 'Surrogate Based' (for example: Price indicates quality), and 'Other Based' (that the consumer subcontracts the decision to somebody else via following recommendations or imitation, that is, he chooses not to handle his perceived risk himself). 'Other Based' decision strategy is different from word-of-mouth and opinion-leadership in that the consumer does not make a decision himself whereas word-of-mouth and opinion-leader information acts as input to the consumer's (own) decision process. In an empirical test of the validity of the concepts, consumers actually used some of these alternatives to 'Decision Making' either instead of or along with 'Decision Making'. The 'Other Based' decision strategy might be connected to the need for conformity. This need is stronger when the buying task is difficult (Lascu and Zinkhan, 1999).

Chhabra and Olshavsky's (1986) research suggests that the relationship between functional risk and perceived benefit of search might not be linear but rather a curve with a maximum at some level of perceived risk

and declining at larger levels of perceived risk as search then is replaced by 'Other Based' decision making.

The Product is an Experience Good

A second angle, from which lack of search is explained, has been to devote the lack of search to attributes of the product. The idea in this line of thought is that some products have attributes that are searchable while others do not. Some researchers (*e.g.* Ward and Lee, 1999) emphasize that transferability of experience depends on the kind of experience the consumer is looking for. Referring to the model in figure 2.2, the consequence of this is that the relations in the model will not be the same regardless of the product class. This is because the perceived benefit in searching for information connected to products classified as search goods probably is larger because they are evaluated on objective and functional information. In contrast to this, evaluation of experience products is based primarily on emotional cues which the consumer has to experience to evaluate them. The emotional experiences are hard to transfer, but the consumer might find interest in asking others, whom he knows share the same preferences, concerning the experience at hand, as he does. Therefore, the perceived benefit form searching for information as expected in the model is probably lower for this kind of product.

Prior Knowledge Due to Continuous Search

The third angle, from which lack of search has been explained, has been to focus on limitations in the empirical research which found this lack of search. Hereby, the observed lack of search is explained by the fact that the empirical research has concentrated on search behaviour in relation to an upcoming purchase. The consumer may not search for information in this situation because he feels that he already is in possession of the information that he needs. This can be the case because he is generally interested in the product class and likes to search for information about the product class in general (in stores, but also among friends and other sources) and, therefore, has acquired information continuously through browsing (shopping without intention to buy) as explained by Bloch et al (1986). Bloch and Richins (1983) investigated this hypothesis empiri-

cally[3] and found significant correlation between product interest, general information seeking about the product, knowledge about the product, and word-of-mouth activity regarding the product, on the one hand, and browsing behaviour on the other hand. This result shows that some consumers engage in external search without intention to buy, resulting in a high level of knowledge at the time the decision process is initiated. In such cases, there will be no or only a limited amount of search for information during the actual buying process.

Revision of the Model due to the Explanations to Lack of Search

The explanations for lack of search before an upcoming purchase put forward in the preceding chapters point to needed alterations in the model of information search last shown in figure 2.2.

First, there is probably not a linear but rather a curvilinear relationship between perceived functional risk and perceived benefits from search, as was concluded in the section about relying on others to make the decision.

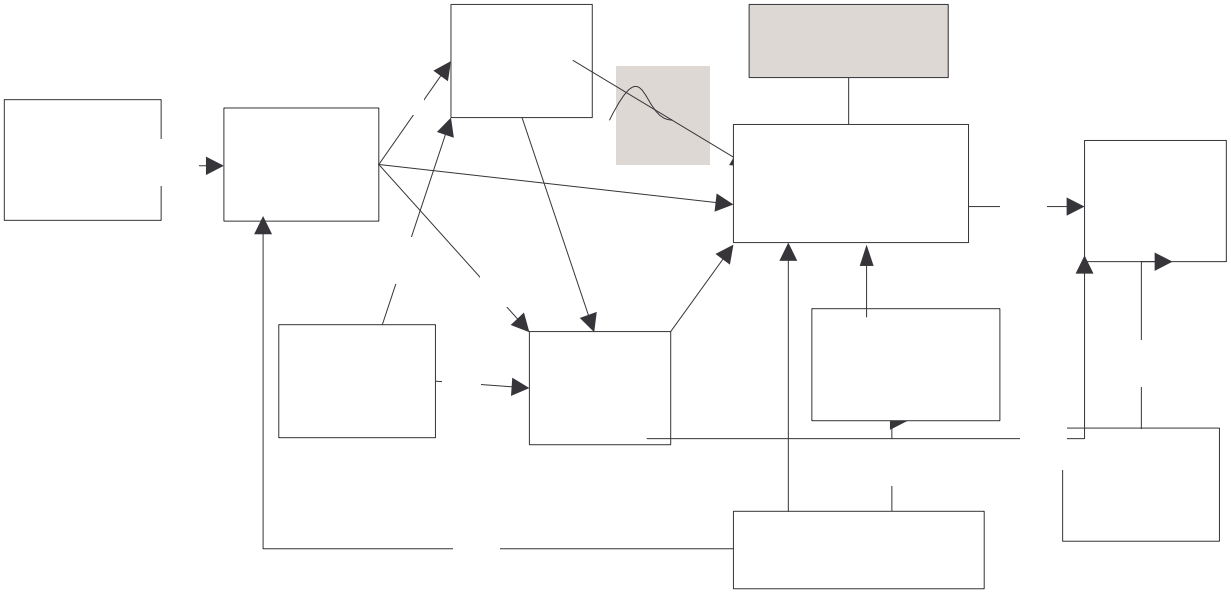
The type of product is also added to the model, as the perceived benefit from search will depend on the type of product as described in the section discussing differences between experience and search goods.

Finally, a feedback arrow from 'Hedonic Value of Product' to 'Product Knowledge', as ongoing search because of interest in the product, as discussed in the section covering continuous search, results in a higher level of product knowledge in general and therefore, a smaller need for additional information when a need for a purchase arises.

Figure 2.3 below incorporates the changes in the model.

[3] For clothing and cars

Figure 2.3: A model of external search - revision 2



Alterations highlighted in grey.

2.2.3 Conclusions on Consumer Search Behaviour

The essence of the findings cited in the above is that although the consumer does not always engage in information search before he makes a purchase, evidence confirms that the consumer engages in search under certain circumstances. These are, if he faces a decision which is perceived as risky (because the product is perceived as functionally or emotionally important), and if he feels that he is not in possession of enough knowledge to be able to choose, but on the other hand feels that he has enough knowledge to know what he is looking for and that the information he needs is searchable. The consumer also seeks information on products without an upcoming purchase in mind. This is primarily the case if he is interested in the product class. If the consumer feels he knows too little, he may retreat to others to make them make the decision based on their knowledge. If the product is not important, he may also base his decision on surrogate information as price or labels.

Expected benefit in terms of risk reduction has proven to be an important determinant of the level of search for information performed by the consumer. The consumer can perform this search using a number of dif-

ferent types of information sources, and among these, the Internet. The benefit from the various types of sources differs, as do the cost of using the source. Therefore, the next topic is to investigate why the consumer would use the Internet to find the information he needs to reduce his functional and emotional risk as described in this chapter before deciding which alternative to purchase.

2.3 The Internet as One Single Source of Information

The Internet contains vast amounts of information. Some of this information is targeted at consumers. This chapter tries to heighten the understanding of why researchers, on the basis of economic theory, expect consumers to use the Internet in their search for information. The first section looks into the general rationale for the consumer to use the Internet for information search, whereas the second section presents theory on explanations for differences in consumer use of the Internet.

2.3.1 A Place for Exchange of Consumer Information

To the consumer, there is a cost involved in getting the information needed in order to evaluate alternatives, as assumed in theories describing the buying decision process. The expectation for the consumer to use the Internet for information search is often connected to an expected lower cost of using the Internet for information search than what is incurred from using the traditional, off-line sources. However, some of the contributions also focus on the Internet giving the consumer a larger benefit in his search for information. In the following, some of these theoretical contributions on consumer use of the Internet are presented.

Butler and Peppard (1998) used the simple 5-stage buying decision model as a framework helping them to understand consumer search on the Internet imagining what consumers need from the marketer during their problem solving process. On this basis, Butler and Peppard propose how these needs should be fulfilled in marketspace, as they call marketplace in cyberspace. Butler and Peppard emphasize the need for the marketer to interact with the consumer and use the information he can collect about the individuals who visit his website(s). If marketers act according to these recommendations, information on the Internet is

seen as an advantage to both consumers and marketers because the Internet provides a place where consumers and marketers are able to meet at a lower cost.

If the consumer has limited cognitive capacity and needs help to make good decisions as Bettman (1998) states, the Internet should make the consumer able to make better decisions at the same cost, because the consumer is able to gain more information at the same cost than he is by searching among traditional sources. Bakos (1997) investigates this hypothesis using the tools from traditional microeconomics. He finds that lower search costs on electronic markets will make the market more effective and, therefore, create social surplus. This is the result of two things: One is that the individual consumer is able to get more information at the same cost and, therefore is able to make a better choice. The second is that there will be more room for companies offering products directed to a limited amount of buyers because a sufficient amount of these buyers, if they search the Internet, will be able to find these producers. The result is that there will be room for a broader range of offerings in the market. Furthermore, as the consumer is able to compare offerings and prices from a broader range of companies at the same cost, prices will decrease. However, this will happen only for products that are not differentiated. Products that are differentiated will be able to attract consumers who get higher utility from the product and, therefore, the company with the differentiated offering still has a small monopoly but with a different, more satisfied set of customers. As can be seen, this way of looking at the Internet also supports the thought that the Internet will be used by the consumer even though the argument is larger benefit in terms of availability of information and products and not smaller cost.

From the above, it is clear that many researchers see the Internet as enhancing the possibility of information exchange between consumer, producer, and retailer. However, others have shown that the markets on the Internet are far from transparent and, therefore, also far from the ideal market imagined by these researchers who base their conclusions on traditional economic theory. Bakos' hypotheses are contradicted in part by two comprehensive empirical studies carried out by Brynjolfsson and Smith (1999, 2000). These studies show that even for ho-

mogenous products like Music CD's and books, there are a large number of consumers who are willing to pay more at branded retailers (5% higher price at Amazon.com on average, compared to non-branded retailers in the study). If the consumer is loyal to a certain retailer on the Internet, he pays an even larger surplus (on average, 6.8% higher price at Amozon.com than at non-branded retailers). This is also the case for consumers who visit a shopbot[4] (Brynjolfsson and Smith 2000). The price difference is larger for consumers who visit the web site directly, probably because these consumers do not see any price comparison prior to visiting the site (and maybe not at all). The difference can, of course, stem from the fact that the branded retailers provide higher utility by delivering a better product in terms of service, delivery time, and so on, but a more plausible explanation is that consumers reduce their perceived risk concerning payment and delivery by using a well-known retailer[5]. However, this strength of (retailer) brands may deteriorate as consumers get more experienced on the Internet and thereby get better at finding the right information about the various retailers and brands (Ward and Lee, 1999).

The above concentrated mainly on the benefit from the Internet's ability to act as a place of exchanging information between consumer and marketer at a low cost. There are also additional reasons why the Internet is seen as an advantage to both consumers and producers in their exchange of products and information. One reason put forward is that the consumer can get in direct contact with the producer and thereby reduce the need for middlemen. This saves both money and time and may provide better information. Some consumers actually feel that they get better advice in a better atmosphere on a website than by a salesperson in a retail store (Wolfinbarger and Gilly, 2000).

Alba *et al.* (1997) use the benefit angle in their discussion on a) what conditions would make the consumer search the Internet for information

---

[4]Shopbots are 'Internet-based services that provide one-click access to price and product information from numerous competing retailers" (Brynjolfsson and Smith, 2000: 5).

[5] They use what Duhan, Johnson, Wilcox and Harrell (1997) called 'surrogate based decision making".

about and make purchases of products, and b) what conditions would make retailers place their offers on the Internet. The interesting part of their discussion in relation to this thesis is the part on information acquisition and entertainment during the process. Alba et al. argued that the Internet compared to traditional retailers and catalogue dealers offers more information. But the information was considered to be of a lower quality and there were problems in comparing alternatives, especially from different dealers. Alba et al. (1997) consider entertainment and social interaction to be low. This could act counter to using the Internet for information search. On the other hand, personal security is high, as one remains at home while shopping. This will enhance the use of the Internet. Fear of personal security is probably a more relevant issue in some countries (USA) than in other countries (Denmark). In order to overcome the prohibitions against using the Internet for information search and purchasing Alba et al. envisioned a system with true interaction between consumer and retailer. They argued that in order to make the Internet a feasible alternative to traditional retailing it has to offer effective screening and selection devices and credible information. These attributes would make the Internet more effective in delivery of information and also more entertaining and, therefore, more attractive. This is in accordance with results from other researchers in the field (Burke, 1997; Hoffman and Novak, 1997; Korkaonkar and Wolin; 2000; Wolfinbarger and Gilly, 2000).

Another benefit in using the Internet as information provider is the selection available to the consumer as described when discussing Bakos (1997). The consumer can visit many different producers directly at their various websites, or he can visit a cyberstore (i.e. a virtual store acting as a traditional retailer). The cyberstore has fewer expenses in stocking and premises (rent, furniture and so on) than the traditional store, and it can more easily attract customers from a large geographical area, and is therefore more effective than a traditional store (Hoffman and Novak, 1996, Alba et al. 1997).

Also focusing on the larger benefit from search on the Internet rather than the smaller cost are Hoffman and Novak. Hoffman and Novak consider themselves as belonging to Postmodernism and thus treat the ad-

vantages of the Internet from another perspective than does researchers who belong to the cognitive school. As postmodernists, Hoffman and Novak, in their research, focus on the hyperral world on the Internet and how the consumer can experience that he is 'there'– a feeling termed 'telepresence'. As described in the introduction, this is not contradictory to the cognitive school of research, especially not when dealing with life on the Internet.

In many ways Postmodernism fits the Internet. Life on the Internet is hyperreal; it is also fragmented as the consumer can easily visit many different stores. Thereby, the consumer has far more possibilities of mixing styles in products and shopping than in the physical world where visiting stores is limited not only by distance between the stores, but also by the way the consumer is perceived by the salesperson. Furthermore, the view of the consumer as playing an active part in forming society fits very well with the much more active way the consumer can choose and use his information sources Internet when searching for information than what is the understanding of the consumer in traditional mass-marketing. The Internet also gives the consumer good possibilities of juxtaposing opposites, because the consumer is more free in selecting products from several suppliers and put them together as he pleases. Finally, on the Internet, the consumer does not know the retailers, he visits, in person. Therefore, commitment to retailers on the Internet is also expected to be less than commitment to traditional retailers.

Hoffman and Novak argue that the traditional one-to-many model of communication between company and consumer does not fit reality on the Internet. Instead, a many-to-many model should replace it (Hoffman and Novak 1996). The many-to-many model incorporates the fact that both consumers and firms can put communication on the Internet as well as use information from the Internet. Some consumers will do both and others will do only one of the two (but the provider will be able to see that they took the information). This property means that word-of-mouth has good opportunities on the Internet as does communication directly between company and consumer (one-to-one). In this early work, Hoffman and Novak envisaged a consumer gateway where the

consumer can navigate by point-and-click. These gateways are increasingly common today.

In their early work, Hoffman and Novak (1996, 1997) also emphasized that the Internet is an alternative to, and not a simulation of, the real world because the consumer interacts with the computer and thereby is able to create his or her 'own' store. Furthermore, Hoffman and Novak (1996, 1997, 2000) state that the environment on the Internet is suited for both experimental (surfing, shopping without a purchase in mind), and more goal oriented behaviour (information search in relation to an upcoming purchase). Experimental shopping on the Internet is associated with surprise, uniqueness, and excitement, and with positive sociality (chatting with other shoppers sharing the same interest), online deal searching (auctions, bargain hunters), and involvement with a product class (hobby-type) (Wolfinbarger and Gilly, 2000).

Hoffman and Novak acknowledge the real world, but they hypothesize that the consumer, with the right skills and the right computer environment, will be able to experience what they call 'flow' while using the Internet. Flow is defined as 'an online experience where the consumer is completely engaged with his or her interaction online to the extent that they loose contact with their immediate physical surroundings' (Hoffman and Novak, 1999).

Hoffman and Novak thus emphasize the benefits to the consumer from the hyperreal world which the he actively can create and become a part of when searching for information on the Internet. They envision consumers playing a much more active role on the Internet than in the traditional marketplace. In continuation of this, they argue that the Internet will be more successful as a market and a medium if the consumer is freed from his passive role as a receiver (Hoffman and Novak 1996). This is in accordance with the thoughts of a researcher, belonging to the cognitive school, Peterson (1997), who emphasizes the Internet's potential benefits for consumers: convenience, current and more complete information, along with potentially lower prices. Other researchers, belonging to Postmodernism, as Venkatesh (1998) emphasize that the consumer, on the Internet, is freed from behavioural norms, as he is 'alone'

in the store (he cannot be seen by the other individuals in the store), and the personnel cannot see him, either.

The above shows that both postmodernist researchers as Hoffman and Novak and Venkatesh, along with traditional researchers as Bakos and Peterson see the Internet as an opportunity for the consumer to become more active in shaping his search for information. However, they see different motives for the use of the Internet. The first group of researchers primarily see the Internet as a means of entertainment, communication and shaping identity, while the second group of researchers primarily see information search using the Internet as a means of buying the 'best' product cheapest in terms of cost of the product and the search for information prior to the decision . These different perspectives on consumers' advantage of using the Internet are not contradictions, however, as both views can be a part of the whole truth: that the consumer sometimes seeks the one, sometimes the other depending on the product and the situation and that some consumers mostly seek information and others entertainment.

Returning to the model in figure 2.3 explaining consumer search for information as motivated by perceived benefit in terms of risk reduction and interesting information, the theory presented in this chapter emphasizes benefits of search using the Internet by availability of information and entertainment while finding it, not just because of hedonic value of the product also because of hedonic value of searching the Internet. Also, if the cost of searching is lower, the consumer can get more benefit using the same amount of resources.

The theories presented in this chapter are far too general to help us on the question which consumers are more likely to use the Internet than others. To answer this question we will have to look more into research on the level of the individual consumer and his use of the Internet.

2.3.2 Research into Consumer Behaviour on the Internet

In relation to the consumer searching for information, the Internet, compared to traditional marketing communication channels, can be characterized as follows (inspired by Bauer et al., 2001):

- *No physical contact*. The contact exists in cyberspace where both parties are anonymous. This has both physical and psychological consequences. The physical consequence of not meeting in person is that the consumer does not have to travel to the person or company he wants to communicate with. This makes it less costly in terms of both time, money, and for some also effort to use the Internet than the traditional information sources. The psychological consequence of not meeting in person is due to the fact that social interaction is different – and for some less satisfying than the social interaction experienced through a traditional meeting. This means that the psychological distance may be experienced as larger by some.

- *Constant availability of information.* The consumer can access the information when he wants and regardless of physical location. This means an easier access to information.

- *Information can be addressed directly.* The Internet enables people to communicate directly and immediately regardless of their spatial and temporal distance.

- *Interactivity*. The information the user gets is often dependent on the input he has given to the computer either as a question via e-mail, via clicking the various options on-screen or entering a search term. He is, therefore, at least partly, able to control the information flow.

- *Efficient transfer of information*. Information can be updated frequently at a low cost and products and services can be presented visually and using sound. In some cases, a demo can be transferred via the Internet.

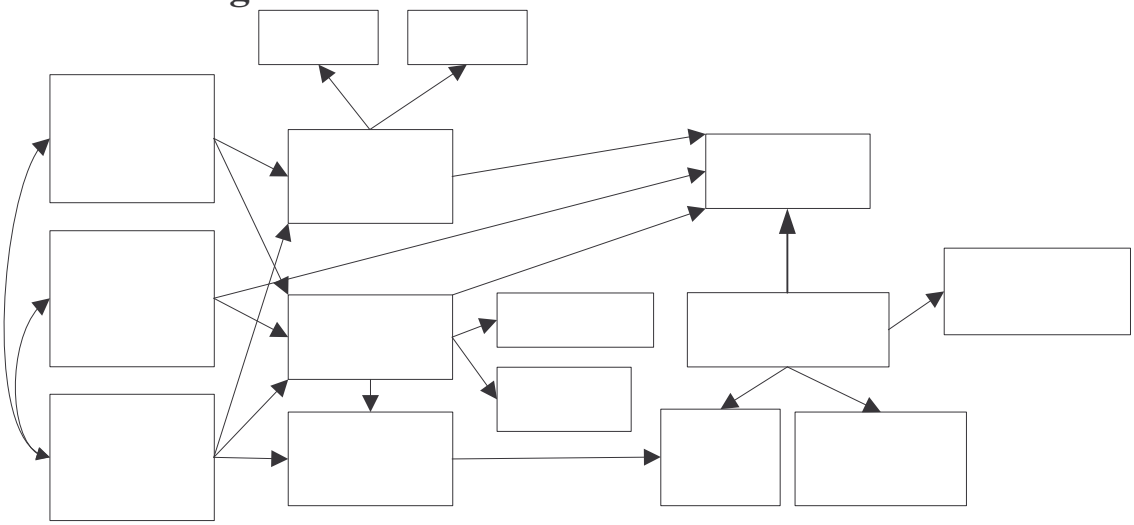
- *Individuality.* Interactivity and efficiency in information transfer allows individualized information and communication at a sustainable cost.

However, not all consumers experience these characteristics as equally advantageous. As described in the former chapter, Hoffman and Novak believe that consumers will perceive the search process on the Internet as more rewarding than the search process in the physical world, and incorporated this in their 'Model of Network Navigation' (Hoffman and Novak, 1996). The model presented in figure 2.4 on the next page, emphasizes that the consumer is on the Internet for either goal-directed (functional) or experimental (emotional) reasons, much like the different perceived benefits from search we saw in figure 2.3.

According to the model, the consumer will focus his attention on navigating on the Internet if he meets content which allows interaction and/or is fun to look at. The outcome of this process depends on the relationship between challenges (how hard it is to navigate) and skills (what the consumer knows about navigating the Internet). If there is congruence between skill and challenge, the consumer can experience 'flow', which means that he surfs the internet without paying attention to the physical world around him. A part of the good feeling is due to experiencing 'telepresence'. If the consumer experiences flow, he will learn about products, perceive control of the process, get a positive experience, and want to explore the Internet again. Flow, thereby, lowers the perceived cost and augments the perceived benefit of searching for information on the Internet.

In addition to the concepts discussed in the above, the model in figure 2.4 also explains (some of) the causes for differences in the level of perceived skill and control, challenge and arousal, and focused attention. Time on the Internet, perceived interactive speed, and importance of Internet in the life of the consumer are factors that explain some of these variables. A slightly revised (simplified) model was tested using GVU data from 1997 and the empirical data mostly supported the model (Novak and Hoffman, 2000).

Figure 2.4: Hoffman and Novak's model



(Hoffman and Novak, 2000)

The model in figure 2.4 does not incorporate any demographic or personality variables. This is a weakness, as there might be differences in the use of the Internet among individuals due to differences in their personality or socio-demographic characteristics. Also, the model was tested on data that were based on self-selection of the respondents on the Internet. Therefore, the respondents are probably not representative of the consumer in general and probably not of the user of the Internet, either.

A qualitative study on consumers' use of the Internet (Wolfinbarger and Gilly, 2000) also gave support to the hypothesis that consumers search for information on the Internet because they are more in control over the situation and freer do to as they like – shop when and how they want to. In other words – they value the availability of the information on the Internet. Furthermore, the consumers participating in the 9 focus groups in the study emphasized the availability of information on the Internet at a possible cost and effort. They also found that some consumers actually prefer the lack of social interaction on the Internet. Especially, they appreciate the lack of salespeople, because these put on pressure to make a purchase, and are perceived as limited in knowledge. This perception of sales personnel might be a consequence of the fact that consumers who use and shop on the Internet are different from the population in general; they have the personality and demographic characteristics of innovators. These include high internal locus of control, better income, and higher

Time since start using the Web

Percieved interactive speed

Importance of Web in life

education (Forrester, 1999; ProActive International; 2000; Rogers, 1995; Schiffman and Kanuk, 2000).

Recently, Hoffman, Novak and Schlosser (2000) investigated the impact of personality on the desire for control over the information search process. The empirical evidence from this study shows that consumers with high internal locus of control[6] will use the Internet more actively and goal directed. The opposite is individuals with high external locus of control[7] who, as they feel that they have no control over what happens, are much more passive in their use of the Internet. This is because they do not believe that, for instance, product search will help them make a better purchase. This lack of benefit from active search for information perceived by consumers with high external locus of control is expected to explain why these consumers use the Internet in an experimental (entertainment) way rather than in a goal-oriented way. The study also showed that individuals with a high internal locus of control had had access to the Internet for a longer period of time than had the individuals with a low internal locus of control. The consequence of this is that it would probably be dangerous to expect that if individuals have been longer on the Internet, they will start to use it in a more goal-oriented way. This is because the difference in use is due to a difference in personality and not to a difference in experience and skill.

The findings on relationship between locus of control and length of time the consumer has been using the Internet are in accordance with theory on adopter categories and their behaviour. This is because individuals with internal locus of control most likely are innovators or early adopters, whereas individuals with external locus of control more likely belong to the later adopter categories. This again means that, bringing the

---

[6] Internal Locus of Control means that the consumer thinks he himself is able to control (part of) the environment and the outcome of actions. Internals are older, more educated and with higher household income. They have used the web longer and are more satisfied with their skills.

[7] External Locus of Control means that the consumer thinks that outcomes are due to outside forces over which he has no power. That means fate, luck or powerful others. Demography and web usage is the opposite of Internal Locus of Control individuals.

‘Multi-Step Flow of Communication’ (Schiffman and Kanuk, 2000) into mind, it will be individuals with internal locus of control who actively seek the Internet for information and bring this information to individuals who act far more passively towards the Internet and to mass information sources in general. Therefore, individuals with external locus of control are mainly reached through individuals with internal locus of control or via entertainment sites. The fact that most on-line purchasing is planned and that the consumers only stay at e-commerce sites for a limited amount of time (Wolfinbarger and Gilly, 2000) also supports the hypothesis that individuals who search for product information on the Internet are goal-directed and thus mainly individuals with internal locus of control. Another reason for the short stay could be that on-line shoppers often may consider time a scarce resource because time on-line often is (was) paid by the minute. Hoffman and Novak’s thesis that entertainment is an important element in the consumer’s perception of the usability of the Internet is confirmed by a study by Eighmey and McCord (1998). This study revealed that the most important factors in individuals’ evaluation of websites were entertainment, personal involvement, and personal relevance. Less important was information involvement, clarity of purpose, controversy, and credibility, continuing relationship, and purchase interest.

Based on the theories and empirical evidence put forward in this chapter, one should expect that the consumer who has some skills in using the Internet would find information search on the Internet more entertaining because he is able to find more information and a larger number of rewarding sites. Also, he is probably also able to locate the needed information faster than the non-skilled user. Therefore, he more often gets a positive experience which means he has a higher expected utility of searching the Internet for information.

This chapter also showed that the consumer’s personality influences how he perceives the opportunities on the Internet. Consumers with a high internal locus of control are goal oriented and therefore seek information in the Internet while consumers with a high external locus of control value entertainment.

The theories presented finally show a link between personality and weighting of attributes of possibilities for information search. For some consumers the possibility of searching fast, without interruption, anonymous, and goal-oriented is important, whereas others put value to the contact they can get with other people while searching for information.

Again returning to figure 2.4, the additional information from this chapter as compared to the contributions from chapter 2.3.1 is that not all consumers will perceive the same benefit from searching for information on the Internet. Consumers, to whom the Internet is more important, who have used the Internet for a longer period of time can be expected to perceive a larger benefit from using the Internet for information search because they will find the needed information more easily and because they will find it entertaining to search for information as such. Also, the consumer's locus of control and attitude towards contact with sales personnel may be possible determinants of the level of use of the Internet for information search prior to a purchase.

This chapter moved one level down from consumer information search in general to looking at consumers' benefit from using the Internet as such for search for information. However, the Internet is not one single source of information but consists of several types of information sources. The next task is therefore to move down another level and look into the characteristics of the various types of sources of information on the Internet.

2.4 The Internet as Several Types of Sources of Information

When the consumer searches for information, many types of sources are available to him. The sources differ in several ways. Traditionally (*e.g.* in common textbooks like Kotler's Marketing Management (Kotler, 2000)), classification of sources takes place on two dimensions: The first dimension relates to who dominates the source – marketer or non-marketer based on the sources' possible motive in own commercial gain from giving the information. The second dimension relates to the question of whether the communication is directed to the individual – if it is

personal or non-personal. In figure 2.5, the most common sources are put into this framework.

Figure 2.5: Traditional sources of information for the consumer

	Non-marketer dominated	**Marketer-dominated**
Personal	Membership groups	Retail Sales Personnel
Non-personal	Consumer information from consumer groups or state institutions Product reviews and tests in printed magazines and papers	Direct mail Advertising

(Based on Blackwell et al., 2001; Kotler, 2000; Schiffman & Kanuk, 1997;Sheth et al., 1999)

The general idea behind the framework is that the source of the message is perceived as credible if the 'person' is believed to be knowing, objective and trustworthy (Solomon, 1999). If the source is credible, it is also more persuasive (Campbell and Kirmani, 2000), and is then also probably more used in the search for information. The more interest the sender can be expected to have in showing the product in a positive light, the less trustworthy, the source is perceived to be by consumers. This is because marketers live from selling products to consumers who evaluate these products positively. However, this lack of trustworthiness resulting in loss of credibility can be offset by a perceived high level of knowledge and is less severe if the information sought is objective. The second characteristic of the framework is the level, to which the message is tailored to the individual receiving the information. Tailored information is, of course, more precisely fulfilling a perceived need for information. However, precise information is not always needed – or asked for, for that matter - and sometimes non-personal sources possess a higher level of knowledge and are, therefore, perceived as equally good or maybe even better as information source, than personal sources are.

In this thesis, we are trying to shed a light on why consumers would use the various sources available to him on the Internet. According to the theories presented in chapter 2.2 and 2.3, the consumer will use these sources if they provide a higher expected benefit in terms of risk reduction and/or benefit from search as such than does the traditional sources. To get a better idea about how the sources on the Internet can be perceived when performing as information sources compared to the off-line sources, it is necessary to compare the two. To understand more about the ability of the sources on the Internet to give consumers the perceived benefit they search for, we have to know more about the different sources on the Internet. A step in this process is to clarify what the corresponding sources on the Internet to the sources displayed in figure 2.5 are. The question thus is: What is the content of figure 2.5 if the same classification is used on information sources on the Internet? The result is shown in figure 2.6 for information sources on the Internet, and the contents of the cells are discussed in detail in the paragraphs to follow.

Figure 2.6: Information sources on the Internet

	Non-marketer dominated	**Marketer-dominated**
Individualized	E-mail with traditional membership groups Newsgroups Chat Brand Communities	Questions via e-mail to producer or retailer Direct e-mailing
Non-individualized	Independent search agents Consumer website information from consumer groups or state institutions Product reviews and tests published on sites on the Internet	Consumer related commercial sites Sponsored search agents Company websites

(Developed from figure 2.5)

Figure 2.6 above illustrates that it is possible to use the same framework as the one used in figure 2.5 on the information sources on the Internet. There is, however, one difference between the two tables:
It was necessary to substitute the term 'personal' by the term 'individualized'. This is due to the fact that 'personal' implies that consumers meet physically and this is not the case on the Internet. Rather, there is a difference as to whether the communication is directed to an individual with respect to contents and maybe also style, or if this is not the case.

The framework presented here has much in common with the framework for online personal influence presented by Sénécal and Nantel (2000). Sénécal and Nantel work in three dimensions: The first dimension captures whether the consumer is the active part or not. That is, the dimension distinguishes between websites and e-mails, as the consumer visits the first and receives the latter. The second dimension captures the medium distinguished as seller, commercially linked 3rd party or non-

commercially linked 3rd party. The third and last dimension relates to whether the sender can be classified as other consumers, an expert, or a decision and support system.

Although this framework is, in a way, more elaborate than the one presented in figure 2.6, it is not considered better for the question treated in this thesis. The reasons for this are as follows. First, the first dimension distinguishes who is the active part in the buyer-seller dyad. This dimension is irrelevant in this study because it has already been chosen that the topic is active search and, therefore, the consumer is the active part. The second dimension – sender type – has, compared to the dimension in figure 2.6, an extra distinction regarding the commercially linked sources of information. This does not really add anything to figure 2.6 as these are the opportunities in the lower left cell. The third dimension is related to the type of knowledge that can be gained from the opportunity. This is linked to the consumer's search for information, but still the framework from figure 2.6 seems to fit the problem better because the view here is the consumer's perception of which sources provides the better information. For example, he might find that other consumers are the experts on a topic involving subjective judgements rather than technical experts on the subject.

The conclusion to the discussion in the above is that the framework presented in figure 2.6 is the one selected as a guideline in the following discussion on the content of the possibilities. It is important, however, to realize that this framework is a simplification, as the true scales of the two characteristics in figure 2.5 and 2.6 are, in reality, continua and not just the two options. In the chapters below the content of the four cells are discussed in detail.

The following chapters discuss strengths and weaknesses of the types of information sources available to the consumer on the Internet.

2.4.1 Individualized, Non-marketer Dominated Sources

According to figure 2.6, sources of information which are individualized and where the source is not marketer dominated includes e-mail with

‘traditional’ reference groups and various types of virtual communities (also called forums or electronic communities) on the Internet.

E-mail is the type of communication on the Internet most commonly used. Approximately two thirds of Internet users use e-mail (Nie and Erbring; 2000; Netvalue, Nov 2000). According to a report from UCLA (UCLA, 2000), 42% of e-mail users in USA use e-mail every day, and according to the same report, 25% have friends they have only met on-line. Research has shown that face-to-face meetings and telephone conversations with traditional membership groups are replaced (partly) by e-mailing with the same group (Nie and Erbring, 2000; Pew Internet, 2000; Stanford, 1999). E-mail is a rather active way of searching for information as the consumer has to pose his question in writing and, consequently, e-mail requires at least some effort of the consumer. As e-mailing, in the current context, substitutes for meeting and telephoning friends and family (Boston Consulting Group, 2000; Stanford, 2000), it is often not less costly than the traditional means of communication unless there is a large distance between the parties, but e-mailing might be considered quicker by the consumer. E-mailing is not considered to be much different from contacting the same groups using traditional methods. Therefore, this type of information source will not be studied in further detail.

Communication in virtual communities is a less individualized way of communication and is used less, than are person-to-person e-mails. Communication in virtual communities can take place via Bulletin Board System (BBS) or instant messaging. BBS is a service for posting and reading messages posted by others to the service whereas instant messaging is a service enabling the consumer to find others on the Internet and communicate directly with them. There are several types of virtual communities on the Internet not involving marketers. Information exchange can take place via lists, chat, and in newsgroups on USENET (Blanchard, 1998; Kozinets, 1999). In principle, lists, newsgroups, and chat work in about the same way: It is possible to post a question to the other participants and read questions and answers from other participants. By subscribing to a list the consumer receives all e-mail sent to the list, and can also post e-mails to the list containing questions or

comments. Membership of newsgroups also require subscription or membership and in addition a newsreader (which is a facility in most e-mail software) or access via a website. The consumer visits or logs on to the newsgroup and is then able to read and post messages from and to the group. Communication thus does not take place in real-time. Communication on lists and in newsgroups is organized as discussions about subjects and runs as comments to subjects. Approximately 5% of active Danish Internet users use newsgroups[8]. Chat can take place in a chat-room, on a website, or via instant messaging and is real-time and runs progressively, and not in threads. The consumer can both be an active participant in the chat and a passive reader of other people's chat. In Denmark, 3.7% of Internet users use chat (Netvalue, Nov. 2000). Chat is used considerably more in France (11.1%) and Spain (35.5%) than in Denmark. Chat is used more by people under 30 (Stanford, 2000). In the countries where the population uses e-mail less than the Danish population there might be some substitution between chat and e-mail. In terms of cost in comparison to meeting face-to-face, participating in virtual communities is often cheaper even if the contact is not over really large distances, depending on the price structure of using the Internet. Virtual communities are used more by experienced Internet users. These consumers are more likely to see the Internet as a place for contact with others who share the same interest (Maignan and Lukas, 1997; Wellman and Gulia, 1999). Hoffman and Novak's many-to-many model seem to be well-suited for this type of communication[9].

Newsgroups and lists both are normally centred on a specific subject, often related to a product class – *e.g.* computer hardware, consumerism, or pets (Fisher et al., 1998). In both, communication is organized in subjects. On lists, the referred-to subject appears in the subject line, whereas in newsgroups, comments are placed under the relevant header. This way, communication in newsgroups about one subject is kept together and it is, this way, easier to survey and to comment on specific subjects it is on lists. Also, communication, this way, resembles 'real' discussions more than do communication on lists. Further, participation

[8] Based on Tønnesen (2000) counting 143.182 users combined with Danmarks Statistik (2000) saying that 63% of the population are active Internet users.

[9] See chapter 2.3.1 about Hoffman and Novak's many-to-many model

in lists is far more passive than participation in newsgroups, as the e-mails are received passively. Lists and newsgroups also differ in the ability to search for and see 'old' communication in the group. When logging on to a newsgroup, the consumer downloads as many past questions and answers as he has asked his newsreader to see. He can then read these questions and answers and also post questions and comments himself via e-mail or a webserver set up to this purpose. This is not possible on a list. However, some lists provide websites containing 'old' communication from the list. Chat and instant messaging differ from lists and newsgroups in terms of consumption relevant content, as they are more often centred on a specific segment of individuals than on consumption issues. Also, the topic of chat is more often 'loose talk' in general and on personal matters. Considering the organization of communication in discussions and the easy downloading of past discussions, newsgroups seem to be more relevant for information search prior to a purchase than do lists and chat.

Contact via lists, newsgroups, and chat is often with people whom the individual do not know in real life. Often, the individual does not know the true identity of the person. This seems to be the case more often in chat than in lists and newsgroups (based on unstructured observation by the author). The fact that the information comes from people whom the individual does not know in real life means that, in general, the source probably is considered less credible than traditional reference groups because the consumer is not able to judge the trustworthiness of the source. However, if virtual communities can be regarded as 'real' communities of which the consumer perceives himself a member, the consumer may see this differently. Therefore, in relation to this project, virtual communities are mainly interesting if they can be regarded as 'real' communities, as this is an important assumption in seeing them as a supplementary reference group to family, friends etc. This is because it may then be that a perceived high level of knowledge about the product class at hand is able to outweigh the smaller credibility that stems from not knowing the individual in real life. This, of course, will mainly be the case if the consumer feels that he knows the other members of the community, because the credibility will then be larger than if he does not feel that he knows the other persons in the community. The conse-

quence of the above discussion is that it is important to consider whether virtual communities can be regarded as 'real' communities.

Communities are characterized by a shared sense of belonging, by rituals and traditions, and by moral responsibility (Muniz et al. 2001). In chapter 1.2, virtual communities were defined as 'social aggregations that emerge from the net when enough people carry on … public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace' (Rheingold, 1993, cited in Kozinets, 1999). A similar notion of communities in cyberspace[10] can be seen in a later definition by Armstrong et al. (1996) pointing to virtual communities as characterised by '...aggregations of content and communications offerings in a single virtual space....', '... [a] focus on user generated content....', and having a '...primary focus that draws users into the community and encourages them to stay'. The fact that virtual communities have a primary focus renders them interesting to the study of consumer behaviour if some of the communities focus on consumption-related issues. This may be the reason for the lesser focus on social relations, and more on content, in this, more marketing-oriented definition. However, returning to the definition of community, sense of belonging and moral responsibility are also necessary elements for the virtual community to be 'real'.

Wellman and Gulia (1999) elaborate on this question and challenge the Internet's ability to maintain a sense of belonging which, in their terminology is the same as the members feeling strong ties to the community. Wellman and Gulia point to the inability for the members to see each other in person, and to online relationships being based more on shared interests, often related to consumption (Fischer et al., 1996) than on shared social characteristics. On the other hand, Wellman and Gulia argue, it is easier to get in touch with (all) of the other members of the community than it is offline. Further, virtual communities only exist if a sufficient number of members put content into the community, as Arm-

[10] In the article referred to as 'electronic community'. However, as the article is from 1996 when the terminology concerning the Internet was more loose than today and thus it seems safe to refer to the definition in spite of this difference in terms

strong and Hagel (1996) emphasize. This entails that members are actively engaged which is not necessarily the case in traditional communities. Wellman and Gulia conclude that virtual communities can be classified as intimate, secondary groups to individuals because members are very homogenous in terms of interests but not necessarily in terms of social background. Because of the shared interests, members can easily relate to each other even though ties are weakened by the lack of ability to meet in person and the ease of going in and out of the community compared to real-life communities. This is supported by the fact that persons using the Internet have been found to have a greater tendency to base their sense of belonging on the basis of shared interests rather than shared social characteristics (Wellman and Gulia, 1999). Based on these assumptions, Wellman and Gulia, in line with other researchers (Bickart, 2002; Fischer, 1996; Granitz, 1996; Muniz, 2001), conclude that virtual communities qualify as 'real' communities concerning sense of belonging.

A second characteristic of a community was 'moral responsibility'. The question of moral responsibility toward other members of the community can be seen as related to theories on information flow between consumers. The virtual community, in contrast to many 'real' communities, exist because the participants voluntarily choose to be members (Bagozzi, 2002, Wellman and Gulia, 1999) and the community only exists through the communication which members place there. Therefore, there has to be communication in the community for the community to stay alive. Such communication will be enhanced if there are members who support reciprocity in the community (Rehm, 2000, Bagozzi, 2002) – that is, if they feel moral responsibility towards the community. Opinion seekers have been found to listen to opinion leaders provided they are perceived as knowledgeable (Aaker et al. 1992, Flynn et al., 1996; Solomon et al., 1999) and it is, therefore, also a necessary prerequisite that such (perceived) knowledgeable consumers take active part in virtual communities of consumption is these are to be relevant for information search. This implies that in the community there has to be consumers seeking information and other consumers who are willing and able to provide the needed information (feel morally responsible to do so). This pattern was found in newsgroups centred on consumption of coffee

(Ganitz and Ward, 1996; Kozinets, 1999) and in general, as cited in Wellman and Gulia (1999). If virtual communities thus consist of both opinion seekers and opinion leaders[11] and fulfil the needs of both groups, this will enhance the use of these communities as sources of consumer information because the virtual community is then close to a 'real' community.

The third and final characteristic of a community is shared rituals and traditions. Virtual communities live up to this characteristic in the form of 'Netiquette' which consists of rules on what is appropriate behaviour and what is not in virtual communities. If attendants of a community do not comply with these rules, they are told so and are shut out of the community if he or she persists in not obeying Netiquette. This also is a sign showing that members feel responsibility towards the community because they, this way, help sustaining the community as a serious forum for discussion if this is the purpose of the community.

Based on the above, it seems fair to classify at least some virtual communities as 'real' communities. Which types of communities can then be expected to be more relevant in terms of consumer search for information in relation to an upcoming purchase? According to the above, these are virtual communities centred on consumption issues, in which the members feel a sense of belonging and in which a sufficient amount of relevant communication takes place. Lists are not considered to be interesting in the current study because lists are quite closed, and because the list format does not induce sense of belonging to a group and is not suitable for searching for past discussions. In contrast to subscription to lists, participation in newsgroups requires that the consumer is active because it is necessary to log on to the group to see and post messages. The consumer thus can be active but he also has the opportunity of being a passive reader. Newsgroups are often based on a product class and are thus consumption related. Also, communication is organized in a way which makes information search among 'old' discussion in addition to posing questions to the newsgroup. In addition, several

[11] Defined as 'individuals who lead in influencing other's opinion about innovations'(Rogers, 1995)

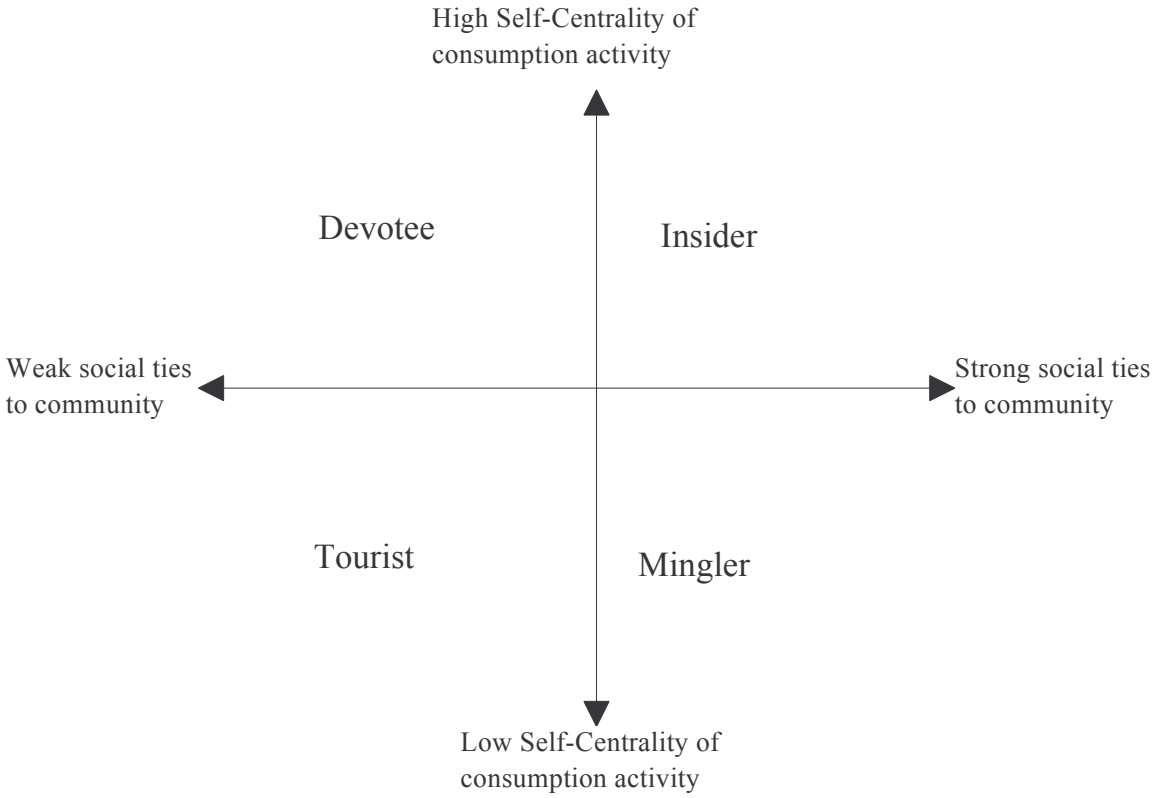
studies have found newsgroups to be capable of qualifying as virtual communities (Fischer et al., 1996, Granitz and Ward, 1996), because they are built around a common consumption of a good or service. Newsgroups are consequently considered to be interesting for further studies in this thesis. Chat is not expected to be considered as a relevant source by the consumer searching for information mostly because of the lack of concentration on one consumption related subject and because of the way the communication is organized, but also because of the lack of credibility due to the anonymity in cyberspace which seems to be a larger problem in open chat than is newsgroups.

Having established that at least some virtual communities qualify as 'real' communities, the next question then is why the consumer would use virtual communities instead of traditional reference groups. If the consumption activity or product is rare, virtual communities might have an advantage in comparison with traditional reference groups because virtual communities give an opportunity to communicate with persons over large distances at a small cost. It might also be easier to use the Internet to find and communicate with persons with the same interest than to find them via traditional sources. Regarding expectations in terms of quality and quantity of information, these are probably quite large, as there are more people to ask, and in some communities there probably are people, who are knowledgeable on the product class. On the other hand, in virtual communities on the Internet, it is more difficult to asses the quality (credibility) of the answers as they will mostly come from individuals not known in person by the information seeker. Asking for credentials might help but in cyberspace there is no guarantee that these are real and not made up. Therefore, building trust and maintaining long lasting relationships are even more important in cyberspace than in the physical world (Morrison and Firmstone, 2000). That it is possible that the information in virtual communities is perceived as usable and exiting is confirmed by research by Bickart and Schindler (2002). An experiment showed that consumers who have gathered information from Internet forums are more involved in the product than are consumers who have searched for information using other sources on the Internet.

Concluding to these remarks on virtual communities, it seems evident that the influence from these communities can only be expected to be of any significance if the information sought is rather objective and/or if the consumer has a relatively strong tie to the community and therefore trusts that the other members are able and willing to provide the needed information. It has already been put forward that the kind of community most likely to be able to produce such a feeling when it comes to exchange of consumer information is the newsgroup. However, not all members play the same role in newsgroups, and in groups centred on consumption, this is also not desirable as consumer exchange of information is dependent on some seeking, some possessing information.

A good framework for understanding different types of consumers participating in a virtual community is developed by Kozinets (1999). His framework is shown in figure 2.7 below.

Figure 2.7: Membership Segments in a Virtual Community



According to the framework presented in figure 2.7 there are two distinctions in membership of a virtual community: The first distinction refers to how central the consumption activity (or symbolic meaning of

the product) is to the person's self-image. This dimension is analogue to hedonic value of the product class to the consumer which was introduced in chapter 2.2.1. The second distinction refers to the persons' social tie to the community. The two axes result in four different membership segments in a virtual community:

The first type of member is the Tourist who is an individual with no strong interest in the product and a weak tie to the community. This could be a person who visits different virtual communities from time to time. According to Kozinets (1999), a Tourist can visit virtual communities in two different modes: recreational – for fun with no intention of social bonds – or informational – with the purpose of getting information in order to reach some goal not connected to the interaction. Based on the discussions in the above of the different characteristics of chatrooms, lists, and newsgroups, it is my hypothesis is that recreational mode happens more in chatrooms whereas informational mode is more common in newsgroups. Therefore, newsgroups are expected to be used for information search by the Tourist whose primary motivation for entering the newsgroup is to seek for advice there.

The second type of member is the Mingler. This is a person with strong social ties to the community but with no special interest in the product class discussed in the virtual community in question. This means that the social aspect in the communication is the most important motive for entering the community. Therefore, Minglers might be individuals who visit chatrooms and not so much newsgroups. This is in accordance with Kozinet's expectation that Minglers can be in recreational mode or in relational mode. Recreational mode happens when the Mingler enters the community for short-term personal gain (seeking entertainment) whereas relational mode happens when the Mingler wants long-term personal gain in terms of long-term social relationships. As Minglers have no deep interest in the product and mostly enters the community for social interaction, they probably do not use virtual communities primarily for search for product information and therefore, information search in virtual communities are not expected to be important to Minglers.

The third type of member is the Devotee. This person is very much linked to the product/consumption activity but has a weak tie to the community. Devotees can enter a community in Informational mode – for information for themselves – or in Transformational mode – wanting to change others' opinions on the product. Thus, a Devotee could be a person who visits a newsgroup in order to pose questions, answer to other members' questions, and follow the discussion. The Devotee can be expected to be quite active in the virtual communities he attends and he is expected to visit more communities related to the consumption activity. If he enters a newsgroup in transformational mode, he seeks to be an opinion leader whereas the Devotee entering the newsgroup in informational mode is seeking advice. As Devotees are involved in the product class, it seems more likely that they act as opinion leaders than opinion seekers in the communities they attend. This is supported by their weak tie to the community resulting in that they may not find the other members trustworthy. Also the weak tie means that they probably participate in discussions about the product class in question more out of self interest in the product than out of moral responsibility towards the community.

The fourth and final type of member is the Insider. The Insider is very interested in the product/consumption activity. In addition to this, the Insider feels strong ties to the community. An Insider could be a person visiting the same newsgroup often out of interest and, as time goes by, gets to know (some of) the other visitors. He can enter in two different modes: relational and transformational, depending on the motive for entering the newsgroup besides the interest in the product. The Insider will enter in relational mode if the desire to share his interest with others is the most important motive – that is, if he is socially integrated (Flynn et al., 1996) whereas he will enter in transformational mode if his primary purpose is to influence other individuals – that is, if he is socially independent. Therefore, the mode must partly depend on the personality of the Insider. These, very interested, persons may form a group of their own in time. One would also expect them to seek information from each other. Concluding to this, I would expect Insiders to be active in the virtual communities of which they are a member, both in answering and asking questions. This way, Insiders are important in keeping communi-

ties alive. As Insiders is the segment of members the most tied to the newsgroup this segment is expected to be the most influenced by information from the newsgroup and, therefore, also to be segment where substitution from search using traditional sources to search in the newsgroup is expected to be the largest. Furthermore, as the product class is important to this segment the segment may also be opinion leaders in the newsgroup along with Devotees. However, Devotees are less tied to the newsgroup and, therefore, may be less willing to spend time giving advice to others in the group.

The above shows that some virtual communities and especially newsgroups have potential to be places where consumers seek for information in relation to an upcoming purchase if the consumer feels that the other members of the virtual community are knowledgeable of the product and that the information they give is credible. As consumers visiting newsgroups are not equally knowledgeable of the product class in question and not equally tied to the group, the influence of and in the group is expected to be related to the person's role in the group, as defined by Kozinets (1999). Also, not all virtual communities qualify as communities in a traditional sense and, therefore, not all types of communities are expected to be equally important. Based on the research studied and observations on the Internet, I believe newsgroups to be the most interesting kind of virtual community to study in further detail as a source of consumer information prior to an upcoming purchase. Also, I expect the substitution to be largest for Insiders, for whom the ties to the newsgroup may be sufficiently strong to substitute even primary reference groups.

2.4.2 Non-individualized, Non-marketer Sources

The second type of information source displayed in figure 2.6 is non-individualized, non-marketer-dominated sources. On the Internet these are independent search agents, standardized information on products that consumer groups, magazines and state institutions place there, and information, put there by individuals to inform other individuals about their experiences with a product, a brand, or a retailer. The latter is sometimes in the form of the so-called 'hate-pages'.

Concerning the information from state institutions and consumer groups, the Internet is a way to make information available to the public. Information from such sources is often considered more credible than marketer-dominated information as there is no economic gain in question. However, for some institutions there may be a political aspect that makes the information perceived as subjective and thus less credible by the general audience. The information on these sites is often quite fact-oriented and there are often links to more sources of information.

The better possibilities for publishing information provided by the Internet means that the consumer will have far easier access to such non-marketer dominated structured information than has previously been the case. Therefore, the expectation is that the consumer uses more of this information on the Internet than via traditional sources.

Independent search agents find information on the Internet for the consumer and supplement the available information by helping the consumer to gather and compare information from numerous sources. The information that a search agent finds often is marketer-dominated information but the search agent as such is not marketer-dominated.

The reason for the consumer to use a search agent would be that he could get the information for comparison of alternatives at a lower cost than by visiting retail stores physically him-self, or even by accessing producers' websites directly via a browser. This issue was discussed theoretically in an influential article by Alba et al. (1997) and tested empirically using shopping for wine as an example, among others, by Lynch and Ariely (1998). Lynch and Ariely's study showed that transparency of information made shopping more enjoyable for the consumers and made them make better choices. As the search agent has no economic interest in giving false information, the consumer probably perceives information from a search agent as credible information. A study made by Urban et al. (1999) supported this notion. As the information is thus both usable and credible to the consumer, most consumers should perceive some benefit from using search agents in their information search.

The relative cost of using a search agent compared to the use of traditional means of information sources for comparison of products depends on whether the consumer perceives difficulties in using (and maybe finding) the search agent. The perceived difficulty, or cost, in using a search agent depends on the perceived cost in using the computer and the perceived cost of using the tool – the search agent. As the search agent is an easy way to get credible information, one should expect consumers to use search agents. I have not been able to find secondary data which confirm this.

The hate-pages are not considered to be influential in consumer search for information. This is mainly because the sender is probably not considered credible because of his self-interest in exposing negative information about the object in question.

The conclusion to this chapter is that the consumer can be expected to use non-individualized, non-marketer dominated sources on the Internet when searching for information because the availability of these sources has been increased by the Internet and because these sources are, in general, perceived as relatively credible. However, the use of these sources is limited by possible problems in finding the sources. Use of the sources for information search in relation to an upcoming purchase is, therefore, assumed to be dependent on use of the sources regularly.

2.4.3 Individualized, Marketer-Dominated Sources

Individualized, marketer-dominated sources on the Internet are represented by the possibility for the consumer to get in direct contact with producers and retailers via e-mail, as opposed to meeting in person, via telephone, or through mail. The Internet, this way, conveys easier contact between marketer (especially producer) and consumer. The consumer can contact the producer directly and get a direct answer. This is more feasible and more anonymous for the consumer than visiting in person or making a telephone call. Marketers also sponsor the so-called '(On-line) Brand Communities'. In figure 2.6, these touch on the non-marketer dominated, individualized sources. This is because the marketer can be more or less visible in the brand community. If he is not visible, consumers may perceive the community as 'their own'. In con-

trast to this, the marketer can be very visible in the community, and that, of course, must alter the perception of the community.

Using e-mail for communication with marketers is probably not seen as an advantage by all consumers. For some persons, having to use a computer and having to pose a question to a stranger in writing makes the possibility far less attractive than talking to a salesperson face-to-face or on the telephone. Still, by some consumers is it deemed convenient not to meet salespersons face-to-face (Venkatesh, 1998; Wolfinbarger and Gilly, 2000). Wolfinbarger and Gilly's study showed that the attitude towards meeting sales personnel is closely linked to the personality of the consumer. Persons with high internal locus of control prefer the absence of physical contact with salespersons whereas individuals with high external locus of control like the social contact with sales personnel. Data from the GVU (1998) study show the same kind of result as 62.7% of the respondents indicate to shop (or search for information) on the Internet because they feel no pressure from sales people.

Several of the researchers referred to in chapter 2.3.2 emphasized that if consumers were to use the Internet for information search, marketers would have to activate consumers visiting their sites. One way of activating consumers which is tied to relationship marketing is running a 'Brand Community'. Brand Communities are online communities hosted by companies (McWilliam, 2000). A more narrow definition is given by Muniz et al. (2001): 'a specialized, non-geographically bound community, based on a structured set of social relationships among admirers of a brand'. Muniz et al. thus emphasize that the setting has to qualify as a community in terms of social relations, and that the members are not just buyers, but admirers of the brand in question.

Brand communities are, in some ways, similar to the communities discussed in chapter 2.4.1 and yet, in other ways, they are completely different. Returning to the characteristics of virtual communities as seen by Armstrong et al. (1996) which were presented in chapter 2.4.1, there does not seem to be a difference between the two kinds of communities concerning being a 'real' community. However, there are two important differences with respect to their usability for the 'average' consumer in

his search for information: Brand communities are centred on a specific brand and, therefore, mostly interesting to consumers who feel connected to a specific brand. In Kozinet's terminology, we would expect many Devotees in Brand Communities. Secondly, brand communities are rooted in the interest of the marketer acting as a tool for creating and maintaining loyalty (McAlexancer, 2002; McWilliam, 2000). Because of the need for interest in a specific brand and not a product class, the potential user group for this kind of community is smaller than the open communities. The credibility of the community in terms of providing objective information on which brand to choose may be hampered because consumers may think that the community is mainly for very loyal consumers ('admirers') who are not objective in their evaluations of the products. In addition to this, credibility of the communication is hampered by the fact that the sponsor is able to control the discussion going on in the community. On the other hand, the awareness of the groups can be quite large as marketers often promote the communities as part of their marketing effort.

In relation to the current study, the interesting aspect of this type of source is not judged to be the ability to write an e-mail to a producer or a retailer instead of meeting them face-to-face or contacting them per telephone, as sending an e-mail, after all, is not very different from calling on the telephone, which is also a rather anonymous way of contact between people, not knowing each other. Furthermore, because brand communities are targeted at consumers who are, in advance, interested in a specific brand and are sponsored by marketers, these are not expected to be important sources of information in relation to an upcoming purchase. However, they can be important to consumers after the purchase, as argued by McWilliam (2000), and therefore interesting in relation to ongoing search. More interesting in the current context is if the information formerly gathered by visiting retail shops and talking to the sales personnel has been replaced by information search in newsgroups on the Internet as these groups are also able to provide individualized information.

2.4.4 Non-individualized, Marketer Dominated Sources

Non-individualized, marketer-dominated sources on the Internet are websites run by producers (brands) or retailers and subject-oriented portals. Information search using this type of source is easy for the consumer in terms of psychological and physical effort. This is because visiting the sites is rather passive once the sites have been found unless the consumer engages in one of the individualized possibilities that are often offered on these sites (brand communities, e-mailing and so on).

As producers and retailers have a clear purpose of economic gain from presenting the information to the consumer, this kind of information is generally not considered to be perceived as very credible unless the information concerns objectively measurable attributes. However, websites run by producers and retailers provide much information to the consumer in terms of available products, features, and retail outlets and prices, and thus, the lack of trustworthiness and objectivity may be outweighed by a higher level of knowledge. The importance of trustworthiness on the Internet is supported by Ward and Lee (1999) who showed that, to overcome the perceived risk of using unknown sources of information on the Internet, the consumer uses brands (considered trustworthy) as quality indicators more on the Internet than off-line. Studies made by Brynjolfsson et al. (1999, 2000) also supported the notion that consumers use brands as quality/trust indicators. Data from GVU (1998) show the same pattern. 9.8% of the respondents visit websites for the retail shops they often visit physically and 44.5% state that this happens sometimes. Research from e-bates based on data from Harris Interactive states that 19% of on-line buyers are brand loyalists in the sense that they visit brand-related websites directly on the Internet. This all supports the proposition that consumers use marketer-controlled websites for information purposes, and especially branded websites which are perceived as trustworthy.

In spite of the credibility problem this kind of information source is thus expected to be used by consumers seeking information on the Internet. This is because it is easier to visit many different producers and retailers on the Internet than in the real world and because the level of knowledge is high. Therefore, one would expect the consumer to visit web-

sites controlled by marketers as a part of their information search. As a consequence of the problems in assessing the credibility of the source which have placed the information on the Internet, the use of this kind of source when searching for information in relation to an upcoming purchase is expected to be related to the level of use of the source in general.

A second type of non-individualized, marketer-dominated source is subject-oriented websites. These websites are not (openly) linked to a specific brand and they are often positioned as cybermagazines in the sense that they are a mix of feature articles and commercials (banner-ads and pop up commercials).

The information from this kind of source is probably considered more credible than the information from producer and retailer websites due to their independent status. It is expected that they are used, to some extent, in the search for consumer information even though it often will require some effort to find exact information on these sites. Therefore, one might also expect that the sites are mainly used by consumers who visit the sites regularly.

Concluding to the above, it can be expected that consumers use mass-marketing from companies put on the Internet when searching for information in relation to an upcoming purchase and that the level of this kind of use is related to the level of use in general.

2.4.5 Expected Use of the Sources on the Internet

The various types of information sources on the Internet in many ways resemble the sources available to the consumer off-line. The sources, therefore, are expected to be used in much the same way by the consumer as are the off-line sources, and the use of the sources thus depends on how and to what extent the consumer uses the various types of sources on the Internet in general. The discussion of the determinants of choice of type of information source is delayed to the following chapter 2.5. However, with respect to the use of the sources on the Internet, it is possible to put forward a few issues based on the presentation of the sources in the preceding sections.

First, there is an important difference between non-marketer dominated, personalized sources on the Internet and non-marketer dominated, personalized sources off-line: Mostly, the consumer never meets persons he knows from virtual communities on the Internet face-to-face, or know their real identity for the matter. There is, nonetheless, evidence (*e.g.* Wellman and Gulia, 1999) that persons are able to feel close to the other members of a virtual community in spite of this lack of personal knowledge of the other person. There is also research holding the proposition that consumers use virtual communities in different ways when searching for information depending on their felt tie to the product and to the group (Kozinets, 1999). The question to pose in relation to consumer search for information then is whether newsgroups centred on consumption issues (which is deemed the most interesting type of virtual community when it comes to consumer information search) can be found to produce the needed feeling of belonging and reciprocity and thereby is able to replace sources formerly used and, if this is the case, which of the four types of sources in figure 2.5 the newsgroup has replaced.

Second, it was put forward that the consumer is able to visit more non-personalized sources on the Internet than off-line and that producers, retailers, and consumer information bodies have easier access to give consumers opportunity to see their communication on the Internet than they are using traditional media. Especially, consumer information bodies should experience an advantage from the Internet because the consumer can more easily find them and visit them on the Internet. However, if the consumer does not know that this type of source is available, he probably will not use this source. It is therefore to be expected that the pattern in the type of sources used for information search in relation to an upcoming purchase is not much different from pattern of sources on the Internet used in general. Therefore, the pattern of use of the Internet in general is expected to be important in the explanation of use of the various types of information sources on the Internet when searching for information in relation to an upcoming purchase.

2.5 Determinants of Choice of Information Source

In chapter 2.4, the consumer's use of different types of information sources has been discussed mainly based on characteristics of the source and the expected benefit and cost of using the source to the consumer in general.

However, characteristics on the receiving side of the communication dyad also play a role in the choice of information source. This is at the heart of this thesis, as we want to find determinants of consumer choice of the Internet, and especially, virtual communities on the Internet, as sources of information needed prior to making a purchase decision instead of other sources available to him. Therefore, this chapter is designated to a discussion of differences among consumers as possible determinants of choice of information source.

In chapter 2.2, the theory presented focused on search performed in order to make a choice of brand rather than a choice of source of information. The reason behind this was that the goal in that chapter was to get a better understanding of the determinants of search to take place at all. However, as the goal in this chapter is to achieve a better understanding of choice of source we shall turn to theory focusing on this. Theorizing on choice of information source on a general level, Wilkie and Dickson (1985) focus on the shopping process seen as search for information, using various sources, before making a purchase decision. Wilkie and Dickson do so in order to find out which role the store visit and the information from the store in terms of advertisements and personnel plays during the buying process. Like other researchers, belonging to the cognitive school, Wilkie and Dickson assume that the consumer engages in the shopping process out of need rather than for hedonic reasons, and that the consumer is quite conscious about what he does during this process. Wilkie and Dickson (1985) also state that the consumer will use several sources of information during the process depending on the benefit he expects from the information he gets relatively to the cost there is involved in getting the information[12].

[12] By cost is meant the broader definition: Money, time and psychological effort

The next step is to investigate why the same information source is perceived differently by different consumers. That is: How are these costs and benefits perceived differently by consumers depending on context and personality? The determinants of perception of different information sources are in the following put into three categories: Difficulty in making the decision, differences in the perceived cost and benefit of using a given source, and characteristics of the product in question. In addition to these determinants, it is necessary to consider that using the Internet for information search requires a change in behaviour from using the traditional sources to using the sources on the Internet. For some consumers this change may be easier than it is for others. This is the topic of the fourth section in this chapter. Finally, expectations regarding determinants of level of use of the Internet and especially virtual communities for information search prior to making a purchase decision are summed up in the last section.

2.5.1 The Decision and Tie to the Source

In chapter 2.4 it was discussed how the consumer relates differently to the various types of sources on the Internet. Especially, different modes of entering the virtual communities on the Internet were highlighted. What remains is to look into the question of what makes the consumer decide what kind of relation is deemed necessary in order to find the needed information. This is the subject of this chapter.

In a study from 1997, Duhan et al. investigated which factors influence whether the consumer seeking for information uses what is termed strong-tie sources (family and friends) or weak-tie sources (acquaintances, for instance in the form of retail personnel). They found that the likelihood of using a strong-tie source increased if the task was perceived as difficult (determined by lack of prior knowledge) and that the likelihood of using a weak-tie source increased if objective information was considered important, or if the consumer had high subjective knowledge (felt, she had high knowledge). This indicates that consumers seek information with the aim of reducing the two different types of risk presented in chapter 2.2.1 – functional and emotional – from different types of sources. Information on functional risk aspects can be gathered among weak-tie sources because the information sought is rather

objective and, therefore, the important thing is that the source possesses this knowledge and not that the source shares values with or knows the preferences of the consumer. In contrast to this, information gathered to reduce emotional risk has to come from sources which are trustworthy, or of which the consumer knows the personal preferences and these are mostly strong-tie sources.

As described in chapter 2.2.2, Chhabra and Olshavsky's (1986) study suggests that the relationship between functional risk and perceived benefit of search might not be linear but rather a curve with a maximum at some level of perceived risk and declining at larger levels of perceived risk as search then is replaced by other-based decision making because the perceived cost of establishing the needed information is high. This means that consumers with very limited knowledge can be expected to seek information from sources which they perceive as knowledgeable of the product class in question, the consumer's preferences, and who are also credible. The kind of source living up to these demands is the primary reference group; that is: family and friends which are strong-tie sources.

Another empirical study (Furse, Punj, and Stewart, 1984) supports the above theory on the use of strong-tie and weak-tie sources. The aim of the study was to find out whether a segmentation of individuals, based on information search before purchase is possible. The researched individuals had recently bought a car. Furse, Punj, and Stewart identified 6 different clusters. These are shown in Figure 2.8 below.

Figure 2.8: Clusters in information search

Group	Percentage	Label	Description
1	26%	Low Search	Experienced. No (or very little) information search due to high experience and therefore little need for information (because of little perceived risk).
2	19%	Purchase Pal Assisted	Information search with a purchase pal. Little experience, little subjective knowledge. Other Based decision (father) – might be asking group 4.
3	5%	High Search	Spend greatest amount of time searching. Uses many sources. Involves others. No confidence in own judgement. High expected utility from search. This group was the least satisfied with their last purchase.
4	12%	Self-Reliant Shopper	Spend much time searching, but mostly out-of-store and not involving others. This group tends to rely on their own ability to judge. Large evoked set.
5	5%	Retail Shopper	Rely heavily on information from the retailer. More than one decision maker (family). Large evoked set.
6	32%	Moderate Search	Moderate on all search variables

Source: Furse, Punj and Stewart, 1984.

Figure 2.8 again confirms the results found in chapter 2.2 and the studies cited in this chapter. Experienced consumers tend to search little. The amount of search among less experienced consumers depends on, among other things, perceived knowledge. Consumers who have little knowledge search for information either on their own using many different sources including retail sales personnel, or with help from others whom they feel confident will help the make the right decision (Purchase Pal – a strong-tie source). This is analogue to the other-based decision discussed in chapter 2.2.2. Consumers with more knowledge, who feel confident in their ability to find the right information, search for information on their own using various sources, but not retail sales personnel. A small group relies heavily on information from the retailer, which is a weak-tie source. This group has a large evoked set and, therefore, also must possess a fair amount of knowledge on the product class.

Relating the information in this chapter to the characteristics of the Internet as an information source, we are able to conclude the following:

It seems that consumers who feel confident in their own knowledge about the product class and choice criteria are more prone to using sources where they are free of personal advice from others and where they believe to find the most information (knowledge). Therefore, it can be expected that knowledge about the product is positively related to use of the Internet for information search. The opposite may be the case if the decision is perceived as difficult, as the consumer then relies on personalized sources.

The insecure consumer seems to rely on two types of individualized sources if the decision is difficult: Strong-tie sources if the consumer possesses very little knowledge, and information from sales personnel if the level of knowledge is higher, or the information sought is objective. Combining this with the framework presented in chapter 2.4.1, strong-tie sources on the Internet relevant to consumer search can be traditional strong-tie sources where e-mail has replaced, or supplemented, traditional means of communication and virtual communities, in which the consumer is an Insider or a Mingler (both have strong social ties to the community). As described in chapter 2.4.1, some consumers may consider newsgroups on the Internet to be a strong-tie source to them. This is expected to be the case mostly for Insiders but maybe also Minglers. Newsgroups may also be used for gathering objective information previously sought in retailing by members who have only weak ties to the group, the segments Tourists, and maybe Devotees, as the ability to judge trustworthiness which comes from knowing persons in real life is then of less importance.

In sum, the preceding discussion shows that different types of decisions require different types of sources for information. Difficult decisions where the consumer possesses only a limited amount of knowledge and/or a decision which is perceived as emotionally risky, require strong-tie sources, whereas other decisions can be based on information from weak-tie sources. Some consumers prefer to seek their information using personal sources while others prefer to use non-personal sources.

The preference for non-personal sources seems to be related to product knowledge. This again leads to expectation of a relation between product knowledge and use of the Internet on the general level. Regarding the use of personalized sources on the Internet, newsgroups may be substituting personalized sources in two different ways: As a strong-tie source to members who feel a strong tie to the virtual community in question (*e.g.* a newsgroup) and as a weak-tie source replacing information from producers and sales personnel for all types of members.

2.5.2 Differences in Costs of Information Search

Till now, the focus has been on differences in the buying decision. There is, however, also evidence showing that different consumers will perceive the cost of using an information source differently. This chapter looks into this evidence.

Some researchers point to the fact that for some people, *time* is the scarce resource, not money as assumed in traditional theory on rational choice (Lehman, 1999). Therefore, the perception of time might play a different role in the decision process regarding where and how to find information for different consumers.

There are also numerous researchers (Andersen et al.2000; Bloch and Richins, 1983; Holbrook and Hirschman, 1982; Malhotra 1983; Nicosia, 1986) who hold – despite different angels on the reasons why - the proposition that, to some consumers at least, *shopping,* and as a part of this information search, is a form of entertainment and an important part of the consumer's social life. Therefore, the perception of benefit and cost incurred by using a given information source may be related to the source's ability to provide entertainment and social contact during the course of search.

Because of this different view on search for information, different consumers will perceive the cost of using the same source differently depending on how they perceive the use of the source and the tasks involved in information search. I have not been able to find theory incorporating this directly into information search. However, as search for

information is often a part of, or the reason for, a shopping trip, research into attitudes towards shopping may be helpful on this issue.

Stone (1954) developed a typology on shoppers based on housewives (women) in Chicago and Lavik (1984) used this typology to classify Norwegian consumers (men and women) into shopping types. The shopping types were[13]:

- The *economic* shopper. This type of shopper is preoccupied with price, quality, and selection. Economic shoppers are younger, well-educated women in the work force.

- The *personalizing* consumer who likes to shop because of the social contact with other people. The personalizing consumer can be stereotyped as elderly women with a small household income. They are more often housewives and possess no or little education.

- The *ethical* consumer who supports certain types of stores. There are no demographic variables designating this consumer.

- The *apathetic (or practical)* consumer who wants shopping to be as simple as possible. Men are overrepresented in this group.

Even though the thesis is concerned with information gathering in extensive problem solving and the framework in the above was implemented for grocery shopping – that is products mostly chosen after only limited problem solving – the idea in the classification can be used in the general understanding of information search.

Looking at the choice of the Internet as information source the possibilities on the Internet must me more appealing to consumers for whom time is a scarce resource, as using the Internet saves time and money for transportation. Using the framework above these are economic and apathetic shoppers.

[13] The demographic description originates from the Norwegian study because it is much more recent and also closer in terms of culture to Denmark.

In contrast to this, the personalizing shopper using the Internet probably would miss the personal contact in the physical world. The question is whether the opportunities on the Internet can match the shopping mall in terms of entertainment for consumers, to whom the search for information is an excuse to go on a shopping trip. On the Internet there is no smell and no physical contact with other people. Especially the last issue is important, according to Miller (1998a and 1998b) and Bloch et al. (1991). On the other hand, Bloch et al. (1991) point to the need for more stores of the same kind in one mall because of the need for information search in different stores. In this sense the Internet may possess an advantage, as the consumer here is able to visit many stores without moving physically.

Maher, Marks and Grimm (1997) also found that shopping enjoyment is positively linked to attitude towards traditional shopping channels: retail and catalogue but not to the alternative channels: television and on-line. Finally, they found a positive link between attitude towards the channel and use of the channel. This means that women who like to shop in the traditional sense tend to choose traditional channels, because the inconvenience in using this channel as compared to using the Internet is not important. The study was not able to conclude on which kind of shopper would choose the new channels and why. This could be because online shopping and, as a part of this, search for information on the Internet, was a relatively new possibility in 1996 when the data was gathered and, therefore, the respondents might have had trouble evaluating this shopping channel. Use of the Internet has become far more convenient and common since then and, therefore, research today may produce different results in this last matter.

Returning to the understanding of the role of the perception of time in the choice of source, Maher et al. (1997) investigated the relation between role overload[14], feeling of time pressure, importance of convenience, and shopping enjoyment on the one hand, and attitude towards and use of non-traditional retail channels on the other hand. Maher et al.

[14] Defined as too many roles: Wife, mother, work etc.

asked women only in their study. They found a relation between 'role overload' and 'perceived time pressure'. Furthermore, they found that 'perceived time pressure' and 'importance of shopping convenience' had a negative impact on shopping enjoyment. This again indicates that women who enjoy shopping do not feel time pressure generally and they do not value convenience in shopping. This is in accordance with general economic theory in the sense that only scarce resources have cost relevance in the choice between alternatives. Maher et al.'s results are in accordance with the above-cited results by Stone and Larvik, because economic shoppers are described as younger, well-educated women in the work force – and, therefore, also are very likely to feel time pressure and role overload.

The studies cited in this chapter point to the fact that different consumers do not evaluate different sources in the same way because they have different scarce resources in terms of money and time and some consumers regard shopping, of which search for information on different alternatives can be a part, as entertainment involving social contact. This means that to consumers, for whom time is a scarce resource, the Internet should be more appealing. In contrast to this, consumers, to whom shopping is entertaining, should use the Internet less, all other things equal, than consumers who do not seek this kind of socially linked entertainment.

2.5.3 Type of Product

Products are not all alike. Some products are mainly chosen based on attributes which are objectively assessable, whereas other products are mainly chosen from emotional cues, or on attributes which are difficult to assess, as was discussed in chapter 2.2.2.

Klein (1998) elaborated on the issue whether the potential of the Internet would be greater for 'search goods' or 'experience goods'. Search goods are goods for which full information for dominant product attributes can be known prior to purchase, whereas experience goods are goods for which the cost/difficulty of achieving information is greater than direct product experience, or where the attributes cannot be known in advance (Klein 1998).The conclusion was that for experience goods

there is a great potential if the Internet makes it possible to experience the good at a reasonable cost before purchasing (a movie for example) and thereby reduces the consumer's perceived risk. Using this opportunity, the consumer's perceived benefit of search will be larger when using the Internet as an information source and, therefore, he will increase the amount of search. Other researchers (Ward and Lee, 1999) emphasize that this certainly depends on the kind of experience the consumer is looking for. Not all experiences (taste and smell, for example) are transferable on the Internet. The emotional experiences are difficult to transfer but the consumer might find interest in asking others, whom he knows share his preferences concerning the experience at hand.

Another, in some ways similar, distinction is made in Ratchford (1987). Here, products are divided into 'feel' versus 'think' products on one dimension, and high versus low involvement on the other dimension. 'Think' products are primarily evaluated on functional facts and not on feelings, whereas 'feel' products are evaluated relatively more on feelings than on facts. Using the theory presented in chapter 2.5.1 in predicting which type of sources on the Internet, consumers would be expected to use for search for information, it is clear that the consumer would tend to use strong-tie sources for products which are selected from feelings, whereas weak-tie sources are more likely to be used for products which are evaluated on functional facts because credibility and knowledge of personal preferences are more important concerning subjective cues than for objective cues. This means that the expectations on use of the sources on the Internet for products evaluated on subjective cues are about the same as the expectations on the use of information sources when the knowledge in advance is small. For this kind of product virtual communities may be substituting personalized sources for members who feel a relatively strong tie to the community; that Insiders and Minglers. For products, on which the choice is based on objective cues the use of virtual communities is expected to be less dependent on felt tie to the group.

The above also means that research trying to predict which consumers will use the Internet for information search should be careful and cover

different types of products in the test of the model, or be aware that the study is only valid for certain products.

2.5.4 Facilitating Factors

Using the Internet for information search requires a shift in behaviour from using the sources that the consumer was socialized into using. Therefore, an important, final issue in the prediction of which consumers would use the Internet for information search is how consumers learn and change between where they seek the information they need to choose between alternatives. The required change in behaviour incurs different costs to different consumers and, therefore, not all consumers will make the change in behaviour equally fast, if at all.

One field in this area is the consumer socialization process – that the child learns consumer skills during childhood (For a review see: John, 1999). Bellenger (1982) has investigated consumer socialization from a cognitivist point of view. He both incorporated intrapersonal theories (that the consumer cognitively makes his choices) and interpersonal (that others influence the choices) theories and tried to develop a general model on how these two fields shape the behaviour of an individual regarding store patronage. The model explains store choice as a result of both socialization and experience via cognitive orientation towards shopping as well as the store. As the individual builds experience socialization plays a lesser role indicating that experienced shoppers would find it easier to shift to the Internet provided they can use their search strategy there. Venkatesh (1998) judges the probability of this as he discusses the influence of former socialization into shopping in department stores and supermarkets on the cost of shifting to the use of the Internet. He concludes that existing social memory is of limited use in the question on how to navigate in the cybermarketscape. This is because the cybermarketscape is so fundamentally different from the physical landscape that the shortest and best paths may be very different than the ones used in the physical world. Therefore, Venkatesh predicts that the cost of shifting to the Internet will be larger for persons with a large amount of shopping experience compared to persons with a more limited amount of shopping experience because the latter have no cheaper alternative in the traditional sources.

Applied to the use of the Internet for information search purposes the above means that one would expect younger consumers to be more prone to using the Internet than older consumers are because they are less socialized into shopping in conventional retail stores than older consumers are. Furthermore, one would expect individuals who generally use the Internet more to be more likely to use the Internet for information search, as well, as they are socialized into using the Internet. This notion has been supported by several empirical studies (Nie and Ebring, 2000, GVU Gatech, 1998).

The concepts of dialectic research can also be helpful in understanding change in the sources the consumer use for Information search. This is because of the emphasis dialectic research places on the material environment helping or slowing change in behaviour. To understand the change in information search patterns as taking place in interaction with the surroundings, and the Internet as a part of these surroundings, is appealing. Relating the concept of totality and, therefore, the need to consider the consumer in his or her social setting to the use of the Internet for information search also seems appealing, as the use of the Internet does not happen in a vacuum but as an integrated part of the life the consumer leads. A change in the way consumers choose to search for information might be induced by contradictions between the life of some individuals and their possibilities for finding information. As discussed in chapter 2.5.2, to some, time, and not money is the scarce resource, and information search in relation to an upcoming purchase may mean spending hours of scarce spare time on transportation in order to get to a shopping area during the time when the shops are open. These contradictions might induce use of the possibilities for finding relevant information on the Internet because this way of finding information saves time for physical transportation, and can also be performed at all hours.

Another aspect of change in consumer behaviour is that consumers do not perceive change equally in terms of incurred cost. Consumers who like to try new places to shop can be expected to use the Internet more for information search. This is because these consumers can be expected

to perceive a change in information search behaviour as less costly than consumers who do not like to change their consumption behaviour.

In the above, cost of behavioural change was considered. However, for this change to happen out of free will the change also has to incur benefits to the consumer which are larger than the costs incurred. One of the benefits from using the Internet for the consumer put forward in chapter 2.3 was the constant availability of information at a low cost regardless of physical distance. However, in order to exploit this opportunity the consumer has to be able to find the information and prefer this freedom in finding information. Ariely (2000) showed that consumer control over information flow, defined as a possibility to choose and alter search strategy during the process, enhanced the consumer's immediate and long-term benefit from the search process. This is especially significant for consumers who engage in the same task numerous times (get experienced). Hodkinson et al. (2000) explain part of this competence by the wayfinding paradigm which states that individuals, as they gain experience, will form knowledge about where they are and how they got there. Very experienced travellers keep a mental image of parts of the topography. When the route is known people have a tendency to keep to familiar routes as long as this is possible although this is under the influence of situation and the person's optimal stimulation level (OSL). On the Internet, this means that the experienced user would know how to get information because he knows where to look. If individuals tend to follow familiar routes, Internet users may, therefore, be loyal towards sites and search engines they know, and search strategies they are familiar with instead of exploring the Internet, as suggested by Hoffman and Novak. Also, persons who have used the Internet longer will easier find the information, they need. This again leads to an expectation of a relation between the amount of time an individual has had access to the Internet and use of the Internet for information search. Therefore, the change is probably perceived as less costly by consumers who know their 'way around' the Internet.

Using the Internet as a consumer information source is, in fact, an innovation in mode of search for the consumer. In speculating on which type of consumer would be the first to use the new possibility for search for

information the characteristics of the innovation presented by Rogers (1968) is a useful framework. Using Rogers' framework a new type of information source will be more easily adopted if it

1. Has a *relative advantage* to the consumer searching for information. Therefore, the type of consumer, who is most likely to adopt the Internet as an information source, is consumers who value the possibilities to choose time, place, and research strategy freely on the Internet. In the chapter 2.5.2 and in the current chapter, this has been shown to relate to role overload. Consumers to whom time is a scarce resource will perceive a larger advantage in terms of lower costs when using the Internet than will consumers to whom time is not a scarce resource. Likewise, not all consumers will expect to find the same amount of credible information on the Internet. Therefore, the relative advantage will also be different in this sense, depending on the consumer seeking for information.

2. Is *compatible* with existing values and past experience. This means that consumers who in general are positive towards the Internet and has experience in using the Internet and/or computers will be more inclined to search for product-related information on the Internet, as this is compatible with what they usually do. Compatibility in terms of computer and/or Internet use is one form of compatibility. Another form of compatibility is that information search using the Internet can be fit in with the way the consumer is used to search for information on alternatives. This must be the case for consumers who are used to do their information search from home as a part of buying from home. The research presented by Hoffman and Novak in chapter 2.3 supports this. Along the same line of arguments is the lack of compatibility of information search on the Internet with shopping, seen as a social activity.

3. Has *low complexity*. This again points to consumers using computers and the Internet in general. These consumers will find it much easier to search for information on the Internet than persons who are not regular Internet users, or who are not used to finding information from home. Also, age will be of importance as younger persons easier learn new skills and have less experience in the traditional sources than older persons do.

4. Can be *tried* on a limited basis. This is related to possession of, or access to, a computer and/or the Internet. This again points to an easier adaptation of information search on the Internet for consumers who, in advance, have access to the Internet.

5. Is high on *communicability*. This has mainly to do with the Internet as such. Individuals with interest in the Internet would be more able to communicate about the subject – how they use the Internet, how they find information or visit many stores than individuals who do not share these interests.

The results presented in this chapter have shown that it is not equally appealing to all consumers to adopt a new way of searching for information. Consumers, to whom the adoption is more advantageous, or to whom the adoption requires less change in established forms of behaviour are expected to use the Internet more for information search than other groups. As younger consumers have less established patterns of behaviour, use of the Internet for information search is expected to be inversely related to age. Consumers who like to use the Internet and/or are skilled users of the Internet can more easily use the Internet for information search than non-users and inexperienced users are able to. Consumers who experience that time is a scarce resource, or who like to find information from home will experience a larger advantage from using the Internet for information search. Finally, consumers who like to try new ways of finding information will perceive a switch to using the Internet for information search less costly than consumers who do not like new ways of finding information. Summing up, all these concepts must be positively related to the use of the Internet for information search.

2.5.5 Expectations About Use of Internet Sources

The discussion of the determinants of use of the Internet has led to expectations in terms of the consumer using the Internet as such for information search, and in terms of the use of the various types of information sources on the Internet.

The type of decision is an important determinant of the choice of information source. If the decision is difficult to make for the consumer, if he is not knowledgeable of the product class, or the choice has to be made based on subjective information he will use strong-tie sources. Strong-tie sources in the physical world are mostly primary reference groups as friends and family. If the information sought for is objective, or the decision is not important to the consumer, the consumer is more likely to use weak-tie sources when searching for information. Some of the consumers who are insecure rely on retailers to give the information needed. Again, it is possible that some of the personal information given this way is replaced by information from newsgroups on the Internet. The question here is if virtual communities on the Internet are able to replace strong-tie sources, or sources perceived as knowledgeable by the consumer.

If the consumer feels confident in finding the needed information himself, either because he is knowledgeable about the product, or the information sought for is objective, he will search using weak-tie sources like retail personnel, brochures, and information in magazines. The weak-tie sources on the Internet are, on average, less familiar to the consumer – or have a weaker tie - than the weak-tie sources off-line because of the larger psychological distance and the inability to judge the source's trustworthiness. Therefore, it can be expected that consumers will use the Internet more for Information search if they are knowledgeable about the product and therefore also to a higher degree able to judge the information given. If the consumer is interested in the product class, hed has larger advantage of the constant availability of the information on the Internet, and, accordingly, interest in the product class is expected to augment use of the Internet for information search.

Different types of shoppers have a different perception of the costs involved in information search. To some individuals searching for information is not regarded as a necessary job but rather as a chance to get in contact with other people, while others regard search for information in relation to an upcoming search as a task to be performed at the least cost possible. Use of the Internet for information search implies less cost to the consumer in terms of time, money, and needed effort to move physically between the places he wants to visit. Based on the above, this advantage may, to some consumers, be off-set by lack of social contact as a part of search for information, while using the Internet.

Not all products are equally suited for information search on the Internet. Some product features which require experience in real life cannot be displayed on the Internet, and therefore, it does not make sense to search for these features on the Internet directly. What does make sense is for the consumer to rely on other people's experience, and their opinion. This is a possibility on the Internet. Along the same line of reasoning is the distinction between features that can be objectively evaluated, and features that are evaluated based on emotions. For subjective and experience based evaluation to be useful, the source has to be perceived as trustworthy and knowing about the information seekers' preferences or vice versa. Therefore, the usefulness of the various sources on the Internet probably depends on the product class in question.

Use of the Internet for information search in relation to an upcoming purchase requires a change from the consumption behaviour he has been socialized into. Roger's framework states that this change will be easier if the change implies an advantage to the consumer, is simple to perform, is not hard to do, and is compatible with existing values and past behaviour. It is self evident that for consumers who are used to and like using the Internet, the shift is far easier than for consumers who first have to learn to use the Internet and how to find information on the Internet. Also, consumers to whom the advantages of using the Internet are larger than for others, can be expected to use the Internet more for information search. The advantages of the Internet are, as mentioned in the above that the consumer is able to search from his home and thereby saves time. This means that consumers, who like to shop from home

and/or who feel that time is a scarce resource, can be expected to use the Internet more as an information source. Finally, consumers who see an advantage in trying new ways of shopping can be expected to feel that the cost of changing behaviour is smaller, and therefore also are expected to use the Internet more for information search purposes.

2.6 Hypotheses about Consumer Use of the Internet

Consumers seek for information because they perceive this as useful. In chapter 2.2 we saw that traditional research in consumer behaviour defines usefulness mainly in terms of ability to reduce functional and emotional risk related to an upcoming purchase. However, the consumer sometimes also seeks information on products without an upcoming purchase in mind. The reason for this can lie in a general interest in the product class, resulting in that the consumer perceives search for information concerning the product interesting in itself. The consumer thus sometimes searches for information without an upcoming purchase in mind. Likewise, he sometimes does not search for information when he has an upcoming purchase in mind. There are several explanations to this. If the consumer finds the decision very difficult to make, he will sometimes retreat to others to make the decision for him. Another explanation for lack of search is that some products are not suitable for search as their features are not searchable. A third explanation is that the consumer, if he is interested in the product as described above, often is in possession of enough knowledge in order to make the decision because he generally is open to/seeking for information on the product class in question. All this was illustrated in the model displayed in figure 2.3.

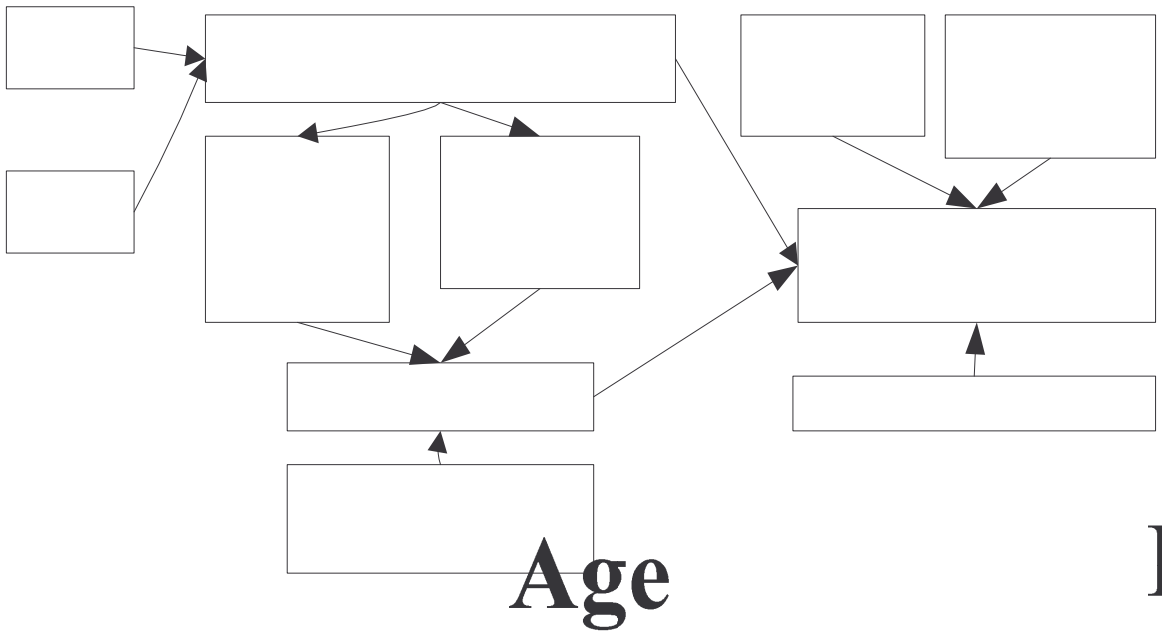
Answering the questions whether and why the consumer searches for information do, however, not answer the main question in this thesis: Why would a consumer use the Internet for finding this information? Returning to figure 2.3, the answer to this question would be: Because the consumer would expect a higher benefit from using the Internet than from using traditional sources. Based on the discussion in chapter 2.2, this expected benefit will be higher if the Internet is perceived as providing better risk reduction in terms of increased knowledge about func-

tional and/or emotional consequences of choosing each of the available alternatives provided the consumer sees these consequences as important. Benefit can also arise from learning about the product class if the consumer finds this of interest in general as described in chapter 2.2.2. Likewise, it was put forward in section 2.3.2 that the consumer can find pleasure in using the Internet for finding information, and thereby increasing the benefit of the search in itself. In contrast to this, it was argued in chapter 2.5.2 that there may be other factors such as enjoyment in shopping, and as a part of/excuse for that, information search, in the physical world lowering the perceived benefit of using the Internet for information search.

In chapter 2.3.1 and 2.5.2, it was put forward that an important reason for using the Internet for information search can be that on the Internet there is much information available at a low cost. However, in chapter 2.5.1, it was argued that it is important that the information is perceived as credible, especially if the consumer is not knowledgeable of the product class in question and, therefore, has problems in evaluating the quality of the information. Furthermore, it was argued in chapter 2.5.4 that the perceived cost of using the Internet may not be equal for all consumers, as some consumers like using the Internet and, therefore, may see a gain in benefit from using the Internet in itself. Based on the theory presented in chapter 2.3.2, it can be expected that mainly younger persons and more experienced users of the Internet will find such a pleasure in using the Internet. Likewise, not all consumers are expected to perceive the same cost of using the Internet when adapting the new source and when searching for information, as was discussed in section 2.5.4.

Based on the above, we are now ready to propose a model explaining the determinants of consumer use of the Internet for information search. This model is presented in figure 2.9 below.

Figure 2.9: Internet Search Model (ISM)



As figure 2.9 shows, use of the Internet as an information source is expected to be determined by perceived (relative) cost of using the Internet, by the importance of the Internet in the person's life, by interest in (hedonic value) and by knowledge about the product, and by the type of shopper, the person is.

The relations in the model are expected to be as follows, based on the previous sections in this chapter and as explained in the above:

If a person is interested in the product class, he will find more use for the constant availability of the information on the Internet, and therefore, also want to use the Internet more for information search. Likewise, if the consumer has much knowledge about a product, he will have more skill in locating relevant information and therefore gain more benefit from being able to control his information flow via the interactivity on the Internet because his level of knowledge is higher. As there are vast amounts of information on the Internet, making it hard to locate and evaluate specific information, it is hypothesized that consumers who have knowledge about the product class will use the Internet more for information search than persons who are not knowledgeable about the product class. Hypothesis 1 therefore states:

1. Consumers who possess a high level of knowledge of the product class will use the Internet more when searching for information in relation to an upcoming purchase. The same is the case for consumers having much interest in the product class.

According to theory on information cost, and as was discussed in chapter 2.3 in relation to the Internet, search for information takes place using the sources from which the consumer perceives the biggest relative benefit in terms of gained knowledge and entertainment compared to the relative cost of using the source in terms of financial, psychological, and time resources. As was discussed in chapter 2.5.2 these sources are not perceived as equally important by all consumers, and therefore, using the Internet is not equally advantageous to all consumers. Hypotheses 2 therefore states:

2. The level of use of the Internet is, for information search purposes, determined by perceived low cost and high benefit of using this source of information.

Hedonic value of the product is expected to augment external search because of a rise in expected benefit from the search through reduction of emotional risk and directly by an interesting experience in finding the information (chapter 2.2.2). The latter is also expected to be the case for use of the source. Also, consumers who use the Internet much, are more skilled in doing so and, therefore, use less effort to find the needed information. This is the topic of hypothesis 3:

3. Consumers, to whom the Internet is important in his everyday life, use the Internet more for information search purposes and perceive lower costs of using the Internet.

Consumers use sources that are credible; that is, trustworthy, objective, and knowledgeable, as was discussed in chapters 2.4 and 2.5.1. Therefore, consumer will use information on the Internet if he or she believes the information is there and if it is perceived as trustworthy. This is expected to be related to the consumer finding the Internet important in his or her life.

4. Consumers, for whom the Internet plays an important part in their life, tend to believe that the needed information is on the Internet, and that the information is credible.

A side-effect of the advantage of using familiar sources is that the consumer experiences the least cost, incurred from search, when using sources that he is familiar with. Therefore, hypothesis 5 (not a part of the model), based on the theory discussed in chapter 2.4.5, states the following:

5. When searching for decision relevant information on the Internet, consumers tend to use the sources on the Internet, they ordinarily use.

That the Internet already occupies an important place in the person's life is one kind of facilitator as Rogers (1966) describes them, and as was discussed in chapter 2.5.4. In addition to knowing how to find the information on the Internet, the consumer, who feels comfortable with the kind of shopping behaviour implied in using the Internet, feels the change in information search behaviour less than consumers, not familiar with searching for information from their residence. Also, to some consumers, the advantage of being able to shop from home is larger than it is to others. Some consumers even like changes in their shopping behaviour. Therefore, hypothesis 6, based on chapter 2.5.4, states:

6. The adaptation of the Internet is easier for consumers who like to shop from home, who are innovative in shopping, and/or feel role overload.

Different consumers experience different costs when using the same source. Sources, in which the consumer finds enjoyment while using them, are less costly to use for the consumer. An important part of shopping enjoyment, which was discussed in chapter 2.5.2, is to be with co-shoppers. The Internet is not able to provide shopping enjoyment as well as, or at least not in the same way as 'old-fashioned' shopping, and, therefore, the Internet is hypothesized to be used lesser for information

search the more the consumer likes to shop with others. Hypothesis 7 therefore states:

7. Enjoyment in shopping is offsetting the benefit and lowering the cost of off-line search for information and therefore discourages the use of the Internet for information search

Internet access is not evenly spread across the population. Younger persons have access to the Internet to a much larger extent than do older persons according to Statistics Denmark (Danmarks Statistik, 2002a). It is also known that the longer a person has had access to the Internet, the more he uses it, and the more it is important to him (chapter 2.3.2). Therefore, it is not possible to rule out age and time the consumer has used the Internet as related to importance of the Internet to the consumer. If these factors are related, it means that the result of the model will change automatically as a larger and larger part of the population has had access to the Internet for a long time and as the older persons 'die out' and are replaced by a generation more involved in the Internet from the beginning. These relations were discussed in chapter 2.3 and 2.5.4 and are captured by hypotheses 8 and 9:

8. Perceived importance of the Internet is related to the length in time, the consumer has used the Internet.

9. Perceived importance of the Internet is related to the person's age. The younger the person is, the more the Internet is important in his life.

Likewise, especially the attitude towards shopping may not be the same across genders, as was touched upon in chapter 2.5.2. Therefore, it is necessary to rule out that the model is not the same for women and men.

10. The relations in the model are not dependent on gender.

The above 10 propositions explain the consumer's use of the Internet for information search and the use of the different sources on a general level. However, the Internet Search Model does not really explain any-

thing about the choice among the different types of sources available to the consumer on the Internet, if the information gathered from these sources has any impact on the consumer, or whether these sources have replaced or substituted the off-line sources formerly used. As the resources for this thesis are too limited to investigate the use of all types of sources, it is necessary to make a choice of which source(s) to look into.

In order to pinpoint which possibilities are more interesting, it is relevant to return to the description of the various types of sources on the Internet in chapter 2.4. In this chapter, it was clear that the option for using the Internet which will produce the largest shift in consumer behaviour is the use of individualized, non-marketer dominated information sources. This is because this option requires a shift from the real face-to-face contact to virtual contact in a group, of which the consumer does not physically know the members – the virtual community. The question then is if the loss in personal contact is outweighed by the advantages of the virtual community in terms of accessibility to the other members from home and at all times and the possibility for the community to enclose a sufficient amount of people sharing a focused interest. If (a segment of) consumers rely heavily on information gathered in such virtual communities when making other-based decisions or decisions in general based more or less on word-of-mouth, the implication is that marketers have to consider these communities in their marketing effort both as pro- and reaction.

As stated in chapter 2.4 on the possibilities on the Internet, the other (new) possibilities on the Internet are far closer to the 'old' ways of acquiring information, as it is primarily the accessibility of the information provided which is new. The information provider is still the same, it is just his retail outlet or the medium which has moved, or multiplied, to a new location in cyberspace.

Among the available sources on the Internet, virtual communities thus seem to be the most interesting from a theoretical point of view, as this is the information source most different from the traditional sources. It is, therefore, interesting to discover whether the use of virtual communi-

ties is influential on individuals participating in them. Some of the communities are difficult to access for outsiders, some types of communities are not discussing topics that are related to the buying process, and other types of communities are not expected to be regarded as delivering objective information. As stated in chapter 2.4.1 and 2.4.3, the type of community that I find to be most relevant for the current study is newsgroups, as these groups are fairly easy to find and access for ordinary Internet users, quite a large part of the groups focus on discussions relating to product choice, and the requirement to enlist gives a feeling of membership (belonging). Therefore, the following hypotheses are related only to the use of newsgroups and not to the use of virtual communities as such. As newsgroups are not representative of all virtual communities, the conclusions in this thesis will, of course, only by applicable to newsgroups, and not to all kinds of virtual communities on the Internet.

As discussed in chapter 2.4.1, it is the expectation that consumers generally will use newsgroups more than chat-rooms for information search purposes because the discussion taking place in newsgroups, in general are more serious and also, the groups are more centred on consumer topics. Therefore, the first hypothesis on the consumer's information search is:

11. Consumers use consumption oriented newsgroups more than chat-rooms for information search purposes

The main reason for the interest in newsgroups is to find out whether newsgroups have replaced or supplemented formerly used marketer- and non-marketer dominated sources. Before being able to answer this question, it is necessary to know more about different types of newsgroups, the different types of members visiting the groups and the relation between the two. This is important because theory presented in chapter 2.5.1 and 2.5.3 tells us that there is a relation between the tie, a person feels to a source, and the extent, to which the source is used, at least when the information needed is not straightforward to give. In addition to this it is interesting to find out if Danish newsgroups can be classified

as ‘real’ communities in terms of ‘sense of belonging’ as was discussed in chapter 2.4.1.

In chapter 2.5.1 it was shown that consumers rely on strong-tie sources when they seek for advice which has to be given on a subjective basis, either because the product has to be chosen based on subjective cues or because they find the decision so hard to make that they choose to rely on another person to choose for them. Therefore, in order to explain differences in the use of newsgroups, it is important to find out if there is a connection between the tie to the group and the use of the group. As the tie is not expected to be equally important concerning different kinds of products, it is consequently relevant to find out if newsgroups centring on different types of products exhibit different usage patterns in general, and if there is a connection between this and the search for information in the newsgroup in question. Hypothesis 12 accordingly states that:

12. Newsgroups centred on products which have higher self-centrality to the consumer are expected to be tighter knit than newsgroups not centring on products which are personally important to the individual.

In newsgroups, communication takes place as postings to the group. Therefore, users of newsgroups have to be motivated to actively enter the group. Kozinets (1999) have stated, as was described in chapter 2.4.1, that the motivation to enter consumption-oriented newsgroups can be described by a 4-factor model. The factors are: Entertainment, Social Relations, Giving Advice and Getting Advice. The model has, to my knowledge, not been tested empirically. Therefore, it necessary to test the validity of the model before doing tests of possible relations between motivation and other concepts. Hypothesis 13 thus states the following:

13. The motivation to enter consumption-oriented newsgroups can be described by a 4-factor model. The factors are: Entertainment, Social Relations, Giving Advice and Getting Advice.

The purpose of the investigation of motives for entering newsgroups and the different composition of the newsgroups is to be able to conclude on

whether consumers, who have different ties to a newsgroup and /or to the product class, use the group for different kinds of information gathering. In the discussion on the different types of consumers entering the newsgroups, Kozinet's framework was put forward in chapter 2.4.1 as a way to segment the different type of users of newsgroups in relation to communication regarding consumption issues. It was also put forward that different segments in newsgroups may have different motivations for entering the newsgroup, and also different roles in the communication, centring on giving and receiving consumer advice that is going on in the newsgroups. Hypothesis 14 consequently states:

14. The motives for entering the group are related to the person's relation to the group and interest in the product and, therefore, individuals who belong to different segments, as defined by Kozinets (1999), also have different motives for entering the groups.

In order for newsgroups to be used for exchange of consumer information, they have to be populated by consumers, who are motivated by search for information, and (other) consumers who are capable and willing to give this information. To find out if this is the case, it is necessary to clarify the motivations of the members of the group, and the connection between motivation, and asking for and giving advice in the group. Motivation to give or seek for information is related to personality, as was also discussed in chapter 2.4.1. In the context of this thesis it is, therefore interesting to investigate whether the knowledgeable members of the newsgroups are motivated to give the needed information to the persons, seeking this information – in other words, if there are Opinion Leaders in newsgroups. Likewise, it is interesting to reveal if there are opinion seekers in newsgroups. The right combination of opinion seekers and opinion leaders in newsgroups, each motivated by their separate need to acquire and give information, is bound to motivate both types of consumers to use newsgroups for exchange of consumer information, thereby probably making the use of newsgroups for search for consumer information in the future more common than it is today. Hypothesis 15 states the following on the composition of the newsgroups:

15. The level of communication in newsgroups can be explained by consumers having different motives to visit the group which fit each other well. The motives are governed by different types of personality in terms of seeking and/or wanting to give advice.

After this look into newsgroups, the types of members they have, and consumers' different motives for entering newsgroups, we are hopefully ready to try and answer the main question in this relation: Have newsgroups replaced off-line sources formerly used for information search? According to the discussion of determinants of the use of different types of sources mainly in chapters 2.4.1 and 2.5.1, and the investigation into the membership segments in newsgroups, and their motivations for using newsgroups, the final hypotheses state as follows:

16. Use of newsgroups primarily replaces marketer-dominated sources. This is more prevalent for Tourists than for the other segments.

17. Use of newsgroups has also replaced use of advice from primary reference groups. This replacement is larger for Insiders than for the other segments and larger in newsgroups on products with high self-centrality to the consumer than in other newsgroups.

The discussion in this chapter made the picture of the antecedents of information search in terms of both kind and amount of search, clearer. The discussion further made it clear that it is imperative to consider that search for information is not always regarded as a job by the consumer but can also be seen as entertainment. Finally, there is a strong possibility of socioeconomic and psychological differences in the motivation for and actual use of the Internet for search for product-related information.

As the variables and the hypothesized relations between them have been found by studying general literature on search behaviour and specific literature on consumer behaviour on the Internet, there is a risk that the relations found are quite different from what is expected from the theoretical study.

Another reason why this study may present different results than the studies used to develop the model is that these studies were based on much skewed samples based on self-selection among international – but mostly American – users on the Internet in 1998 as the latest. This means that a study performed in Denmark in 2002 can produce different results because the population investigated is different from the population investigated in the published studies.

Therefore, it is – if not exiting – then interesting to see if the hypotheses stated in the above can stand for an empirical test. The design of this test is the topic of the next chapter. The hypotheses stated in this chapter operate on two different levels. The first 10 hypotheses are related to the Internet Search Model, which is supposed to capture determinants of use of the Internet for information search on the general level. The next 7 hypotheses are related to use of a special kind of information source on the Internet: The newsgroup. Such different levels of investigation put different demands on the empirical research designed to test the hypotheses. Therefore, the solution to this problem is an important decision to make in the following chapter on research design.

3. Research Design

At the end of the former chapter, several hypotheses were stated regarding the consumer's use of the Internet for information search in relation to an upcoming purchase. The hypotheses were related to two different levels of detail: A general model explaining the determinants of the extent, to which the consumer uses the Internet for information search (the Internet Search Model) and hypotheses regarding the search for information taking place in newsgroups on the Internet which are related to consumption and purchase of products.

The next step is to test the hypotheses through empirical research. The research paradigm is neopositivism. Therefore, the demands of good research are the ones applied to this kind of research: Objectivity, validity and reliability (Guba, 1990). The research design is developed bearing these demands in mind. The following chapters discuss necessary choices in research design: How to measure the concepts in the model and the hypotheses, how to gather the relevant data, and how to select individuals to participate in the empirical research. As the measurement issue is closely related to the desired way of analysing the data, a preliminary plan for data analysis is described prior to the chapters covering measurement, questionnaire design, and sampling.

Before undertaking the confirmative research, for which the plan is designed in the following, I made observations of life on the Internet to try and grasp a preliminary understanding of this life. I have made numerous exploratory visits to websites, both commercial and non-commercial, and also to chatrooms and newsgroups. These visits have, together with theory on research design, laid the ground for determining how to design the quantitative study aimed at testing the Internet Search Model and the hypotheses regarding communication related to consumer search for information in the virtual communities. I have chosen not to supplement these observations with qualitative interviews before proceeding to the quantitative study because the hypotheses stated are based on extensive literature studies which also include empirical research. This research has been partial and mostly based on samples which were collected by self-selection of the respondents. It is, there-

fore, my opinion that the need is an empirical study of the whole Internet Search Model based on a sample which is representative for Internet users in general, and that further qualitative research would probably not bring in any new elements into this thesis. With respect to the hypotheses on search in the virtual communities, qualitative research has been done in the form of unstructured and structured observations of the communication in the newsgroups.

As the hypotheses regarding consumer search relate to two different levels of detail, there is, in my opinion, a need to perform two separate empirical studies: First, and most importantly, a general study in order to test the Internet Search Model, and to investigate the importance of newsgroups on the general level. The second, empirical study will look into search for information going on in newsgroups to test the hypotheses about use of newsgroups for search for consumer information.

There are three reasons why it is preferable to make two separate studies rather than one study: the first reason is linked to the sampling issue, the second reason is related to the questionnaire issue, and the third reason is related to a special aspect of the newsgroup study.

The first reason for making two, more or less unrelated studies is that it is important to get a general picture of the antecedents of use of the possible information sources on the Internet in order to find out where the 'average' consumer finds his information. Therefore, the main study has to be able to grasp the behaviour of this average consumer. Such a picture cannot be made based on a sample drawn from individuals who participate in newsgroups, as only a smaller percentage (according to NetValue around 8%) uses newsgroups and the individuals who do this are probably not average Internet users, but more skilled Internet users. Therefore, a sample taken among members of virtual communities is not representative for the average Internet user. On the other hand, a sample drawn from the general population probably will contain too few newsgroup users to be able to do any analysis on this group as a separate entity unless the total sample is extremely large. Therefore, it seems appropriate to draw a separate sample among newsgroup users in order to challenge the hypotheses regarding influence on consumer behaviour

from this group, and another sample drawn from the population of Internet users in order to test the Internet Search Model.

The second reason for making two separate studies is related to the questionnaire issue. It seems inefficient to include questions aimed at testing hypotheses on search for information in the newsgroups in all questionnaires, because the questionnaire will get unnecessarily long and the content inconsistent if a single questionnaire tries to cover both issues, as many of the questions will be irrelevant to, by far, the largest part of the sample.

A third but related reason for making two separate studies is that in order to look into exchange of consumer information, it is interesting to gather data on who influences who. In order to gather such network data, it is necessary to work within a group and draw the sample from this group. The consequence of this is that the importance of the possibility of saying much about a few groups is higher than the importance assigned to the sample being representative. This is in contrast to the goal of the empirical study undertaken to test the Internet Search Model to sample the population of Internet users because representativeness in this study is very important.

In the newsgroup study, the focus is on determining if the hypotheses concerning segments, roles and influence in the newsgroup hold. Therefore, the quantitative study in the newsgroups is preceded by observations in a number of newsgroups, at first unstructured and later structured in order to be able to select newsgroups which are relevant for sampling.

The type of study chosen is in both cases a survey. This method of collecting data is chosen because of the ability to study larger and, therefore, usually more representative groups, than one possibly is able to achieve using qualitative data. Furthermore, using a questionnaire makes it possible to do a structured analysis. Also, quantitative studies have the advantage of making comparison with other, related studies easier.

To study communication in newsgroups, an alternative to the survey method is to use the qualitative methods put forward by Kozinets as 'Netnography' (Kozinets, 2002). These methods include mostly lurking in newsgroups combined with more or less structured analysis of the textual information posted in the newsgroups. This approach has the advantage of being very flexible and open-minded. There is, however several problems in using this approach: It is difficult to aggregate the information to draw a fuller picture of the society and the approach is too loose to be able to follow-up with other, related studies in other communities to see if the patterns are the same there. Also, as with all qualitative research it is hard, if not impossible, to validate the results found; a problem that is accentuated if the researcher lacks skill in doing ethnographic studies as is the case here. Finally, the purpose of the research in connection to this thesis is to study why and if consumers use newsgroups in their search for information and not their actual communication, while doing so.

Therefore, the choice has been to do a study that is quantitative by nature. As a consequence of this, the survey instrument will be a questionnaire with mostly closed questions combined with observations in the chosen newsgroups. This way, it is possible to analyse the material using statistical methods. Also, the observations done in the newsgroups will be mainly quantitative focusing on numbers of meetings and postings as in structural network analysis as described in Borgatti (2001) and Wellman (2001). In short, structural network analysis studies networks using quantitative measures to characterize the network in question.

3.1 Preliminary Plan for Data Analysis

The aim of the empirical study is to produce data that can be analyzed in order to test the proposed hypotheses. In order to perform the necessary analyses, it is often a prerequisite that the data possess certain characteristics. As a consequence of this it is important to consider the foreseen data analyses during the development of the research design. Therefore, this chapter lays out the preliminary plan for data analysis.

On the general level, the procedure followed is the procedure described in most common textbooks on marketing research, *e.g.* Hair (1998) or Malhotra (1999).

According to this procedure, the first step after data collection is to validate and clean the data. This will also be the first step in the current study. Therefore, the data are first inspected for possible problem areas and for missing data. For observations missing only a few answers, the missing answer is imputed as the average answer of the individuals who have answered to the question. This method is a rather simple way of dealing with the problem, and can cause problems if used on more than a few cases (Byrne, 2001; Hair, 1998; Malhotra, 1999). If there are too many missing items the observation is, therefore, deleted. The exact definition of 'many' is postponed to the treatment of the relevant sample.

The next step is to evaluate the data set in terms of external validity. Is the data set representative of the population we are trying to measure? Tests for representativity of the general sample is done using simple goodness-of-fit tests on demographic variables comparing the data set to the population of Internet users in Denmark, whereas the newsgroup sample, as far as possible, is compared to the relevant newsgroup. How this is done is discussed in further detail in the chapter on analysis of the newsgroup data. If the data set is not representative, it is necessary to discuss the implications and possibilities of off-setting this.

After the evaluation of the data set, the next step is initial description and analysis of the data to acquire an initial understanding of the contents and patterns in the data – that is: differences between groups, distribution on variables etc.

As a part of, or in connection with, these analyses, tests of the hypotheses which do not require analysis of structural equations are done. The analyses done in this part of the thesis are carried out using common statistical techniques as χ^2 -analysis, regression, and analysis of variance depending on the type of data involved in the analysis. Therefore, these analyses per se do not require specific scales. However, in order to get

the maximum amount of information, the highest possible level of scaling is always used in the questionnaire – of course with regard for the consumer's possibility of and willingness to answer the question as precisely as demanded by the response format and what is relevant in terms of detail. For example, a person's exact age is not relevant. To measure the age of a person in years or maybe 5-year intervals is mostly a sufficient level of detail for adults.

A special issue is test of the Internet Search Model and the measurement models used in the newsgroup study. As the Internet Search Model is a model well suited for multivariate analysis, it is the intention to submit the model to a test using Structural Equation Modelling.

An important choice to make in Structural Equation Modelling is the choice of which estimation method to use as there are several alternatives, depending on the characteristics of the input data. Byrne (2001, p. 71) states that estimation by the Maximum Likelihood method is the best alternative, provided that the scale has at least 5 categories, and the variables are not skewed in different directions. This implies that the variables measuring the unobservable constructs in the models should be measured on such a scale whenever possible. Also, the model has to be identified. A rule of thumb used to ensure that a model is identified is to measure each latent variable by 3 to 5 observed variables, whenever possible (Bagozzi and Yi, 1988). This is also important to consider when choosing measurement scales for the empirical study.

After this short description of the preliminary plan for data analysis, we are ready for the next step in research design: Development of a measurement model which meets the criteria put forward in the above.

3.2 Issues in the Development of a Measurement Model

In order to be able to test the Internet Search Model, it is necessary to find measures for the concepts in the model meeting the criteria stated in the above and also, in reality, measuring the concept in question and nothing else. This chapter treats this issue.

An often-cited reference on how to develop scales is Churchill (1979). In his article, Churchill recommends a multi-step model for development of measurement scales. The process is very resource demanding as it involves several tests of the scales on samples of considerable sizes. This is probably the reason why, although many researchers cite Churchill, only a few go through the steps in the model. Either the scales are not tested at all, or they are tested on samples which are much smaller than Churchill recommends (Flynn and Pearcy, 2001). This shows that considering alternatives to development of scales for this study from scratch is probably preferable as the development of scales from scratch may be too resource demanding if this is not the core research issue.

A viable alternative to self-development of scales is to use scales developed and thoroughly tested by others. To ensure that the scales used are of an adequate quality, most of them have been found in the compilation 'Marketing Scales Handbook' edited by Bruner and Hensel (1992). In this compilation of scales, the scales are all evaluated based on scientific standards as described by for example Peters (1979, 1981, and 1986). The newest edition of 'Marketing Scales Handbook' available was the 1992 edition. To ensure that the scales are not outdated by more recently developed scales a search was made in Proquest (www.proquest.com) to ensure that the scales are still widely used in studies published in peer reviewed journals. To validate the selection of scales in Bruner and Hensel, a crosscheck was made in the much smaller compilation of scales, 'Handbook of Marketing Scales' (Bearden, Netemeyer and Mobley, 1993). This cross-validation gave no reason to believe that the 'Marketing Scales Handbook' does not cover the field adequately.

Due to the time of publishing, the Marketing Scales Handbook does not contain scales measuring Internet usage. Scales measuring this was therefore inspired by the work of Hoffman and Novak and their associates as they, in their effort to model the flow construct, have developed several scales designed for measuring constructs related to usage of the Internet. Their work on development and testing of the scales has been described in several working papers and often-cited articles (*e.g.* Novak,

Hoffman and Yung, 2000). Therefore, these scales are deemed suitable for use in this study.

As this text is written in English and the questions are posed to Danes and, accordingly, in Danish, translation of the questions is a special issue in the development of measures for the elements in the models. I have chosen to do the operationalization in English, and afterwards translate this only in the questionnaire. The two main reasons for this are that the scales developed by others on the subject are in English and that this text is written in English. One can argue that there is no absolute certainty that questions translated to Danish are perceived exactly the same as they are in English – especially when translated by an amateur - but there is not any bullet-proof way to overcome this problem. If the whole thing is done in Danish there will still be a considerable amount of translation of scales, developed in English, and tested in their English version on American or English respondents. Therefore, the best option is to stay in English except when it is necessary to translate into Danish, which is the case only when contacting the Danish respondents.

The scales were, after translation into Danish, all tested in several pilot studies using students and colleagues as respondents. The samples, taken this way, were too small to meaningfully calculate Cronbach's Alpha. Therefore, I had to rely on that the Alphas reproduced in Bruner et al. (1992) and Novak, Hoffman and Yung (2000), would be valid, also for the final sample in this study.

3.3 Variables Related to Internet Use on the General Level

This chapter discusses operationalization of the variables in the Internet Search Model. The central question in the empirical test of the model is if we are able to determine a relation between how much the consumer uses possibilities for information search on the Internet and the explaining variables in the Internet Search Model which was presented in figure 2.9. In order to carry out such an empirical test, it is necessary to find out how to measure the theoretical constructs in the model.

The chapter starts with a discussion of how to ensure that the consumer thinks of the right kind of problem while answering the questions. After this, measurement of the variables in the model is discussed. The discussion of measurements starts with the concept 'Use of the Internet as an Information Source', and then works its way backwards through the model. The chapter closes with a consideration of which measures, not directly related to constructs in the model, it is relevant to add to the questionnaire.

3.3.1 Setting the Context for the Respondent

It is important to ensure that all respondents think of the same kind of buying process while answering questions. This is not automatically the case, as the search process is different for different kinds of products. Furthermore, it is possible to deal with past behaviour as well as intended behaviour.

Introduction to the topic of the survey and to the questionnaire is, of course, a tool used for setting the context for the respondents. However, as this project deals only with extensive problem solving, the consumer has to think of a situation, in which such a process has or will take place, while answering questions. Therefore, an introduction describing the research project is not enough to ensure that all consumers think of the same kind of process while answering the questionnaire.

One way of ensuring that all consumers participating in the survey think of the same kind of buying process is to use a screening process, selecting only consumers who have bought a certain product (for example a car or a dishwasher) during a certain time span in the past. The problem in using such an approach is twofold. The first issue is one of getting a sample of an adequate size. This is because only a limited number of people buy a certain product during the time span in the past, of which the consumer can be expected to have sufficient recollection to be able to answer correctly to questions about the process that took place. Therefore, only a small fraction of the original sample would pass such screening criteria. The second problem is one of generalization, as the possibility to generalize from the study is dependent, not only on the sample being representative for the population, but also on the buying

process studied being representative for extensive problem solving. Therefore, doing a study considering only one kind of product seems to be too limited. On the other hand making the consumer think of a specific buying process helps recollection. In addition to this, when interpreting results of the study, it makes the results more valid if we know from which products we infer results to the buying process in general. The conclusion to this short discussion is that the adequate number of products to put into investigation must be limited but larger than one and also that products chosen have to be products, for which a sufficient number of consumers can be expected to have been through a buying process during a shorter time span, for example the past year.

The practical solution to the problem discussed in the above is to ask respondents to state what products, from a range of products, he or she has purchased during a period in the past and then, from the products selected, to think of the process leading to purchase of one of these while filling in the questionnaire. Which product the consumer has bought has to be indicated in the questionnaire. This way, it is possible to do both a general analysis of the process as such and partial analyses concerning certain products – provided, of course that the total sample is of a sufficient size and the partial samples are also of sufficient sizes. The span of products was selected based on results in Jensen (1990), observations on the Internet, and an attempt to cover products which are different from each other in terms of where and how consumers buy them, and how 'natural' it would be to search the Internet for information on the product. The selection was tested along with pilot testing of the questionnaire. The final selection of products can be seen in the questionnaire printed in appendix A.1.

The second issue pointed to in the beginning of this chapter is the question of actual (past) behaviour and intended (future) behaviour. The reason why we are facing this problem is that it is not possible to observe the actual process, as it is hard to find individuals who are running through the process at a certain point in time and after this to follow these individuals through the whole decision process. In addition, observation of actual behaviour suffers from the problems put forward in chapter 1.1 on Radical Behaviourism.

Because of the problems stated in the above, it is customary to settle for the respondent's recollection of past behaviour and/or an indication of intended future behaviour even though also these approaches pose problems which have to be overcome. In the present case, this seems to be the sensible solution, too. As intended behaviour often is subject to wishful thinking, I choose to measure reported past behaviour in this study. Regarding past behaviour, a general problem is that the consumer cannot remember what actually happened or remembers what he would have liked to have done, rather than what he actually did. To overcome these problems, I have asked for events which have taken place during the past year. It will also be of help to the consumer's recollection of events that the questions are related to a buying process which presumably activated the consumer's cognitive capacity. This should make remembrance of the incident easier than for habitual behaviour. Finally, I do not think subjects in this study are prone to the respondents wishing for social acceptance, even though there is a risk that use of the Internet may be over-reported, either because it makes the respondent look 'smart' (socially desirable), or because it is better remembered because it is a deviance from formerly ordinary behaviour.

3.3.2 Use of the Internet as an Information Source

Use of the Internet as an information source can be split into two parts: Use of the Internet as an information source in general, and a sum of use of the different types of information sources on the Internet. If the consumer is consistent in his answers, correlation between the two ways of measuring use of the Internet for information search should be high.

Use of the Internet as an information source can be looked at in two different ways: an absolute and a relative way.

The absolute amount of use of the Internet for information search can be measured as time spent using the Internet for search purposes in relation to an upcoming purchase. However, this measure cannot alone measure importance of the information sources available on the Internet. This is because even a consumer who spends many hours visiting companies' websites may still only do this as a supplement to other sources, and

may not even consider information found on the Internet to be important concerning the actual choice to be made by the consumer. Furthermore, for most consumers, it is probably difficult to estimate how much time they spent using each source of information, especially if they experienced 'flow' as Hoffman and Novak describes it in their work (1997, 1999).

A relative measure of how much the Internet was used for information search in relation to the selected choice process will measure how much the consumer used the Internet compared to off-line sources. This is relevant information, as the topic of interest here is mostly to explain use of sources on the Internet instead of, or in addition to, off-line sources. In favour of measuring relative use of the Internet for search for information speaks, in addition, that to the consumer it is probably easier to estimate how much he or she used sources online in total compared to off-line sources than to estimate amount of time he spent searching on the Internet.

The next problem is to find a scale, on which to measure relative use of the Internet for information search. The possibility finally chosen was to let the consumer state on a 5-point scale whether he used the source much or little compared to other sources. This way of measuring is, of course, not very accurate, as much is not the same to two different consumers. However, after thorough consideration and testing of alternative measures, this seems to be the best way, after all. One of the alternatives tried was to use a constant-sum scale. However, this was too complicated, even for students taking a course in marketing research.

The second way to measure use of the Internet for decision relevant information is, as mentioned at the beginning of this chapter, a measure summing use of each of the possible sources on the Internet. It was decided to measure absolute use of each source on a 5-point scale. The measure used in structural equation modelling is then the sum of use of each possibility standardized to the same mean as the general measure. This is because it is easier to interpret results when all measures use the same scale.

An additional reason for collecting information on use of different sources on the Internet in relation to an upcoming purchase is the possibility to get a general picture of which sources are used for information search, as well as a wish to test whether the consumer uses the same type of sources when he seeks for information for evaluation of alternatives, as he does in general.

As a supplement to measuring relative use of the sources on the Internet, it is relevant to ask the consumer where he actually found information, and from which sources he found useful information as this will probably influence future behaviour. This is because finding useful information is rewarding and, therefore, supports this behaviour whereas spending time and not finding anything works the opposite way around. Amount and usefulness of information were measured in the same way as use of each source, using a 5-point scale.

3.3.3 Knowledge about and Interest in the Product

During development of the model it was shown that the search process is influenced by product interest and product knowledge. It is expected that amount of search in general is positively linked to both product interest and product knowledge. In the Internet Search Model, knowledge about the product class as well as interest in the product class is expected to lead to more extensive use of the Internet for information search because the consumer knows more about what to search for and finds the opportunity to search how and when he prefers more rewarding.

Product knowledge can be measured as indicated by Bloch, Ridgeway and Sherrell (1989). They used a 2-item 5-point Likert-type scale (Alpha=0.86) to measure knowledge of a product compared to others in general, and compared to their friends. The items are displayed in table 3.1 at the end of this chapter. As product knowledge, in the model in figure 2.3, product interest is also deemed important in level of use of the Internet for search purposes. There was no evident scale for measurement of product interest, and therefore, I chose to measure this construct using the same scale as the one used for measuring product knowledge.

Importance of the purchase decision is not directly represented in the model, but may be relevant in use of the Internet for information search as decisions which are important require more credible information sources. This may be off-setting to the influence from a high level of product interest and/or knowledge. Importance of the decision can be measured by a 3-item scale developed by Ratchford (1987). The reported Alpha-level for this scale was 0.77. The scale was a semantic differential scale. It was tested for both criterion and discriminant validity and use of the scale was much the same as in the current study.

Total amount of search poses a problem of its own, as this is expected to be very hard to estimate for consumers. First, the search process is composed of many little searches and, therefore, total time spent can be very difficult to estimate for the consumer. Second, the relevant time span may be different for different products in the study. On the other hand – asking in terms of 'much' or 'little' may be too inaccurate. Therefore, after all, I decided to ask the question using exact time spans in hours. As total time for search for information includes transportation the question concerned time used including time for transportation. As one trip to a shopping area including visiting stores easily takes around 3 hours, I decided to let the lowest time bracket be less than 10 hours.

3.3.4 Perceived Cost of Search Using the Internet

As discussed earlier, classical economic decision theory states that the consumer will use certain information sources if the benefit he expects from using the sources is larger than the cost he expects is implied in using the source. Therefore, the consumer is expected to use the Internet for search for information if he perceives that the cost of doing so is lower than the cost incurred by traditional sources, or if he expects to incur a larger benefit at the same cost.

There are three elements in perceived cost of using the source: Time, money and effort. That money is a part of the cost element is self-evident. However, money is not the only scarce resource relevant to consider in this case. Time is also a scarce resource and, therefore, a cost that has to be accounted for. While searching for information, time

is spent locating the information source, accessing the information source (that is, going to the retailer or accessing a website), and acquiring and comprehending the information retrieved. This time could have been spent doing other things. Likewise, psychological and physiological effort is a scarce resource which has to be distributed among various tasks. If one task occupies the brain, there is less space for other tasks.

Direct questions regarding lower perceived cost in terms of money, time, and psychological effort of gathering information using sources on the Internet measure the cost elements. Again, measurements can be relative – compared on a general level to other ways of gathering information – or they can be absolute. In accordance with generally accepted economic theory, the hypothesis is that the consumer will use the Internet for information search if this is, in some way, perceived as cheaper than the alternatives. Therefore, a relative measure is used with a high value indicating lower cost of search.

Not all cost elements are equally scarce to all consumers. As discussed in chapter 2.5.2 on differences in experienced cost, for some consumers, time, and not money, is the scarce resource. If this is the case, cost elements may not exercise an equal amount of influence on the decision of which source to use when searching for consumer information. Therefore, weights should be assigned to the cost elements. This weight can be indicated by each consumer. This way, the cost variable used in the analysis can be weighted by importance. This means that costs which are considered large but not important will be perceived as smaller than they actually are, while costs elements which are considered important will be perceived larger than they really are.

The direct questions about perceived cost elements are supplemented by questions regarding overall effort needed and benefit provided from using the Internet for information search.

3.3.5 The Shopping Aspect

In the Internet Search Model, enjoyment in shopping is assumed to be an offsetting factor to using the Internet when searching for information. This is because some people like to go shopping regardless of the cost

incurred. Therefore, the assumption is that consumers, who like shopping, will use the Internet less at the same perceived cost than other consumers.

There are several scales measuring shopping enjoyment. O'Guinn and Faber (1989) developed a 3-item scale for this purpose. The scale achieved an alpha value of 0.89. However, the O'Guinn and Faber scale only measures shopping enjoyment in itself and was developed to study compulsory shopping behaviour. As an important difference between shopping off-line and on-line is the social aspect, this measure needs to be supplemented by items capturing enjoyment connected to being with others physically, while shopping.

It is important to consider that the 'others' involved can be both 'co-shoppers' and retail staff. Several studies related to research done by Hoffman and Novak (Hoffman, Novak and Schlosser, 2000; Wolfinbarger and Gilly, 2001) showed that part of the enjoyment in using possibilities for information search provided on the Internet is freedom from physical limitations, and lack of influence from and contact with salespeople. However, there are also consumers who prefer contact to, and advice given by sales personnel. The 'Personalized Shopper' scale developed by Lumpkin (1985) takes the importance placed on contact with retail staff into account. The reported alpha-value for this scale was 0.84, and the scale was validated using factor analysis.

The scale was developed to measure interest in shopping in stores, in which the consumer is known. This is not quite the issue for this study. However, the questions cover interest in having personal contact with sales staff and, therefore, the scale is deemed usable in this study, too. Lumbkin and Hawes (1984) also developed a scale for shopping involvement (referred to as 'Scale for Shopping Enjoyment'). This scale covers, in 9 items, both the interest in shopping per se, and the social aspects involved in shopping. The reported alpha-value of the scale was 0.83. The scale was validated using factor analysis on a sample of people aged 65 and above. As the scale covers both the social and entertainment aspect, it seems to be a better scale than the O'Guinn and Faber scale for the purpose of this study. It is not considered a major prob-

lem that the scale was tested on a sample different from the sample to be used in this study.

It was, however, the object of some concern that the scale consists of 9 items, which is quite a large number, and that Bruner et al. (1992) point to the fact that two of the items relate to innovativeness in shopping. As innovativeness in shopping is seen as a facilitating factor in the Internet Search Model, it seems a viable solution to move these two questions to the facilitating factors and thereby reduce the total number of items in the questionnaire. The conclusion, therefore, is that Lumbkins' two scales are usable to measure shopping enjoyment and enjoyment in contact with sales staff, respectively, with the suggested modification in the items relating to innovativeness in shopping.

3.3.6 Expected Usability of Information on the Internet

Perceived cost of finding useful information on the Internet is assumed to be dependent on belief in that relevant information is there (that the source is "knowing"), and that the information is credible (trustworthy). This, in turn, is assumed to be dependent on importance of the Internet in general.

To measure expectation of needed information being available on the Internet, it is necessary to first establish what information the consumer is in need of. This is information relevant for decision-making. Based on the theoretical considerations in chapter 2.3, it seems safe to assume that the consumer seeks information on product features, prices, and, depending on the consumer's personality, knowledge about, and interest in the product, also comparison of brands (objective features) and/or advice (subjective). Therefore, the scale measuring availability of information must include items covering these areas. I found no previously developed scale measuring this concept. Therefore, 4 items were developed covering the areas listed above.

Credibility of the source can be measured using a 5-item, 9-point scale, developed by Lichtenstein and Bearden (1989). The scale has a reported alpha value of 0.78. The items are displayed at the end of this chapter.

Due to translation problems, two of the items were reduced to one, as there is no difference in the meaning of the words in Danish.

3.3.7 Importance of the Internet

Belief that the Internet contains the information needed and that the information is credible is assumed to be dependent on the place the Internet occupies in the person's life. The place the Internet has in the person's life, or how important the Internet is to the individual, is also assumed to influence use of the Internet for information search directly because it eases adaptation of the behaviour, as described in chapter 2.5.4 on adaptability of the Internet as an information source.

Novak, Hoffman and Yung developed a 5-item scale aimed at measuring importance of the Internet in the individuals' life (Alpha = 0.9) (Novak, Hoffman and Yung, 2000). In their study they found that importance of the Internet is related to flow which, in turn, mediates experimental behaviour (which could also be search behaviour) on the Internet. Measurement of how important the respondent perceives the Internet to be in his life can thus be carried out by asking the respondent directly about this, using the Hoffman and Novak scale.

One thing is how important the consumer perceives the Internet to be in his life; another thing is how much space the Internet actually occupies in the person's life. It has already been stated that amount and frequency in use of the Internet is related to the amount of time the consumer has been using the Internet. Therefore, objective measures of the time the person has used the Internet, and how much time he spends on the Internet might be useful additional measures in evaluation of how important the Internet is in the consumer's life. To measure the time the consumer has been using the Internet, Hoffman and Novak's time brackets were used as they proved to be usable in their research. It is, of course, a question whether this scale is interval scaled, even though the time brackets relate to years. Hoffman and Novak do not comment on the issue when using the variable in the empirical test of their model. Nonetheless, this variable has to be used with caution, especially if it is found to be skewed. The same applies to the scale measuring time on-line per week.

As described in chapters 3.2, Hoffman and Novak, along with Schlosser (2000), investigated if enjoyment in using the Internet is related to personality. They found that individuals with high internal locus of control enjoyed using the Internet more because they felt more in control on the Internet than they did off-line. Therefore, locus of control might play a part in level of importance of the Internet in the person's life in this study, too. To this end there is thus a need for a measurement scale for Locus of Control. Lumbkin and Hunt (1989) used the Rotter (1966) scale. Rotter tested the scale using factor analysis. The alpha value is 0.62 which is on the low side. Hoffman Novak, therefore, used a scale developed by Levenson (1974) in their questionnaire in the GVU-study from 1998. This scale measures several aspects of locus of control separately in, in total, 19 items. This seems to be many items for this study, as the concept is not central in the model. Because of this, and because the items in the Rotter scale are rather straightforward, and Rotter's scale is also widely used in spite of the low alpha value, the Rotter scale was selected for use in the questionnaire.

The final measure relating to importance of the Internet is the consumer's use of the Internet. If the consumer uses the Internet almost exclusively for work purposes, this might offset the impact of the time the consumer has been using the Internet on importance of the Internet in the person's life. Therefore, it is interesting to be able to control for use of the Internet for private or work-related purposes. It was chosen to let the consumer choose between 5 different brackets, ranging from 100% private to 100% work-related use.

The last variable assumed to be related to importance of the Internet is age. The measurement of this variable poses no problems and is therefore not discussed any further.

3.3.8 Facilitating Factors

The last element determining perceived relative cost of using the Internet for search for product information is adaptability of the source, as described in chapter 2.5.4 discussing change in consumption behaviour.

The hypothesis is that consumers who have a positive attitude towards shopping from home will more easily adopt search on the Internet as a possibility because the Internet share many attributes with home shopping. As shopping on the Internet is also a form of home-shopping, this was added to the list of possibilities for home-shopping in the model. Hawes and Lumpkin (1984) developed a 7-item, 6-point Likert scale designated to measure attitude towards shopping from home. The scale had an alpha value of 0.755, and was validated by factor analysis with a satisfactory result. Therefore, this scale was chosen for measurement of proneness to in-home shopping.

Innovativeness in shopping was measured by the two questions relating to this concept as described along with the discussion on measurement of shopping enjoyment.

As stated in chapter 2.5.2, perception of time as a scarce resource is linked to role overload. As searching for information on the Internet can be done from home at the time which best suits the consumer, role overload is expected to act as a facilitator for using the Internet for search for consumer information. Role Overload can be measured by a 13-item scale developed by Rielly (1982). The scale was recently used in a study of women's attitudes toward shopping channels (Maher et al., 1997). In this study, and in the original study, the reported alpha value was quite high. There are, however, two problems in using this scale directly in the study. First, as role overload is not a central concept in this study, 13 items is considered to be too many. Second, the scale was directed to housewives, and some of the questions are a bit odd to pose to men and to women in the work force, in Denmark. Therefore, the scale has to be changed, mainly towards fewer items, if it is to be used to measure role overload in this study. Thorough consideration resulted in a measurement scale keeping only two items of the original 13. To validate this scale, a check of the correlation with socio-economic indicators as weekly workload, number of children living at home, and gender can be done, as role overload should correlate positively with at least workload and number of children living at home. These questions related to socio-economic status are posed anyway for other reasons and, therefore, the

variables needed for validation do not contribute to the total number of questions in the questionnaire.

3.3.9 Background Variables

In addition to the variables which are a part of the model or relating to measures in the model, there are a few additional variables which it is natural to include in the survey. These variables relate mostly to description of the resulting data set and control for representativeness. These variables are:

- Where the respondent lives. This is measured by postal code, as this number can easily be recoded into relevant geographic areas. The variable is used for test of representativeness of the sample.

- Number of persons in the household split into adults, number of children, and age of members of the family. All these variables can be measured directly. These variables are used for test of representativeness of the sample and in test of validity of the scale measuring role overload.

- Workload pr. adult in the household is also assumed to be related to role overload along with number and age of children. It has been decided to ask for workload for each adult if there is more than one adult.

- Gender is relevant for description of the data. In addition to this, hypothesis 10 states that the relations in the Internet Search model are the same for both men and women. To test this, we have to know gender of the respondent. Finally, role overload is expected to differ between men and women because handling of the children is still mainly a female responsibility according to Statistics Denmark (Levevilkår i Danmark, 1997).

- Gross income for the household is to be used for check for representativeness of the data. Gross income is measured in terms of earnings in the household (by adults) per year.

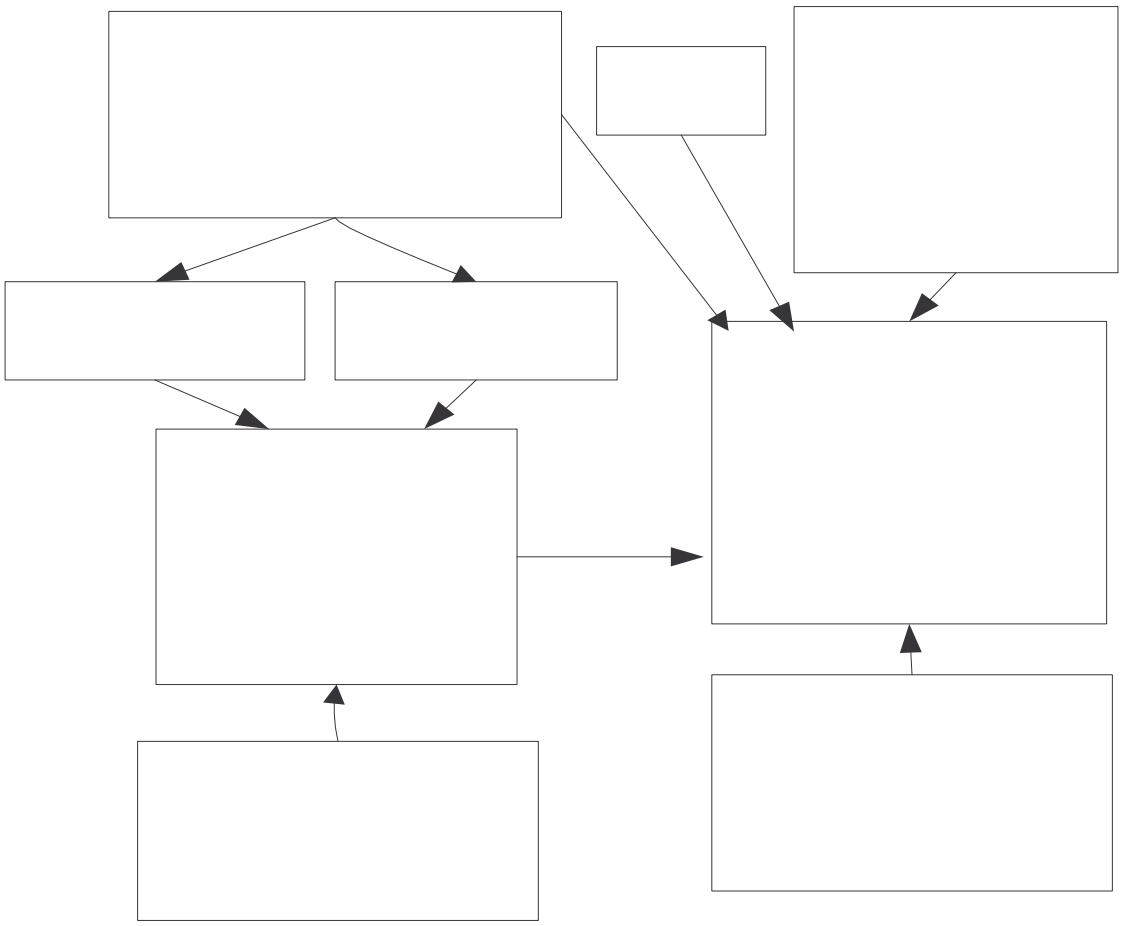
- Level of education is also to be used to test for representativeness of the data set. The common brackets used by Statistics Denmark are used as these are well known, also by the respondents.

The concepts selected to cover the elements in the Internet Search Model and the belonging measurement scales are summated in figure 3.1 and table 3.1, respectively, in the following chapter.

3.3.10 Summing Up Model Operationalization

The discussion on measurement of the unobservable variables has resulted in a precision of the Internet Search Model in terms of observed variables. To make the picture of the model clearer, the model is presented in figure 3.1 below.

Figure 3.1: The Internet Search Model with measurement variables



The table below contains the variables in the model along with the measurement scales used. For each measurement scale, the items of which the scale consists are shown to the right of the scale.

Table 3.1: Variables for the general study

Variable		Items
Use of information source	**On-line versus off-line** **Newsgroups** **Producers' websites** **Retailer's websites** **Consumer oriented commercial websites** **Independent Search Agents** **Chat** **Independent consumer websites** **Other sources**	Relative use information search is measured as amount of time used and information gathered on the Internet versus off-line and split between possibilities. For each possibility is indicated relative use (s20a-s211a), amount of information found (s20b-s211b) and value of acquired information in making purchase decision(s 20c-s211c). Scale from 1 to 5 indicating not important to very important General use is measured as how often, the source is used. In addition is asked for regular use of e-mail, gaming and mobile phone software and downloading of music (s 41-s412).

To be continued…

Table 3.1: Variables for the general study

Variable	Items
Amount of search	
Time used in total	Categories in hours (s1e)
Product knowledge (Bloch, Ridgeway and Sherrell) 5-point Little to a great deal	Relative to other people (s1b1) Relative to friends (s1b2)
Product interest	As Product knowledge (s1a1 and s1a2)
Purchase experience (never, once, twice, three times+)	Bought previously (s1c)
Importance of decision 7-point scale (Ratchford)	Very important/very unimportant (s1d1) Decision requires a lot/little thought (s1d2) A lot/little to loose if you choose the wrong brand (sd3)
Belief in availability of information 5-point Likert scales	I can easily find information on products on the Internet (s52) When I am to choose between goods, there is much helpful information on the Internet (s51) It is possible to get advice on product choice on the Internet (s53) There is good information on prices available on the Internet (s54)
Perceived credibility of source (Lichtenstein Bearden) As the last two items were the same when translated, only 4 items were in the questionnaire	Insincere/sincere (s55) Honest/dishonest (s56) Dependable/not dependable (s57) Not trustworthy/trustworthy (s58) Not credible/credible
Cost of use of source Scale	Economically more/less expensive (s59) Timewise more/less demanding (s510) More or less psychological effort (s511)

To be continued……

Table 3.1: Variables for the general study

Variable	Items
Importance of cost elements Scale	Economy (s512) Time (s513) Psychological effort (s514)
Perceived relative cost and benefit in general in using the Internet for information search Scale	More or less costly in terms of total effort (s3e, s3f)
Type of shopper	
Importance of contact with sales staff (Lumbkin)	I like to shop where people know me (s61) I like to shop where clerks know my name (s62) I try to get to know the clerks in the store where I shop (s69)
Shopping enjoyment (Lumbkin)	I like to go shopping with a friend (s64) When I talk to my friends, shopping is a topic of conversation (s65) (R) I enjoy going to regional shopping centres (s66) I often combine shopping with dinner or lunch at a restaurant (s67) I get a psychological lift from shopping (s68) Shopping gives me a chance to get out and do something (s63) I am interesting in shopping (s612)

To be continued…

Table 3.1: Variables for the general study

Variable	Items
Importance of the Internet	
Importance of the Internet (Hoffman and Novak) Due to translation problem the items were reduced to three	Important/unimportant (s52) Irrelevant(relevant (R) (s53) Means a lot to me/means nothing to me (s51) Matters to me/doesn't matter to me Of no concern/of concern to me (R)
Locus of Control (Rotter)	What happens is my own doing (s621) Getting people to do the right things depends upon ability not luck (s624) When I make plans, I am certain that I can make them work (s622)
Web usage (Hoffman and Novak)	Startweb 6 categories (s3a) Timeuse 6 categories (hours/week) (s3b)
Additional variables	
Use of Internet private or work related purposes	Percentage brackets (s3c)
Gender	(s71b)
Age	Number of years (s71a)
Workload	Hours pr. Week (s71c)
Household Income	Income brackets pr. 100.000 DKK/year before tax (s78)
Level of education	According to the Danish system (s71d)
Postage Number	(s79)
Number and age of children	(s73-77, a-b)
Partner/no partner	(s72a-d)

to be continued…

Table 3.1: Variables for the general study

Variable	Items
Facilitating factors	
Innovativeness in shopping (2 items from Lumbkin's scale for shopping enjoyment)	When a new store opens, I am among the first to try it (s610) I like to try new and different places to shop (s611)
In-home Shopper (Hawes and Lumbkin) (Internet possibility added)	I love to browse through catalogues (s619) I am ordering more things from home to save energy (s613) By shopping at home through mail/phone order/internet saves me a lot of time (s617) I don't like to shop at home through mail/phone order/Internet (R) (s615) I use mail/phone order/Internet from home because I can't find what I want at local stores (s618) Mail/phone order/Internet from home is more convenient than going to the store (s614) In-home shopping via mail/phone order/Internet costs too much for what you get (R) (s616)
Role overload – reduced to two items	I feel I have to do things hastily and maybe less carefully in order to get everything done (s620) I am too busy in my every day life (new item) (s623)

Rem.: (R) indicates scale is reversed

3.4 Variables Relating to Use of Newsgroups

Hypotheses 11-17 stated in chapter 2.6 relate to consumer information search in newsgroups. This chapter discusses how to measure the concepts in the hypotheses in order to make these hypotheses testable.

Hypothesis 11 regarding use of newsgroups in general is tested in the general study. The remaining hypotheses include the following concepts: User segments, motivation factors, sources of information, and opinion-leadership and – seekership. In addition to these concepts, the same socioeconomic variables and measures of relation to the Internet in general as the ones chosen for the general study are included in the study. This is because these measures are necessary in order to be able to compare the newsgroup user to the 'average' user of the Internet. The items for all concepts are displayed in table 3.2 in chapter 3.4.7 at the end of this chapter.

3.4.1 Setting the context

As discussed in chapter 3.3.1 concerning the same issue in the general study, it is necessary to set the context for the respondent before he starts answering the questions. Again, it is important that the respondents understand the topic of the study before starting the survey, as this makes understanding the individual questions easier.

In the newsgroup study, it is possible to develop questionnaires tailored for the topic, which is the focus of the newsgroup in question. This makes questions more understandable than generalized questions and, therefore, this opportunity was taken.

In addition to this, the context was set by a thorough introduction to the subject of the study and how to fill in the questionnaire.

3.4.2 Determination of Segment Membership

As described in the theoretical chapter and the hypotheses, Kozinets (1999) works with two dimensions in membership of the group: Self-

centrality of the product to the consumer and relation to the group. This again means that it is important to be able to define how the consumer feels he is related to the newsgroup and to the product in question.

On the general level, level of attachment to the newsgroups can be studied by observing frequency in the communication going on between the individuals in the newsgroup. This communication tells us if the newsgroup in question is tightly knit or the ties to the group are weak. This way of measuring ties in communities is recommended by Wellman (1990). In this study, he found that frequent contact is related to giving material aid and in newsgroups, the contact between individuals arises from participation in the same discussions, or threads as discussions taking place in virtual communities are called. Frequency of communication can be measured as number of encounters the same individuals have with each other over a given period of time and number of times each individual in the group is posting or answering questions in the group. Other measures could be communication between the persons taking place by other means than by posting in the group – that is via telephone, e-mail, or in person. The latter way of measuring ties to the group was rejected because observation in the newsgroups revealed that communication outside the group is quite rare and, therefore, not a good measure of tie to the group. Also, the topic of interest in this study is the power of newsgroups as an information source in itself and, therefore, the interesting thing to measure is tie to the group as such and not to persons in the group via contact outside the group.

On the individual level, tie to the group can be measured in several ways: A more objective measure relating to behaviour or subjective statements from the respondents on felt tie to the group. Objective measures of strength of ties are used in structural network analysis as described in Wellman (1999, 2001). The strength of ties between people is typically measured by number of meetings in some way – being on the same boards, in the same club, or number of times families are connected by marriage. As this study is quantitative in nature even though the hypotheses are inspired by Kozinets, this way of measuring social tie is the one mostly in accordance with the study. Ties can then be measured by registering which members participate in which threads. A mul-

tiplication of the resulting matrix by its transpose gives us number of times the same individuals meet each other in threads in the newsgroup.

If we are able to observe how often all members of a newsgroup access the newsgroup, such a registration would be an objective way to measure strength of tie to the newsgroup, and between individuals in the newsgroup. However, in the newsgroup there are lurkers who do not communicate to the group, but only read what others write. The consequence of relying only on observations of activity in the group would be that lurkers would be measured to have no tie to the group. In fact, they would not be perceived as members of the group at all. Therefore, it is necessary to ask members of the newsgroups how often they access the group and use the answer to this question as a measure of strength of the tie to the group. The measure can be number of accesses per time unit, or frequency of access related to a time unit. As some members of the newsgroups do not access the group very often, the time unit chosen is a month. Respondents probably do not keep track of number of accesses per month but know if they access the group several times per day, every day, several times per week and so on. Therefore, this scale is chosen although the scale properties of number of accesses per months are better as this measure is a true ratio scale whereas the scale chosen is ordinal, and only with good will can be regarded as interval scaled.

Additional objective measures of tie to the group are how long the respondent has been a member, and how much time he spends in the group while on-line. The last is also an interesting measure in the sense that it gives us knowledge about the way consumers use the group – quick scanning or thorough reading and communication?

The second dimension in Kozinets' model for segmenting members of virtual communities, as shown in figure 2.7, is Self-Centrality of consumption activity. Reading of the article in detail reveals that this, to Kozinets, is the same as hedonic value of the product in the consumers' life. Importance of the product to the respondent is measured using the same scale as in the general study for product interest, as interest indicates how high self-centrality, the product class has to the consumer. The reasoning behind this is that if the product is central to the con-

sumer and his perception of his own identity, this must produce interest in product class. This is supported by Kozinets exemplifying self-centrality of the consumption activity by collecting something as a hobby, and by Chauhuri's research cited in chapter 2.2.1.

In addition to measuring interest in the product, knowledge about the product is measured using the same scale as in the general study. This gives opportunity to compare the newsgroups members to average Internet users in terms of knowledge about the product. This is interesting, as members possessing knowledge in the group is a prerequisite for good advice to be given in the group.

3.4.3 Motivation factors

Individuals enter newsgroups for various reasons. Among them is the opportunity to get advice which can help in choice of brand, product, or solution to a problem. Kozinets (1999) operates with 4 different motivational factors: Informational (seeking information), transformational (giving advice), recreational (getting entertained) and relational (socializing).

I have not been able to find any scales measuring these concepts in relation to consumer information in newsgroups. Therefore, it was necessary to develop scales for this study. It was not possible to test the scales in advance on the relevant audience. Therefore, the measurement model must be tested before testing the hypothesized relations in the data. The items designated to measure the 4 factors are shown in table 3.2. For entertainment and giving advice, there is only one item for each factor. This is due to my own lack of creativity.

3.4.4 Communication and Perceived Influence

As stated before in this chapter, communication in the group can be measured partially by observation of postings in the group. However, the observed communication can only be linked to the unobservable variables for the individuals who are both posting in the group and answer the questionnaire, and not for individuals, who do only one of the two. Therefore, as for measurement of link to the group, also when it

comes to measurement of communication in the group, it is necessary to pose questions relation to this area directly to the respondents in the survey. However, with a sufficient overlap between observable members and respondents, number of postings can be used for validation of the answers in the survey.

Communication in the group is created by persons posting questions to the group and/or answering questions from other members of the group. In the hypothesis on communication in the group, these activities are linked to opinion leader/seeker score, and also to score on the motivation factors and segment membership. The reasoning behind this is that, in order to get communication going, there has to be both persons seeking advice, and persons willing to give advice. To measure if the respondent is active in communication in the group, he is asked if he has posed questions asking for advice on product choice, had this been relevant. The respondent is also asked whether he has answered postings asking for advice on product choice, if he believed to possess knowledge about, or experience with, the product class. The scales used were 6-point scales from never to every time.

The influence advice given in the group exerts on the individual consumer is at the focal point of the research questions regarding consumer search in newsgroups, and the related hypotheses. On this issue, one can lead the same discussion about past or intended behaviour as in chapter 3.3.1. In addition to the arguments in that discussion, it is necessary to consider that purchase decisions may not have been relevant in the period of time the respondent has used the newsgroup. Therefore, lack of relevance must be an option in recording past behaviour. For the same reason, it is relevant to ask, not only for influence in the past, but also for expected influence in the future.

3.4.5 Opinion-leadership and -seekership

Hypothesis 15 states that communication in the newsgroup is enhanced by different motives to visit the group which fit each other well. The motives are governed by different types of personality in terms of seeking and wanting to give advice. Consumers prone to seeking advice are also called opinion seekers, whereas consumers influencing others are

called opinion leaders (Rogers, 1995). To find out whether these types of consumers are in the newsgroup we have to find measures of opinion leadership and opinion seekership. Like some of the other concepts, these concepts can be measured both by observation and questions. As the discussions in newsgroups relate to specific products or issues, it seems most relevant to measure opinion leadership regarding a specific product rather than the more general measurement of whether the consumer is a Market Maven (opinion leader on many products).

A way of observing opinion leadership in a newsgroup is to count number of times a member of the group posts to the group. Likewise, opinion seekership can be measured by counting number of times a person asks questions to the group. These measures cannot stand alone, though, as persons may be opinion seekers in the group, but not in general answers and the opposite way around. Also, that a person is active in the group is not the same as him being influential. Finally, as before, observations cannot always be linked to the unobservable variables, as it is not possible to identify all respondents from the survey in the newsgroups. Therefore, observations of activity in the group may not be the best way to measure these concepts. The task then is to find survey scales for measuring opinion leadership and opinion seekership.

As opinion leaders are people who influence others, one way of identifying opinion leaders is to ask the respondents if they find that there are members in the group who have influenced their choice of product. This is the sociometric method (Rogers, 1995; Schiffman and Kanuk, 2000). If somebody is influential, it is interesting to find out to what extent, they are influential. Also, it is interesting to know whether individuals who are influential are so via only contact through the newsgroup or if the influence has been established or enhanced via more personalized meetings. That is, via meetings in person, on telephone or via e-mail.

To find out if opinion leadership and opinion seekership is linked to activity in the group, it is necessary to be able to link the two. If all opinion leaders, identified by others, can also be identified in the sample, this is possibly the best way to do so. However, there is no certainty that this will be the case. Therefore, it is safest to measure opinion leadership

by asking respondents themselves, even though there is a risk of them over reporting their opinion leadership, as this is a desirable trait (Jacoby, 1972; Schiffman and Kanuk, 2000).

Opinion leadership by self-designation can be measured on a scale developed by Davis and Rubin (1983, a modification of a scale developed by Rogers in 1961) designed to measure tendency to provide information to and influence others about a certain product. The scale has 7-items, each with a yes/no answering opportunity. The items are shown in table 3.2. The scale has been tested for predictive validity and has a reported alpha of 0.82 (Bruner and Hensel, 1992).

The opposite of being an opinion leader, is in some sense, to be an opinion seeker. Bearden, Netemeyer and Tell (1989) have developed a set of two scales measuring respectively if the person is an information seeker (interpersonal) and interpersonal influence susceptibility (normative). The former scale has 4 items and the second 8 items. It is the expectation that individuals entering the community in a transformational mode rate high on opinion leadership whereas individuals entering the community in an informational mode rate high on information seekership, and are susceptible to interpersonal influence. It is possible that the same person can be both an opinion seeker and an opinion leader or neither of the two (Flynn et al, 1996). The scales used were developed using Churchill's (1979) procedure and, therefore, were tested thoroughly. The alpha coefficients were 0.82 and 0.87 respectively. The items for the scales are displayed in table 3.2.

3.4.6 Substitution of Traditional Information Sources

The two final hypotheses states that information sought in newsgroup has, to some extent, replaced information formerly gathered at other personal sources.

Chapter 2.4 discussed use of information sources, and from this chapter and the resulting two hypotheses it is clear that there has to be questions relating to both marketer and non-marketer dominated sources. The relevant sources are presented in chapter 2.4. As replacement of sources is mostly of interest if the source was used before, there has to be ques-

tions on both former use and level of replacement of each source by search in the newsgroup. The scales range from not at all to much and completely, respectively, with 4 points on each scale. I considered asking for percentages, as this would be a scale with higher properties than the one chosen. However, it is not probable that percentages given by the respondents would be accurate and, therefore, precision in the scale is off-set by low reliability in the actual measurement.

3.4.7 Summarizing Questions for Newsgroup Study

The table below summarizes measurement scales chosen for the newsgroup questionnaire.

Table 3.2: Variables for the newsgroup study

Variable	Items
Segment	
Interest in product	(see general study) (g1a1, g1a2)
Knowledge of product	(see general study) (g1b1, g1b2)
Relation to group	How long been a member (related to months) (g3a) How often access the group (related to weeks/days) (g3b) For how long is the group accessed (related to minutes) (g3c)
Web usage	As in the general study (+ general evaluation of cost and benefit of using the Internet for information search)
Socioeconomic variables	As in the general study (adults only)
Opinion leaders in group	
Person	Name (g3f1a-g3f5a)
Influence	Scale from a little to decisive (g3f1b-g3f5b)
Additional contact	E-mail (g3f1c-g3f5c) Telephone (g3f1d-g3f5d) Meeting in person ((g3f1e-g3f5e)

To be continued……

Table 3.2: Variables for the newsgroup study

Variable	Items
Communication with and perceived influence from the group	
Communication	How often do you Ask questions to the newsgroup, if you have a relevant problem regarding purchasing a product? (g4d1) Answer questions on product purchase, if you have knowledge about the subject? (g4d2) Answer questions on product purchase, if you have personal experience in the subject? (g4d3)
Influence from group	Have chosen based on advice from group (g511) Will, in the future, choose based on advice from group (g512)
Motivation factors	
Seeking information	I get information along the way which can help me, if I am going to purchase things, for which consideration in necessary before a choice can be made (g4e1) I like to keep up on the subject because it is of interest to me (g4e4) The advice you get is more serious than other places on the Internet (g4e5) I can get answers to specific questions (g4e9) Guidance is from private persons (g4e10)
Giving advice	I can be of help to others in their choice process (g4e7)
Getting entertained	It is entertaining (g4e6)
Socializing	I get in contact with others, who are interested in the topic (g4e2) It is a way to be 'together' with others who share the same interest (g4e3) I feel, I have come to know the others, and I would like to meet them again (figuratively speaking) (g4e8)

To be continued……

Table 3.2: Variables for the newsgroup

Variable	Items
Supplementing or substitution of sources former used	Used before – scale from not at all to a lot The newsgroup has replaced the source – scale from not at all to completely
Marketer-dominated sources	
Retail sales personnel	g53a,b
Brochures	g55a,b
Advertisements distributed to mailbox	g54a,b
Non marketer dominated sources -	
Friends, you see often	g51a,b
Friends, you see seldom	g52a,b
Family, you see often	
Family, you see seldom	g53a,b
Colleagues	g54a,b
	g56a,b
Role in group Opinion leadership (Generalized) Davis and Ruben (1983) In the questionnaire transformed to 5 point scale	In general do you like to talk about _____with your friends (r)? Yes (2) No(1) (g61) Would you say you give very little information (1), average amount of information (2) or a great deal of information (3) about ____ to your friends? (g62) During the past six months, have you told anyone about some ____? Yes (2) No (1) (g63) Compared with your circle of friends, are you less likely (1), about as likely (2) or more likely (3) to be asked for advice about ____? (g64) If you and your friends where to discuss ____, what part of would you be most likely to play? Would you mainly listen to their ideas (1)? (g65) Would they mainly listen to your ideas (2)? Which happens more often? Do you tell your friends about some ____ (2)? (g66) Do they tell you about some _____ (1)? (g67)

To be continued……

Table 3.2: Variables for the newsgroup study

Variable	Items
Role in group continued…	
Information Seeker Bearden, Netemeyer and Teel (1989)	To make sure I buy the right product, I often observe what others are buying and using (g68) If I have little experience with a product, I often ask my friends about the product (g69) I often consult other people to help choose the best alternatives available from a product class (g610) I frequently gather information from friends or family about a product before I buy (g611)
Interpersonal Influence Susceptibility Bearden, Netemeyer and Teel (1989)	I rarely purchase the latest fashion until I am sure that my friends approve of them (g612) It is important that others like the products and brand I buy (g613) When buying products, I generally purchase those brands that I think others will approve of (g614) If other people can see me using a product, I often purchase the brand they expect me to buy I like to know what brands and products make good impressions on others (g615) I achieve a sense of belonging by purchasing the same products and brands that others purchase (g616) If I want to be like someone, I often try to buy the same products, they buy (g617) I often identify with other people by purchasing the same products and brands they purchase

3.5 Survey Method and Questionnaire Design

In chapter 1.1 discussing research paradigm it was decided that the basis for an empirical test of the hypotheses should be a survey. There are several different ways to execute surveys. Two important aspects are how to contact consumers and how to collect data. In principle, there are four different types of survey methods (Malhotra, 1999): Person-to-person interviewing, telephone interviewing, mail interviewing, and electronic interviewing.

Person-to-person interviewing can be ruled out as a possibility, as the aim is to collect one sample which is fairly representative for the Danish Internet user and another sample among newsgroups users. This means that the first sample should be spread geographically and the second sample probably is spread geographically due to the population in question. Therefore, person-to-person interviewing would be both very time consuming and very expensive. Telephone interviewing is not suitable because of the many questions which are a result of the many items in the measurement scales used. This leaves us with the choice between mail interviewing and electronic interviewing.

A number of things count to the advantage of electronic interviewing: It is faster and cheaper than mail interviewing and, as respondents are Internet users, they all have access to the Internet as is necessary for this kind of interviewing. Also, it would be difficult to get mail addresses for newsgroup users. That electronic interviewing is faster than traditional mail surveying can be seen in table 3.3 below which reports response times in studies comparing mail, electronic and fax-surveying.

Table 3.3: Response time using different survey methods

	Response time, Internet	Response time, mail	Response time, fax
Coupier et al (2001)	50% of responses received within 3 days		
(Cobanoglu, 2001) e-mail with link to questionnaire	5.97 days (mean)	16.46 days (mean)	4 days (mean)
(Schillewaerdt et al. 1998), Advertisement in magazine with link	50% of responses received within 16 days		
Posting in news-groups	50% of responses received within 3 days , 79% of responses received within 1 week		
Advertisement on other website with link	15% within 1 day		
Dommemeyer and Moriarty, (2000)	Between 2.5 and 8 days (mean)	Typically around 12 days (mean)	8.8 days (mean)
Ilieva et al. (2002) (e-mail)	5.59 days on average	12.21 days on average	

There are, however, also reservations to doing surveys electronically: It can be problematic to expose respondents to the questionnaire, and also to develop the questionnaire.

The first problem is how to expose respondents to the questionnaire in a way which makes respondents motivated to answering the questionnaire. Studies (*e.g.* Schillewaerdt et al., 1998) show that the response

rate in electronic surveys is very low if respondents are recruited by advertising on the Internet or in magazines. In order for it to be realistic to expect a response rate in an electronic survey at least as high as in a traditional mail survey, it is necessary to contact respondents, whom we want to participate in the survey, directly (Ilieva et al., 2002; Malhotra, 1999) and encourage them to answer the questionnaire. This is more time consuming than just advertising for participants and offsets some of the cost and speed advantage compared to mail surveys. This issue will also be discussed further in the following chapter which discusses sampling issues in detail.

The second problem pointed to is the problem of development of the questionnaire. In traditional mail surveying the questionnaire can be developed using an ordinary word processor. In electronic surveys, however, the questionnaire has to be developed in such a way that it can be published on the Internet and also record responses in some way. Suitable software for development of electronic questionnaires is available, and it is also possible to buy this service. However, these solutions are not feasible in this kind of project, as they are prohibitively expensive. The price of software is typically around 50.000 DKK if the software is also suited for recording responses. The consequence of this is that the researcher has to develop the questionnaire using software tools that are not especially designed for this kind of task, and which it is not common knowledge how to use.

Research on the topic revealed two different feasible ways to develop questionnaires publishable on the Internet: Using Adobe Acrobat which has good graphical tools and also is able to gather the data, or programming the questionnaire in HTMl using Microsoft FrontPage or another tool like, for instance, Microsoft Notepad, for the HTML-programming. Testing the usability of both options in terms of appearance, time for download, and also considering that the respondent, in order to download a questionnaire developed in Adobe Acrobat, would have to have Acrobat Reader installed, I ended up choosing the HTML option, even though this meant that the questionnaire might not appear quite as nice, and I would have to learn some HTML. The research performed revealed that this was not an insurmountable task, especially considering

that use of Adobe was not straightforward, either, and that download time for a questionnaire in HTML with an ordinary modem was 30 seconds compared to around 2 minutes for the same questionnaire developed using Adobe Acrobat. Therefore, weighting pros and cons, the final choice fell on doing both surveys as electronic surveys with a questionnaire developed directly in HTML.

The next choice is between the two different kinds of electronic surveying: E-mail with the questionnaire embedded or attached, or publishing the questionnaire on the Internet. Because of fear for viruses, it is quite common not to open attachments. To confirm this, an empirical study showed very poor response rates for contact via unsolicited e-mails with attached questionnaire (Dommemeyer and Moriarty, 2002). Because of this, unsolicited e-mail with the questionnaire attached was ruled out. The other possibility, which was to send the questionnaire embedded in an e-mail, was ruled out for two reasons. First and foremost, this requires that the respondent has e-mail software which is able to read HTML. This is quite common, but often the option is shut off without the user knowing it and not knowing how to turn it on, either. Second, once the questionnaire is mailed to the respondent, there is no possibility to correct errors. As my skills in HTML-programming are limited, it seemed to be a good idea to keep an option open to correct errors along the way.

The conclusion to these pros and cons was to do the survey using a web-based questionnaire built in HTML (see for example Powell, 1999). This means that the answers the respondent gives automatically are recorded in a text-file according to values coded along with each option for each variable. This again means that responses are automatically registered in the computer. This is considered an advantage, as there is then no human coding error. The down side to automatic registration is that errors are also repeated automatically. Therefore, one has to be very careful in building the questionnaire.

The questionnaire is designed according to general rules of thumb for questionnaire design as described in Malhotra (1999) and Dillman (2001). This implies that the questionnaires in both surveys are built

much the same way. Prints of the questionnaires (a sample for the newsgroups) are reproduced in appendix A. As the prints do not look exactly as the questionnaires do on-line, the questionnaires are also available online at: www.sam.sdu.dk/~alj/sporgemail.htm and www.sam.sdu.dk/~alj/grupper.htm

The questionnaires start with an introduction telling the respondent about the purpose of the study, and how to fill in the questionnaire. The respondent does not have to be on-line while filling in the questionnaire – only when downloading and submitting. As these processes cost respondents money, it is important to estimate how long they take, in order to be able to inform respondents of this, as required by the ESOMAR Guidelines on Internet Research (Nancarrow et al., 2001). In this case, this was not really an issue, as download time for the questionnaires was less than 30 seconds using a standard modem.

After the introduction the context is set as described in the chapter on operationalization of the concepts in the Internet Search Model. The flow in the following sections is, with a slight modification, in line with the general rule that questions should be sorted in sections containing the different subjects in the questionnaire, and that the flow should be from general to specific. After setting the context by choosing a product, it is natural to start with questions on relation to the product and the decision. The next topic is general interest in the Internet, followed by use of the newsgroup in the newsgroup study. After this, questions relating to beliefs and personality traits are placed, and in the end, questions on socio-demographic variables are posed.

The scales chosen in the previous chapters differ in terms of number of points. As many different scales make reading, understanding, and responding more difficult for the respondent, there has been a modification in some scales. After the modifications, scales are Likert-type scales with 5 boxes as answering opportunity whenever this is possible.

At the end of the questionnaire is placed a 'button' which the respondent has to push to send the response to the researcher. This button can be configured in several ways. The most straightforward from a program-

ming point of view is to start the respondents' e-mailing software. This way, the response is sent as an attachment to an ordinary e-mail. Unfortunately, this way of responding proved not to be without problems: The respondents had problems starting their software, some of the attachments came back blank and some e-mails were without attachments. As a result of this, I had to consider other ways to configure the 'Send' button. There are two possibilities: To receive the response in an attached file, or that the response is written directly into a preconfigured data base at the receiving server. In both cases, the response is sent directly via the Internet without activating any e-mail software. After conferring with the EDP-department, I settled for the attachment solution with subsequent infiling of responses into SAS.

In order to be able to control for non-response and for multiple responses from the same person, it is necessary to be able to control for who each respondent is. To do this, I asked respondents to indicate their e-mail address in the questionnaire. To be able to do follow-up surveys in the future, I also asked for permission to contact them again. Respondents are not completely anonymous this way. However, this is not considered to be a problem as respondents are not recognizable because their e-mail address is known. Furthermore, research has shown that lack of anonymity is often not a problem if the questions are not sensitive (Basi, 1999; Kalafatis, 1996).

The next step is a pilot test of the questionnaire. Initially, a small sample of university staff tested the questionnaire. This test was done mostly in order to test whether answers from the questionnaire would reach me safely. During this test I encountered some problems:

First, that it is not unproblematic to send answers via e-mail as this is dependent on a correctly set-up e-mailing function from within the web-browser.

The second problem was one of missing data. The problem was that if the respondent did not fill in some of the questions, the field was not returned and, therefore, it was hard to read the data automatically because the fields were not always in the same place. Also, if a text field was

empty, it could end up at different places in the data set. This problem was handled by hidden fields containing default values, thereby securing that no fields are empty. Another way to overcome this problem is to show these pre-chosen fields as a 'don't know' option.

After these corrections had been made, the questionnaire was tested on a group of students on the Master Programme (5 respondents) during a session where they answered the questionnaire on the Internet while I was present. The students made oral comments as they answered the questionnaire. This test revealed several problems:

First: The questionnaire did not include a question on whether the respondent had searched for information before buying the product in question.

Second, it was not stated whether the term product (in Danish: 'vare') included services, or if the study was exclusively targeted at product information search in relation to physical products.

Thirdly, the question regarding how often the respondent used the Internet had to be changed, as the respondents all agreed that the difference between 'once every fortnight' and 'more seldom' ending in 'never' was too large, whereas the distances between 'every day', '1-2 times pr. week' and 'several times a week' were far too small. Therefore, the possibility '1-2 times pr. week' was removed and a possibility of 'once a month' was added.

The final and largest problem in the questionnaire was the one which was expected to be the largest one: Allocation of time and importance to the various sources on the Internet. In the first draft of the questionnaire, respondents were asked to indicate the percentage of time, they used on each type of source. However, this was, as was partly expected, far too difficult a task, even for students on a masters' course in Marketing Research. There seemed to be two separate problems: one of splitting percentages between sources and one of filling in one column at a time. Because of these problems it was decided to renounce on the detailed information it is possible to obtain from using a constant sum scale, as

splitting 100% is and, instead, ask respondents to rate use of each type of source on a scale from 1 to 5, where 5 is much and 1 is little. Another possible approach would have been to ask respondents to rank order use of the various types of sources and the information retrieved from these sources. However, this would, for one thing, not reveal whether one type of source was used much more than another (indicating strength of use), and the respondent would still have to answer the question for each column at a time.

In addition to these larger problems, the students, of course, encountered several smaller problems in terms of wording and order of the questions. These comments were also considered and applied to the questionnaire, whenever it was deemed in order to do so. After this process, the questionnaire was, once again tested on the students, and also on a selection of staff at the University of Southern Denmark.

After the final test and resulting small corrections, the questionnaire was ready for a test in the relevant population: Internet users and members of newsgroups. As these tests are related to sampling and recruitment of respondents, results from these tests are reported in the next chapter treating this issue.

3.6 Sampling and Contacting Respondents

At the beginning of this chapter, it was decided to do two separate surveys. The two surveys are very different by nature. The general study tries to test hypotheses about determinants of consumer search processes on the Internet in general, and therefore has to be representative of the average Internet user. Opposite to this is the newsgroup study which tries to shed light on consumer communication and influence within a virtual community. In this last case, the important thing is not that the sample is representative of the Internet user as such, but that the sample covers newsgroups users. Therefore, sampling necessarily will be very different in the two cases, and is therefore treated in two separate chapters below.

3.6.1 The Study on Internet use on the general level

Many surveys on the Internet, *e.g.* the GVU surveys, have used self-selection of respondents via postings on often-visited web-sites. There are several problems in using self-selection. The worst problem is probably that respondents choosing to answer the questionnaire have to actively choose to do so without anybody asking them in person. As participation in surveys requires time and effort, only respondents who are interested in the topic, or who have a lot of time on their hands, are inclined to answer. Therefore, the sample will be skewed when self-selection is involved. In addition to this, the sample becomes even more skewed because only respondents visiting the website which contains a link to the questionnaire get the offer of answering at all. As this study attempts to model behaviour of the common Internet user, this is a serious problem. Therefore, it seems more appropriate to actively seek out persons, whom we want to be respondents in the survey, and then contact each of these persons, encouraging him or her to participate in the survey. This contact could be in person, via telephone or e-mail.

As stated in the above, a main concern in the study concerning Internet use on the general level is to be able to infer from the sample to Internet users as such. A sample suited for this can be obtained by drawing a simple, random sample among the Danish Internet population. There are, however several reasons why this is not quite so simple.

First, it is very hard to draw a completely random sample among the population as such, as it is difficult to establish an adequate frame, from which it is possible to draw the sample, especially with limited financial resources. Therefore, one would always be forced to use some kind of approximate list of the population, as for example a telephone directory. The consequence of drawing a sample this way is that the sample would not be completely representative, as many telephone numbers are not listed. Furthermore, especially young people are often not listed because they have only cell phones, and are not buying their connection from companies which keep phone directories.

Second, in such a sample, there are many respondents who are not relevant for this study, as they do not use the Internet at all. One could say,

though, that as 75% of the population has Internet access at home or at work (Danmarks Statistik, 2002a), and everybody has access via public libraries, this is not an excuse for not using a theoretically good way of drawing a sample. Still, as there are many persons, who are not relevant, and who would therefore probably be non-respondents, the above is considered a valid argument.

Third, in order to be able to contact respondents via e-mail, it seems practical to use a sampling frame which contains e-mail addresses. Surveys have shown that 72% of the individuals who have Internet access use the Internet for e-mail (Danmarks Statistik, 2000a). It therefore seems appropriate, bearing the arguments in the above – especially the practical part - in mind to test the possibility of using a list of e-mail addresses as a sampling frame.

There are, however, several practical problems in using e-mail addresses as a sampling frame. The first problem is that there is no exhaustive directory of e-mail addresses of the Danish Internet population, and therefore, it is necessary to use directories which contain only a part of addresses. There are e-mail panels available which are based on individuals, who have been randomly chosen, and then via telephone have accepted to participate in surveys, but it is prohibitively expensive to get addresses this way. Therefore, in this case, the list of e-mail addresses has to be available to the public for free, or at a little charge.

Research on the Internet at Danish e-mail providers and on the top 10 web properties during September 2001 showed that there are two available directories containing a larger number of Danish e-mail addresses. The directory containing the largest number of e-mail addresses is at www.epost.dk, sponsored by TDC, who also owns the second most visited web property. This directory is built on data from TDC Internet, Tiscali (America Online), and other, smaller providers supplied by listings made by individuals themselves. www.epost.dk reports to contain 930.000 e-mail addresses in November 2000. It is not possible to see the number of addresses listed in the directory at this point in time, as the user can see only 40 addresses at a time due to fear of abuse of the service for marketing purposes. In the evaluation of this directory in terms

of usability as sampling frame for this study, it is important to notice that many of the addresses in this directory have been put there automatically, when individuals acquired their e-mail address at their supplier. Also, e-mail addresses of many individuals have been registered at www.epost.dk by the company, they work for and the address is then their e-mail address at work, and not their personal address. Therefore, they might consider an unsolicited e-mail directed to them in person as spam.

Unfortunately, there are also many e-mail addresses belonging to children in www.epost.dk, as one of the providers of addresses to the directory is provider for public schools (www.skolekom.dk). Also, there are official e-mail addresses on companies. However, it is possible to 'clean' the directory for most of these unwanted addresses as information on the school or company which possesses the e-mailing address in question appears on the screen. For around 5% of the individuals in the directory, there is more than one address. The reason for this is that some individuals have added an address themselves in addition to the automatically posted address. Sometimes, it is possible to see that two different addresses belong to the same individual. However, this is not always the case, as addresses posted by companies are posted along with the company address and addresses posted by individuals mostly indicate home address.

The other Danish provider of an e-mail address directory is joes.jubii.dk. This list is sponsored by www.jubii.dk, owned by Lycos Network. This is the top web property in Denmark. Holders of addresses in this directory have all actively registered the addresses at joes.jubii.dk. Also, the registrant has determined what information he or she wants to appear along with the e-mail address. At joes.jubii.dk, individuals are not required to reveal neither address nor postal code. However observations[15] have shown that many do so. The demand for the persons to register actively has as a consequence that individuals in this directory on average have been more active as regards getting listed

[15] Search on several names and counting number with and without address and/or postal code

in the directory than individuals listed at www.epost.dk. Therefore, individuals on joes.jubii.dk are probably less representative of the average Internet user than individuals listed at www.epost.dk. On the other hand, the directory at joes.jubii.dk may cover the many possible addresses better than the directory at www.epost.dk which, for instance, hardly contains any hotmail addresses.

There is no information on how large the number of addresses registered at joes.jubii.dk is. However, it is possible to make an estimate, based on a search in the directory. A search using the extensions .dk, .com and .net resulted in 128.833 e-mail addresses (112.045, 14.716 and 2.072 respectively). There are other extensions than these, but searches using the names Dorthe, Ole, Michael, Jesper, and Jens, all common, Danish names showed that other extensions than .dk, .net and .com extend to around 1%, and is mainly due to extensions from other countries like .be, .no and .de. Therefore, it seems fair to assume that the total number of e-mail addresses registered at joes.jubii.dk is around 130.000 which, in any case, is a much smaller number than the number of addresses listed at www.epost.dk. There seems to be an overlap of only about 10% between the two directories.

For both directories it poses a problem that most of the addresses are on men. A search at joes.jubii.dk reveals that searches on very common Danish female names result in the following number of addresses: Dorthe: 322, Birgitte: 354, Susanne: 640 and Lone: 546. Compared to the common male names, Jens, Ole, Jesper and Michael which produce from 1755 to 2431 addresses, this is few. In relation to the present study, the problem is different from finding addresses via telephone directories which also mostly list the name of the male in the household. This is because the person selected for participation in an electronic survey is also the respondent whereas a person selected from the telephone directory is representing the sampling unit and is not necessarily the respondent. When using the telephone for initial contact, the researcher is able to control for participation by asking for, for instance, the person in the household who had birthday most recently. As e-mail addresses are more personal than telephone numbers, this is more complicated when using e-mail as a means of contact to possible respondents.

Still another problem concerning use of e-mail addresses drawn from the two directories is that many of the e-mail addresses in the directories are inactive. Especially, this was expected to be the case at www.epost.dk as the addresses in this directory are provided automatically along with Internet access sold. Therefore, some of the individuals on the list are bound to either not use the e-mail address at all or to use another address than the one on the list.

It is important to test for the size of this problem and 'ordinary' non-response using this type of sampling frame, as the initial sample size has to be adjusted for non-response in order to receive a sufficient number of filled-in questionnaires. Also, this is not a viable way of selecting respondents, if the response rate is too low.

To test response rate and possibility of drawing a random sample from the two directories, a pilot test was carried out using an approximately random sample from each directory. Results from this test are shown in table 3.4 below:

Table 3.4: Response using two different sampling frames

	Sent to (number)	Undeliverables	% undeliverables	Responses	Response rate in total (%)	Response rate (% of delivered mails)
www.epost.dk	61	9	14.75	4	6.56	7.69
joes.jubii.dk	51	14	27.45	7	13.73	18.92

An inspection of the table shows that the percentage of undeliverables is quite high. It is larger for the sample drawn from joes.jubii.dk than from www.epost.dk. This is probably because there is no automatic update of the directory at joes.jubii.dk. For both of the samples, the response rate is quite low; about as low as in sampling via ordinary mail with no previous accept of participation (Darmer and Freytag, 1996; Malhotra, 1999). The response rate for the sample taken from the list at joes.jubii.dk is considerably higher than the sample taken from www.epost.dk. This may be because some of the addresses listed at

www.epost.dk are not actively used even though they can receive mail as described above. The explanation to the higher response rate in the sample drawn at joes.jubii.dk can also be that persons on joes.jubii.dk have actively registered their e-mail address in the directory indicating that they are willing to receive e-mail from persons, to whom they did not give the address directly. Still, the response rate found in this small test is too low for the current study.

As a result of the unsatisfying results from testing usability of e-mail address directories as sample frame, I decided to recruit respondents via telephone in spite of the problems pointed to in the beginning of this chapter. To be able to control for non-response and to send reminders, the procedure followed was to ask for an e-mail address, to which I could send an e-mail containing a link to the website, at which the questionnaire was located.

It was still the intention to draw a random sample. As it was not possible to draw a completely random sample, as described at the beginning of this chapter, it was necessary to think of a way to draw a close-to-random sample. The target was to send the questionnaire to at least 500 respondents. As some of the selected respondents would not meet screening criteria (Internet access), others would refuse to participate, and still others would not be reachable, the prior estimate was that it would be necessary to contact about 1500 persons. The procedure, I followed in order to do so, is described in the following. The procedure is a two-stage cluster sampling procedure, based on a combination of random and quota sampling.

First stage was selection of cities. A list of postal codes in Denmark was numbered from 1 and up. 30 postal codes were selected at random outside the centre of Copenhagen and 10 postal codes (streets) were selected in Copenhagen.

Next stage was selection of citizens (households). The list of households was generated as a combination of the postal code and a randomly selected number between 1 and 20. This was to reduce the number of households drawn in the same city and keep the weight of cities relative

to each other on the final sample. As a sampling frame for selecting households and attached telephone numbers, I used www.teledanmark.dk. This, of cause, results in yet another sampling error, as not all numbers are listed in this directory because some persons do not wish to be listed, and others do not own a cable telephone. This is, as mentioned before, especially a problem among young people, as they often have only a cell phone.

Third stage was selection of respondent as a telephone number is for a household and not for individuals. With the help of two research assistants, I tried to get in contact with the households on the list. It was attempted to reach each number at least three times. Calls were made during April and May 2002.

If the attempt resulted in contact, the person who answered the telephone was asked to participate in the survey. If this person was not an Internet user, and someone else in the household was, this other person was asked to participate in the survey. The respondent was explained a little about the research question and method, and asked if he or she would participate. If the answer was yes, their e-mail address was obtained.

An e-mail containing a link to the questionnaire was then sent to the address within 24 hours after the initial contact. If a response was not received within 2 weeks, a follow-up e-mail was sent to the person in question and then yet another one 2 weeks later if there was still no response. In an attempt to encourage the remaining non-respondents to react, the last e-mail contained an embedded version of the e-mail plus a link to the website where the questionnaire was located. Results from the sampling procedure are presented in appendix B.1 and are commented in chapter 4 containing results from the study on Internet use on the general level.

3.6.2 The study in Newsgroups on the Internet

As stated in the above, in order to study interpersonal influence in newsgroups as an example of virtual communities, it is necessary to study these communities in a separate sample. The sampling process

contains two stages: The first stage is one of deciding on sampling method and the next is how to pick the sample from the sample frame.
The first question to decide on is how to pick the sample. The sample can be spread among many newsgroups and thereby be representative for all newsgroups on the Danish USENET; or it can be concentrated in a few newsgroups which has the advantage that members of these groups are well covered. On the other hand, if these groups are not representative of all groups on USENET, conclusions drawn from responses from the sample are not representative for all newsgroups but only for the ones which are similar to the ones chosen for the sample. Still, drawing severely the choice of sampling method towards picking only a few groups is that it is only possible to gather the observation data discussed in chapter 3.4 in a limited number of groups, as these observations require registration of all communication within a group over a period in time.

Weighting pros and cons for the two ways of sampling resulted in a decision to forego the opportunity to acquire a representative sample of newsgroups users to gain the opportunity to observe and look more closely at a few groups. This is considered important as several of the hypotheses to be tested by the empirical study are related to the ties within the newsgroup. As the topic of interest in this thesis is consumer search for information, an important selection criterion in the choice of newsgroups to sample, is that discussions in the group are often (preferable mostly) related to product choice. Research on Usenet confirmed that there are several interesting groups related to consumption. Groups are not equally serious. To get serious answers from respondents, the next criteria for selecting relevant groups was that the group should discuss in a good tone and that users mostly used what seemed like their own name, if not their own e-mail address. In some groups, this seems to be mostly the case, whereas the aliases used in other groups are far from ordinary Danish names.

From the groups remaining after this initial screening, I picked four which differ in topics discussed, have sufficient traffic (based on the numbers in Tønnesen, 2001), and which seem to have high interaction

in the group. These criteria for selecting groups are the ones Kozinets (2002) recommend. The four groups selected are:

- House and Garden (dk.fritid.hus-og-have). This is a group discussing various topics related to redecoration, smaller repair jobs in the home, and gardening. Advice sought and given is on products primarily chosen on objective cues. Also, the consumer is generally not personally closely linked to the products.
- Dog (dk.fritid.dyr.hund). This is a group discussing a more personal 'product', for which the final buying decision is often based on subjective cues. This is an almost exclusively female group – if the names used indicate the true gender of the person posting in the group.
- Cars (dk.fritid.bil). This is a group centred on a product which, for most members, seems to be a time consuming hobby. This group is different from the group on dogs in the respect that choice of product is, to a larger extent, based on objective cues. This makes advice less dependable on trust in the advisor's personal judgement. This is an almost exclusively male group – if names used indicate true gender of the person posting in the group.
- The final group chosen was Consumer Issues (dk.forbruger): This group is different from the other groups as members are discussing consumer issues and products in general. This means that members are probably not closely tied to neither the group nor the product. Also, advice given in this group is mostly on objective cues. The other group most similar to this group is expected to be the group on House and Garden.

To get a representative picture of the pattern of interaction taking place inside the four groups all postings from November 2001 and answers to the postings that were published on USENET before December 9th were downloaded, saved on my computer and registered in terms of participants in each thread. As most of the answers to a posting comes within a few days, December 9th seems to be an adequate cut-off point, even though, theoretically, there can be answers to a posting even months after it is first seen on the server. Nonetheless, for this study, a month was considered long enough. Also, the month of November was considered

to be a normal/ordinary month, and therefore appropriate for the purpose of this study although topics are, of course, influenced by time of year.

As discussed in chapter 3.4, observation data from the newsgroups only cover information exchange going on via postings and answers on the news-server and, therefore, are of only limited use in test of most of the topics in hypotheses 11-17 stated in chapter 2.6. To be able to answer the questions related to these hypotheses it is necessary to communicate directly with users of the newsgroups. The next question then, is how to sample these users.

As stated earlier, the best way to sample respondents for a survey aimed at gathering data for subsequent statistical analysis is to choose these respondents randomly from a sampling frame covering the population in question. From the observation data from the newsgroups, addresses of all members who are active in the group during a certain period of time can be harvested and used as a sampling frame for members in the group. Usability of this sampling frame is dependent on, that members, who are active in posting to the newsgroup, use their correct e-mail address in the group, and not some alias. This is often not the case, as newsgroup users often do not use their true e-mail address because newsgroups are harvested for e-mail addresses, later used for marketing purposes. Also, a sample frame built from postings to the newsgroup, does not contain all members using the newsgroup, as some members are what Wellman (1999) call 'lurkers'. These members follow discussions in the group, but are not showing themselves because they newer post any questions or answers to the group.

Another problem in using a list built from harvested e-mail addresses as a sampling frame is that such a list does not give each (posting) member of the newsgroup equal chance of being selected for the sample. This is because the list of harvested e-mail addresses is composed of the various types of visitors in the group. Members who use the group often will have a bigger chance of getting on the list than members visiting the group more seldom. However, this is not deemed to be a problem in the current study. This is because members selected during a period of suf-

ficient length can be regarded as representative for the various types of members in the group and over time, and there should be many different members visiting a few times and a few visiting many times. Members visiting a few times or only once can be regarded as representative for the many others who visit at another time than the registration period.

Yet another problem in using harvested e-mail addresses from the selected newsgroups as a sampling frame is that, as newsgroups are considered a zone for non-marketers communicating with each other, there might be a problem in harvesting e-mail addresses and using these addresses for sending e-mails containing a link to the questionnaire, as such e-mails might be considered spam. To test the usability of this sampling frame, I decided to test this way of contacting the respondents in one of the newsgroups selected for the empirical study.

In total, the request to participate in the survey was sent to 150 members of the group 'House and Garden'. This resulted in 32 usable answers from users of the group. 25 e-mails were undeliverable and another 18 responses came from users who did not want to participate in the survey. Some of senders of these refuse-to-participate e-mails were very angry, because it is not in accordance with good netiquette to harvest e-mail addresses in newsgroups to be used for sending mails directly to members of the groups.

As a consequence of this rather disappointing result, and because of the 'lurker' problem, I decided to renounce on the decision not to use self-selection. Therefore, the rest of the participants in the survey were recruited by a request to users of the remaining 3 newsgroups to participate in the survey via a posting in the selected groups in April 2002, featuring the request and a link to a tailored questionnaire for each group. To get maximum attention the link was placed in the groups at the beginning of a week-end, as this is the point in time with most visitors in the newsgroups. As the request was off-topic, I asked permission to post the request in the groups, and also for a pre-test of the questionnaire by the contact person for the group.

The consequence of the decision to recruit respondents this way is that the sample from the newsgroups is finally taken in such a way that one has to take precautions when inferring from the sample to newsgroup users in general. Also, it is a problem in the analysis of the data that the sample has been selected differently in one of the groups. This also has to be kept in mind when inferring from the results from this empirical study. However, because of the observations made in the selected newsgroups it is possible to validate the resulting data set which off-sets the problems a little.

After this description of the research design we are ready to turn to results of the empirical study. This is the topic of the following chapters.

4. Results Concerning Internet Users in general

The purpose of the data analyses carried out in this chapter is to test the Internet Search Model and hypotheses linked to this model. Analyses performed on the data from the sample among Internet users in general are carried out as described in chapter 3.2. The chapter starts with a validation and description of the sample. Validation criteria are representativeness of the sample for the Internet user, and quality of responses given in terms of non-missing responses to single items in the questionnaire. Along with this initial analysis, a short demographic profile of the sample is presented.

Next, Internet use on the general level is described and discussed in chapter 4.1.1 before we turn to the focal point of the chapter: Use of the Internet for search for consumer information in relation to an upcoming purchase. Initially, data on the different parts of the model and use of the various types of information sources are described. After this, we are ready to turn to the test of Internet Search Model, and the hypotheses about relations in the model which were stated in chapter 2.6. The empirical test of the model starts with a validation of the measurement model developed in chapter 3.3. The test of the relations in the model is then carried out after necessary adjustments have been made. The procedure and validation criteria which are used for testing the Internet Search Model are described in further detail in appendix E.

4.1 Validation and Initial Description of the Sample

The purpose of this chapter is to learn a bit about the data set to reveal whether the data are suitable for testing the hypotheses stated in chapter 2.6 on the average Internet user's use of the Internet for information search, and to get an overview of the distribution of the sample on the different parts of the Internet Search Model. In order to be suited for the analyses, the data have to be representative for the sample and there has to be some variation in the data.

4.1.1 Sampling Results

As described in chapter 3.6.1, the sample for the test of the hypotheses regarding use of the Internet for information search on the general level was selected as a close-to-random sample of telephone numbers with subsequent screening and recruitment of respondents via telephone. On the initial list of telephone numbers, there were 1673 usable telephone numbers after deletion of double numbers and company numbers. At least 3 attempts to reach each household during a period of 6 weeks in April and May 2002 resulted in contact with 1325 households. This is 79% of the original sample. Initial calls were made mostly between 4 and 8 P.M. with additional attempts to reach the households at other times of the day. Calls were made by me and 2 research assistants[16].

As only Internet Users are relevant for this study, the interview began with a screening question to determine whether the respondent was relevant for the study. Of the 1325, 637 or 48.1% claimed that they had no access to the Internet or did not use the Internet, leaving 51.9% with Internet access and actually using their Internet access. According to Statistics Denmark (2002a), 75% of the Danish population (between age 15 and 74) had access to the Internet at the time of the survey. This percentage is much higher than the 51.9% found in the screening procedure. Such a difference raises the suspicion that there might be respondents, who claim not to have access to the Internet just to escape participation in the survey without directly having to refuse to participate. However, some of the persons with Internet access according to Statistics Denmark had access only from work and some of these are probably not allowed to use the Internet for private purposes. Therefore, the 75% is larger than the part of the population actually using the Internet for private purposes. Looking closer at the statistics reveal that 12% of individuals with Internet access never use the Internet, leaving 63% who use the Internet. Another 7% use the Internet less than once a week, and some never use the Internet for private purposes. It is not possible to know the exact number having Internet access and actually using the Internet for private purposes, but with the above numbers in mind, the 51.9% found in this survey do not seem unlikely, especially as there

[16] The research assistants made about half of the calls

were persons older than 74 in the before screening sample, and Internet access is inversely related to age (Danmarks Statistik, 2002a).

After confirming that the respondent had Internet access, and also used the Internet for private purposes, we asked, if the respondent would be willing to participate in a survey via the Internet. Of the 688 individuals who had Internet access and also used the Internet 282 – or 40.9% refused to participate and 407 or 59.1% accepted to participate in the survey, and gave us their e-mail address. Obtaining the e-mail addresses was no problem. It seems quite common to have an extra e-mail address for such purposes. Using this e-mail address, the person does not really run a risk of misuse of the e-mail address.

To each of the 407 e-mail addresses obtained, a link to the questionnaire was sent shortly after we obtained the address. In appendix B.1, detailed numbers for response rates and response time are presented. In spite of several call-backs and attempts to correct the address, 5 of the e-mails sent remained undeliverable. As this is only about 1% of the e-mails sent, the number of undeliverables is not regarded as a significant problem.

From the 402 respondents who received an e-mail with a link to the questionnaire, an answer was obtained without any reminders from 154. Of these, 139 were usable. There were two main reasons why some answers were not usable: 3 did not want to participate after all, and 12 respondents had problems returning the answer, or returned an empty answer due to technical problems, mainly with their e-mail software, as described in chapter 3.6.1. As a result of these problems, it was, as also described in chapter 3.6.1, decided to use another type of return system, not dependent on the e-mail system but only on the Internet itself.

The first reminder was sent after 2 weeks. This reminder was sent to 248 respondents. The first reminder resulted in 72 usable answers plus 16 answers not usable and 5 individuals indicating that they did not wish to participate after all.

After another 2 weeks, a second reminder was sent to respondents who had not yet reacted in any way to first two mailings. This reminder was sent to 154 respondents. It resulted in 30 usable answers plus 6 respondents indicating that they did not wish to participate after all and 4 not usable answers.

After yet another 2 weeks, a 3rd reminder was sent to respondents. This reminder was sent to the remaining 103 respondents, who had not reacted to the earlier mailings. In order to make it easier for the respondent to answer[17], this last e-mail included the questionnaire in the e-mail. The result from this 3rd mailing was rather poor, as the response rate was 9.5% or 10 usable answers. In addition to this, there were 2 unusable answers and 2 respondents, indicating that they did not wish to participate after all. 9 out of the 10 usable responses were answered using the questionnaire which was included in the e-mail. This indicated that it was a good idea to embed the questionnaire in the e-mail.

In total, 251 usable questionnaires were received resulting in a response rate of 62.19% of e-mails sent. In addition to this, 34 or 8.46% of the persons, to whom a mail was sent, responded to the questionnaire, but their responses were not usable due to technical problems and 16 or 3.98% replied that they did not wish to participate, after all. 3 respondents were not sent reminders because their mail accounts in the mean time were closed.

74.63% of the mails sent resulted in a response. The response rate in total, calculated on the basis of the respondents from the original sample meeting the screening criteria, is 36.48% usable response. This is an acceptable level compared to response rates generally reported in e-mail surveys and especially compared to mail surveys (Cobanoglu, 2001; Malhotra, 1999; Schillewaert, 1998). If responses which were not usable due to technical problems are included in the number of responses, the response rate is 43.6%. Even though the response rate is far from 100%, it is at an acceptable level according to Malhotra (1999).

---

[17] And also because my technical skills had improved in the meantime so I knew how to.

One of the reasons for choosing electronic surveying was that the method has been reported to be quick compared to mail surveys. The response times in this survey were close to the results reported in research, on which the decision to use electronic surveying was based in chapter 3.6.1: Median time from mailing to response is about 3 days and the 75% percentile is reached within 1 week (appendix B.1.). In the survey in total, the median response time was around two weeks and the 75% percentile was reached around 3 weeks after the first mailing. The longer response time for the total number of answers is due to the procedure involving three reminders with a time span of two weeks between the reminders. The numbers show that especially the first, but also the second reminder, augmented the response rate, as these reminders added 17.92% and 7.46% to the response rate, respectively. The third mailing only added 2.24% to the response rate which questions the necessity of this last reminder (appendix B.1)

The conclusion from the analysis of response rate and response time in the present survey is that the collection method has proven to be quite effective in terms of speed. In terms of response rate the result is, if not brilliant, then satisfactory compared to what could be expected from the use of other methods. There were, however, problems encountered in the use of electronic surveys. These were mainly connected to technical problems in sending the response and were correctable.

Having evaluated the sample in terms of response rate, the next step is to evaluate the resulting sample. There are two elements in this evaluation: Quality of the answers and representativeness of the sample. The first is traditionally evaluated in terms of missing answers to items in responses received (Au 1994; Cobanoglu, 2001; Tse, 1998), while the second is based on comparison with data on mostly demographic profile of the population.

The first element evaluated is quality of the data in terms of non-missing answers to items. To this end, the 251 completed questionnaires were scanned for missing answers in the questions where it was possible to detect a missing response (the questions in section 1, 3, 5 and 6). In

total, 142 respondents had answered all questions. Another 47 had missed only one question. In total, 18 respondents had 10 or more unanswered questions. This is considered quite a large number compared to the 54 questions in total and indicates possible problems in wording of the questions. The 18 respondents constitute less than 10% of the dataset and on demographic variables they are very much like the rest of the data set. Because questions in the sections considered are vital in the further analysis, and the respondents do not form a not a significant part of the dataset, the 18 respondents were deleted from the dataset. In the following, the name for variables containing cases with missing values start with a s (*e.g.* s1a1) and the name for the same variables adjusted for missing values start with an e (*e.g.* e1a1). After the deletion on the 18 respondents, the final size of the sample is 233. This is the number used in most of the analyses carried out later in this chapter.

As a result of deletion of the 18 respondents, the mean number of unanswered questions went from 2.39 to 1.18 or from 4.33% to 2.18% of the 54 questions. It is especially interesting that the mean number of unanswered questions in section 5 went from 0.98 to 0.24 or from 7% to 2% of the 14 items in section 5. The remaining percentage of unanswered questions is considered sufficiently low for the dataset to be of an adequate quality for further analysis (Stanton, 1998; Tse, 1995).

In many of the analyses performed in the following chapters, the still missing answers to a few questions will pose a problem, as respondents missing an answer to just one of the questions included in the analysis are omitted from the analysis. As several of the analyses include many questions, this means that many analyses would be based on the 142 respondents, who answered all questions, if nothing is done. As there is only a small amount of missing observations in the remaining dataset, I decided to estimate the missing answers by the mean value of the answers of the respondents who had answered to each question. The consequence of this way of dealing with missing answers to items is that the mean is unchanged but the standard deviation is smaller than it is in the original dataset. This approach is normally considered appropriate when the number of missing answers is small and the number of respondents missing an answer in each question is also small (Hair, 2001; Malhotra,

1999). As this is the case here, I consider substituting missing values with the mean from the responses of the non-missing values a viable solution. Another common practice is to allocate the value in the middle – the 'neutral' value - assuming that individuals, who do not answer the question, are indifferent or neutral. However, this way of dealing with the problem pulls the mean towards the middle and away from the known average. This is, in my opinion, worse than the effect on the variance. Therefore, the first way of imputing the missing values considered is chosen as the better one of the two possibilities. It was chosen not to impute age in the 4 cases where an answer to this question was missing. Therefore, analyses carried out involving this parameter is based on 229 respondents.

In the discussion of the appropriateness of using electronic surveys in chapter 3.6, one of the issues was the lack of anonymity when using e-mail as response medium. According to the sources, I had consulted this lack of anonymity would not be a problem. This was confirmed by the responses from the screening via telephone, and in addition to this, by the fact that all respondents gave their e-mail address in the questionnaire, and 180 or as much as 77.25% of the respondents indicated that I could contact them again for follow-up questions.

4.1.2 Socio-economic Profile and Representativeness

The questionnaire included several questions on socioeconomic and demographic variables. This chapter describes distribution of the sample on these variables and tests representativeness of the sample in terms of these variables. It can, of course, be a topic of discussion whether a sample the size of 233 can be representative of the population as such. That is, whether a sample this size is sufficiently large to cover the different attitudes and lifestyles in addition to being representative on socioeconomic factors. It can be argued that the sample is small considering this kind of representativeness. Also, the sample is too small to carry out analyses on subgroups of the sample. On the other hand, the topics in this study are quite common and not exotic and therefore, I consider it safe to assume that the sample can be representative for Internet users, provided that it is representative regarding socioeconomic factors.

The sample was tested for representativeness in terms of age, family type, education, the respondent living east or west of Storebælt, and income. In order to test whether the sample was representative, is was compared with data from Statistics Denmark (Danmarks Statistik 2002a, 2002c) on the distribution of the Danish Population on the variables in question weighted by Internet access in the groups. The comparison was made, using simple Goodness-of-Fit tests. Results of the tests are reported along with the data on the various variables below. In all tests the data are checked to see if the expected number of respondents in each cell is sufficiently high for the test to be valid according to Yarnold's rule (Kristensen, 1982). If, assuming the hypothesis stating that the two groups are distributed the same way across the variable in question, the expected number of respondents in cells is too small, categories are put together in order to achieve a sufficiently large expected number of respondents in cells in the contingency tables.

The mean age for respondents is 40.72. The youngest respondent was 15 years of age and the eldest was 78 years of age. The sample is representative for the population of Internet users in Denmark regarding age considering individuals of less than 80 years of age. (p-value ≈ 0.06). It is especially interesting that also the older part of the Internet users are represented in the sample, as this part of the population tends to be underrepresented in surveys concerning aspects of Internet use. The sample is also representative for the Internet user regarding geographic distribution (p-value ≈ 0.50).

The mean number of work hours pr. week is 36.88, which is close to the 'normal' full time employment in Denmark, which is 37 hours pr. week. There are, however, large variations, as the range is from 0 to 90 with a standard deviation of 14.29. This means that respondents with very different workloads are covered in the sample making it, at least in this sense, more representative. Numbers for the 200 respondents who are the second adult in the household are very similar to persons who have answered the questionnaire.

230 respondents indicated their gender. Of these, 99 or 42.49% are women. Men and women have almost equal access to the Internet, but

more women than men with access to the Internet do not use this access at all or use the access less than once a week. Using the same approach as when evaluating the number of respondents having Internet access, it seems that the sample is representative with respect to gender. This is important as women traditionally perform a large part of the shopping in the household (Jensen, 1990; Socialforskningsinstituttet, 1997).

Regarding income, the sample is skewed compared to the population of Internet users (p-value ≈ 0.01). Incomes less than 100.000 DKK per year are underrepresented, whereas incomes between 100.000 DKK and 400.000 DKK are overrepresented. The sample is also very far from representative on education, as far too many individuals in the sample are highly educated (p-value ≈ 0.001).

The distribution of the population of Internet users and of the sample on family type is presented in table 4.1. From the table, it is clear that the sample includes far too many couples with children, and far too few singles without children. This is probably due to the initial screening procedure, as the possibility of reaching singles without children on the telephone is smaller than reaching couples with children. This is because the latter have to make dinner and attend to the children in the late afternoon, and are, therefore, more easily reachable.

Table 4.1: Civil status of families in the sample

	Percentage in sample	Percentage among Danish Internet users
Single, no children	12.45	41.63
Couple, no children	36.05	28.53
Single with children	2.15	4.72
Couple with children	47.64	24.48
Other	1.72	0.63

All in all, the conclusion on representativeness of the sample is that the sample appears not to be representative of the part of the population using the Internet but is skewed towards families with a high level of education. However, in terms of age, geographic dispersion, and probably

also workload, and gender, the sample is representative for the population of Internet users.

What is the consequence of the lack of representativeness? There are two solutions to this problem: One is to do nothing, but to be careful in the inferences drawn from the sample. Another is to weigh the sample by making the underrepresented groups weigh more heavily in the analyses and thereby making the sample representative on the weighting variable (Hair, 1998). I see two problems in the last approach: The first is that weighing assumes that individuals, who are not in the sample, would answer as their representatives in the sample have done. This is one thing we do not know anything about. The second is that with the data available, it is not possible to weigh the data on more than one variable at a time. This means that the weighing is only partially correct. Considering these two arguments, I decided not to weigh the data but to bear in mind, during the analysis that the data set is not representative of the Internet population as such but that conclusions are truer for the well educated families with children than for poorly educated singles.

4.1.3 Products Bought

In order to secure that respondents thought about a specific product, each respondent was asked to think about a product which he or she had bought during the last year, while filling in the questionnaire. It was explicitly not a prerequisite that the product had been bought on the Internet.

The distribution on products is shown in table 4.2.

Table 4.2: Product class thought about while answering the questions

Type of product	Percentage	Number of respondents
Holiday trip	26.61	62
Vacuum cleaner	3.00	7
Computer/upgrade	15.02	35
Materials for rebuilding	11.59	27
Mobile phone	9.44	22
White goods	7.73	18
Books	2.15	5
Music centre	3.43	8
Other*	18.45	43

*Examples of other things are software, TV set (2), DVD-films (3), car (2)

The, by far, most common product class to think about is a holiday trip. The second type of product, for which to think about the buying process, is a computer or upgrade of a computer, the third is materials for rebuilding, and the fourth is a mobile phone. Computers, holiday trips, and mobile phones are products, for which there is a vast amount of information on the Internet, and about which consumers, according to secondary data (GVU Gatech, 1998), search for information. Materials for rebuilding are somewhat different and are not expected to generate as much search on the Internet. As much as 18.45% indicated some other product than the ones to choose from on the list in the questionnaire.

The distribution in the table shows that the buying process for different kinds of products are represented in the sample, and therefore, the sample also covers different kinds of decision processes, and also products, for which the amount of information on the Internet is different.

4.1.4 Product Search, Interest and Knowledge

The next latent variables in the Internet Search Model are interest in the product and knowledge about the product. High levels of product interest and/or product knowledge are supposed to encourage use of the Internet when searching for product information, because the consumer

then feels more secure about being able to find and evaluate the needed information on the Internet.

The consumers in the sample consider themselves a bit more interested in and knowledgeable about the product, they bought, both compared to others in general, and compared to their friends. The numbers in table 4.3 show this, as the numbers in the bottom of the table is clearly larger than the numbers at the top.

Table 4.3: Interest in and knowledge about the product

Percentages in columns	Interest in general	Interest compared to friends	Knowledge in general	Knowledge compared to friends
Little	5.80	7.41	7.14	5.56
Some	10.71	12.04	11.16	12.96
As everybody	32.14	45.37	64.82	41.20
Quite much	31.70	25.46	31.70	27.78
A great deal	19.64	9.72	15.18	12.50

More than 50% had bought the type of product in question before three times or more (question 1.C). This confirms that the perception of being knowledgeable about the product is probably true.

In order to search for product information, the consumer has to be interested in the product and posses some knowledge about the product. Otherwise, he will either not consider it worth the effort or be too insecure to search for information. Furthermore, the decision has to be important for the consumer to be willing to invest time in search for the right choice.

The questions in section 1D covered this issue. The results are shown in the following table:

Table 4.4: Importance of the decision

Percentages in columns	It was very important to choose the right thing	It was hard to make the decision	To make the wrong decision would have and impact for a long time
The term suits not at all	0.45	22.69	15.28
The term suits not well	5.36	27.78	14.81
The term suits quite well	9.28	21.76	19.44
The term suits rather well	21.88	16.67	19.44
The term suits well	62.95	11.11	31.02

Table 4.4 shows that the majority of the respondents found the decision they were facing important in the sense that it was important to make the right choice. The distribution is more even for how hard it was to make the decision and also the impact of the decision. This is different than expected, as all product types are choice goods.

Table 4.5: Time used for information search

Percentages in column	Time used for search
I did not search at all	11.61
Less than 10 hours	60.27
Between 10 and 20 hours	13.39
Between 20 and 30 hours	4.91
Between 30 and 40 hours	3.57
More than 40 hours	6.25

Table 4.5 above shows amount of time used, in total, for search for information, before a purchase was made. Most choices were made after less than 10 hours of search. 14 respondents or 6% of the sample used more than 40 hours searching for information before they felt equipped

to make a decision. 7 of these respondents bought a holiday trip and 4 bought a computer, indicating that these types of products are considered objects, for which it is difficult to make the right decision to buy. 26 did not search at all. This was spread all over product classes. A small investigation into the relation between this variable and product knowledge, product interest and importance of the decision revealed a significant correlation between importance of the decision and amount of search but not between amount of search and the remaining two variables. However, in a regression analysis R-square was only 0.07 and therefore, the explanatory power of importance of the decision in relation to amount of search is very low. Another small investigation into the relation between time spent on search and difficulty in making the decision revealed no relation between the two. The explanation to this is probably that type of product is also of some importance when it comes to absolute amount of time, spent on search for information.

The amount of time searched is skewed. Therefore, the variable is not well suited for structural equation analysis. Already during development of the measurement model in chapter 3.3.3, this variable was suspected to pose problems, mainly because it must be very difficult for consumers to estimate exact amount of time spent for product search in retrospective. The skewness can be explained in two ways: Either the time brackets very badly chosen or respondents are not good at estimating the amount of time spent for product search. Pilot testing of the questionnaire did not show any problems concerning the scaling of this variable. However, the results here indicates that it probably would have been wiser to include more time brackets in the low level, especially taking into account that search using the Internet, or the telephone for that matter, does not require time fort transportation. Regardless of the explanation, the consequence of the large skewness of the variable, the Internet Search Model has to be tested without this direct measurement of total amount of search. This is not considered a serious problem, as it is primarily product interest and product knowledge which is supposed to enhance the use of the Internet for information search, as explained in the beginning of this chapter.

4.1.5 Internet Use on the General Level

In the Internet Search Model importance of the Internet in the person's life is supposed to influence use of the Internet directly and through a lower perceived cost of finding information on the Internet.

Importance of the Internet in the person's life was measured in terms of length of time, the person had had access to the Internet, how often, the person uses the Internet, and whether use of the Internet is mainly for work, or for private purposes. On the more subjective level, measures were importance of the Internet in the person's life, and perceived cost and benefit of using the Internet for product information search on a general level. On the more specific level, respondents were asked to indicate their use of the various types of sources of information which can be found on the Internet, in order for us to be able to conclude if respondents use the type of source for information source which they use on a regular basis for various purposes.

In table 4.6 length of time, respondents have had access to the Internet is shown. Quite a large amount of the sample – more than 75% - has had access to the Internet for more than 2 years.

Table 4.6: Time, respondents and population have been on-line

	Percentage	Number of respondents	Population, Percentage
Less than 6 months	2.62	6	4.00
6 months up to 1 year	3.05	7	
1 year to 2 years	16.59	38	10.67
2 year to 3 years	21.40	49	16.00
3 years and more	56.33	129	69.33

In table 4.3 is also shown length of time the population of Internet users in Denmark have had access to the Internet. In 2002, 75% of the Danish population had access to the Internet. In 2001, 2000 and 1999 the same number was 72%, 64% and 52%, respectively (Danmarks Statistik, 2002a). This means that 3% of the Danish population, or 4% of the part (75%) of the Danish population who are Internet users have had access

to the Internet less than a year. The rest of the numbers in the column showing 'Population, percentage' are calculated the same way. To find out, if the sample is representative in terms of length of access to the Internet, a goodness-of-fit test was carried out. The result of the test was that there, compared to the distribution of the Danish population of Internet users, are more inexperienced users in the sample than in the population as such. This is, of course, a problem in terms of representativeness. However, as most of the former research has been based on samples which were severely skewed towards skilled users, and this study tries to find results for average Internet users, it is considered to be a smaller problem that the inexperienced are overrepresented than if the opposite was true.

Table 4.7: Time respondents use the Internet pr. week

	Percentage	**Number of respondents**	**Percentage in Survey made by GVU Gatech**
Less than 1 hour pr. week	15.35	35	9.3
1 hour up to 5 hours pr. week	46.93	107	
More than 5 but less than 10 hours pr. week	19.30	44	25.4
More than 10 but less than 20 hours pr. week	10.53	24	38.1
More than 20 but less than 40 hours pr. week	6.14	14	17.9
More than 40 hours pr. week	1.75	4	9.3

Table 4.7 above this text shows the distribution of Internet use per week in the sample and in the European part of the large, widely used, survey made by GVU Gatech (GVU, 1998). The table shows that many respondents are 'light users' of the Internet using the Internet less than 5 hours per week. A cross-tabulation of the period of time the respondent has been on-line and the time spent on the Internet per week shows a weak, positive correspondence between the two variables indicating that

individuals who have been longer on the Internet also use the Internet more. Table 4.7 also shows that the distribution on Internet use in this study is very different from the distribution found in the GVU Gatech sample, with far fewer heavy users. This may be because the GVU survey was carried out using self-selection of respondents. Therefore, the result from the present sample is probably closer to the true picture of the 'average' Internet user than were the Gatech sample[18].

As search for information on products is a private matter, and use of the Internet might happen only at work or mostly at work, I also asked respondents about the distribution of their Internet use related to private matters and work, respectively. The answers are displayed in 4.8 below:

Table 4.8: Percentage of Internet use that is work-related

	Percentage	Number of respondents
Nothing at all	20.17	47
Less than 10%	20.60	48
More than 10 but less than 40 percent	15.88	37
About half – between 40 and 60%	15.45	36
More than 60 but less than 90%	19.31	45
More than 90%	5.15	12
All of it	1.29	3

Note: 2.15% did not answer this question

Up to ninety percent work-related use of the Internet, the distribution is quite even. This shows that the distribution of the time spent on the Internet for work purposes, and for private purposes it is very different across individuals. A cross-tabulation reveals a weak correspondence between time on the Internet and private use. Respondents who use the Internet very much – that is more than 40 hours per week, mostly use the Internet for private purposes, whereas individuals using the Internet between 20 and 40 hours per week use the Internet a little more for

[18] The GVU Gatech survey was made on the Internet via self-selection in 1998. Much often-cited research on consumer's use of the Internet is based on these data. Among others are the studies carried out by Hoffman and Novak.

work than for private purposes. On the other side of the continuum are very light users of the Internet. 24 of these 35 Internet users spend less than 10% of their time on the Internet for work purposes. This indicates that regarding search for product information, also individuals who spend only little time on the Internet, can be interesting, as they spend most of that time for private purposes like search for information on products.

Perceived importance of the Internet in the person's life was measured by a three-item scale with 5 categories (1 indicating lack of importance and 5 indicating importance). Mean values on the three questions were around 2 with a standard deviation of about 1, indicating that the persons in the study generally do not find the Internet of great importance in their lives, but with differences in strength of importance (the exact numbers can be found in appendix C.1).

In order to find out how the respondents in the sample use the Internet, section 4 in the questionnaire contained questions on this matter (the questionnaire is printed in appendix A.1). Table 4.9 below shows the distribution in use of the relevant types of sources available on the Internet ordered by level of use.

Table 4.9: Use of Internet opportunities in general

Percentage in rows	Every day	At least once pr. week	At least once pr. fort-night	At least once pr. month	More seldom	Never
Search engines	20.6	43.35	18.03	7.73	9.01	1.29
Websites, producers	6.01	21.03	15.45	16.31	24.03	15.02
Consumer information sites	0.86	10.73	12.88	19.74	35.62	17.60
Websites, retailers	0.86	10.30	8.58	18.45	32.19	28.33
Newsgroups	9.44	18.03	7.73	9.87	21.40	31.33
Chat	1.29	3.86	3.00	1.29	5.15	81.55

Note: Other sources mentioned were banking and other financial sources, health information and download of shareware.

The numbers in table 4.9 shows that search engines are used regularly by most Internet users. 89.3% use a search engine at least once pr. month and as much as 20.6% of the sample use search engines every day. The difference to usage level of the second most used kind of offer is rather large. Websites from producers and retailers and general consumer information sites are used by approximately half of the respondents at least once a month. The distribution on frequency in use of these to marketer-dominated information sources differ. Consumers who use producer's websites visit these websites more regularly retailer's websites and consumer information websites.

About half of the respondents use newsgroups at least once a month and at the same time, one third never use newsgroups. The table shows that as much as 9.4% of the sample uses newsgroups every day, and another 18.03% use newsgroups at least once pr. week. It is my impression that this is quite a large percentage compared to other statistics on the use of newsgroups. I, therefore, ask myself if some of the respondents misunderstood this term and thought that they were asked if they read

derstood this term and thought that they were asked if they read regular news on the Internet. Chat, both commercially sponsored and not commercially sponsored, is the source on the Internet used the least by the respondents in the sample, confirming that this type of source is not very interesting to look further into when investigating the behaviour of the 'ordinary' Internet user. The picture is probably quite different in a sample of young individuals.

The final latent variable expected to relate to importance of the Internet in the person's life is Locus of control. Mean values for the three variables measuring this latent variable are displayed in table 4.10.

Table 4.10: Locus of control

Item	Name	Mean (stddev)
What happens is my own doing	E621	4.12 (0.85)
When I make plans, I can make them work	E622	3.47 (0.94)
I have an impact on how others are to me	E624	4.22 (0.74)

The respondents tend to agree on all three items, showing that they have high internal rather than external locus of control (1 indicates total disagreement and 5 indicates total agreement to the item). This is in accordance with findings of Wolfinbarger (2000) that individuals using the Internet generally have high internal locus of control. The standard deviation on this measure is rather low, and because of this, this latent variable may prove not to be useful in analysis of determinants of use of the Internet for information search, after all.

4.1.6 Use of the Internet for Information Search

Use of the Internet for search for information prior to buying was measured in two ways: Directly and by questions about use of the different types of sources on the Internet. In addition to questions about amount of search on each type of information source, respondents were asked to rate on a scale from 1 to 5 for each type of source to what extent they

found material and if the material found was useful in their decision process.

Respondents were asked to rate use of the Internet in general relative to sources not on the Internet and to estimate use of each of the different types of information sources on the Internet on a scale from 1 to 5. In order to make the following text and table easier to read, I interpret a 1 as ‘very little’, 2 as ‘a little’, 3 as ‘some’, 4 as ‘much’ and 5 as ‘very much’.

Table 4.11 shows distribution on time spent relatively on search on the Internet in general and how much each of the various sources was used. The order of sources in this table is the same as the order in table 4.9 containing use of the sources in general, expect for use of Internet for information search on a general level, which is in the first row. This means that the type of source most commonly used in general is in the second row in this table, the source used the second most in the third row, and so on.

Table 4.11: Use of the various types of sources on the Internet

	Not at all Percentage of all responses	**Percentages in rows based on users of source**				
		Very little	**A little**	**Some**	**Much**	**Very Much**
The Internet in total	32.19	27.85	22.79	19.62	13.29	16.46
General search engines	49.79	35.05	19.65	13.68	13.68	17.94
Specialized search engines	64.38	49.41	9.63	13.25	14.46	13.25
Websites, producers	63.95	51.18	23.80	13.09	4.77	7.16
Consumer information sites, sponsored by government	76.82	59.26	16.66	14.80	7.42	1.86
Consumer information sites, other	81.12	88.62	4.55	2.28	2.28	2.28
Websites, retailers	67.81	50.67	17.33	14.66	8.01	9.32
Newsgroups	78.87	61.22	16.30	12.26	8.17	2.04
Chat – commercial sponsors	84.98	85.70	0.00	2.86	5.72	5.72
Chat – other	86.27	90.61	0.00	6.26	0.00	3.13

Table 4.11 shows that 32.19% of the respondents did not spend any of the time used for information search prior to the purchase decision on the Internet. In contrast to this, 20% of the total, or about one third of the respondents who used the Internet for information search, used the Internet much or very much, compared to off-line sources. A Chi-square test on a cross-tabulation of proportion of time spent searching on the Internet and time spent on search in total reveals no dependency between these two variables confirming that absolute amount of time spent

on search should not be a part of the Internet Search Model, as was already decided in the former chapter.

In general, types of sources used much in general, are also the sources used when searching for specific information in relation to an upcoming purchase. An exception to this is retailers' websites, which seem to be used more in connection to this task than otherwise. This is probably because some of the respondents actually bought their product on the Internet. Cross-tabulations of the variables in this table with the corresponding variables in table 4.9 show that the connection between everyday use and use for information search also seems to hold on the individual level, except for the use of newsgroups.

Most types of sources on the Internet have been used only a little. It is clear from table 4.11 that the only source which has been used much is search agents. Producers' websites also seem to be quite popular, compared to the other sources whereas consumer information sources are less popular than the other non-individualized sources, even though this type of source is not marketer dominated. The reason for this could be that consumer information sites are not well-known to consumers compared to brand names of consumer goods, as these are far more heavily marketed than consumer information sites are.

An important question to be answered from the data from the sample on Internet users in general is level of usage of newsgroups for finding consumer information. In this part of the questionnaire, there was no doubt about the meaning of the term, as there was an explicit mentioning of USENET. 21.13% of the respondents used newsgroups for finding product information. Among the respondents who used newsgroups quite much for information search, some used the newsgroup quite regularly, while another group used the newsgroup only seldom. This is in accordance with Kozinet's framework from chapter 2.4 and the hypotheses stated in relation to this: Newsgroups are used for information search by two types of members: Tourists who visit the newsgroup only seldom, and Insiders visiting the newsgroup more often. Looking at table 4.11, it is clear that the largest part of respondents have used newsgroups only a little. However, compared to the other sources quite a

large percentage has used newsgroups much while searching for information.

Chat has been used for search for information by 15% of the respondents. Only a small part of chat users have used these more than a little in their search for information. As the statistical material is very small, it is dangerous to conclude more on these numbers. Still, I looked into the connection between use of chat in general and use of chat rooms for search for product information. The picture seems to resemble the picture of the use of newsgroups: Some of those who used chat a lot for information search in relation to the decision in question also use chat on a regular basis, whereas others also using chat much for search for product information do not normally use chat on the Internet.

In short, it seems that consumers tend to use non-individualized marketer-dominated types of sources on the Internet that they regularly use for various purposes when searching for product information in relation to an upcoming purchase. The most popular type of information source is search engines and as many as one third of the respondents used producer's and retailer's websites in their search for information. A large part of these are regular users of this kind of information source.

Individualized, non-marketer dominated sources on the Internet, (chat and newsgroups) are, as expected from the secondary data, used more seldom. Also, as expected from hypothesis 11, newsgroups are more used for search for product information than chat is. Different individuals use both types of interaction in two different ways. Some individuals use chat or newsgroups regularly and also for search for product information, whereas others use chat or newsgroups only when searching for information and not on a regular basis.

Respondents were asked to rate, not only their use of the various sources on the Internet, but also how much material they found on the various sources and whether the material was helpful in reaching a decision on what to buy. The variables measuring this and the variables measuring use of the various sources are highly correlated. This means that consumers who search for information on the Internet find material which

helps in making a decision at the types of sources where they search the most.

Of interest is also whether there is a relationship between relative use of the Internet for search for information and the type of product bought, as the products were expected to differ in this manner as explained in chapter 2.5.3. A chi-square test revealed that there is a significant dependency between the two variables. The numbers are displayed in table 4.12.

Table 4.12: Product bought and relative use of the Internet for search for information

Use of Internet for search Number	None	A little	Some	Much/very much	Total (%)
A holiday trip	5	27	14	16	**62 (27.31)**
Other	23	20	9	11	**63 (27.75)**
A computer/hardware	11	7	3	14	**35 (15.42)**
Materials for rebuilding	13	12	2	0	**27 (11.89)**
Mobile phone	9	7	2	4	**22 (9.69)**
White goods	10	6	1	1	**18 (7.93)**
Total (%)	**71 (31.28)**	**79 (34.80)**	**31 (13.65)**	**46 (20.26)**	**227 (100)**

Note: Some cells have expected values less than three. However, using Yarnold's rule (Kristensen, 1982) the minimum expectation is 30/24 and therefore, the test is valid. Chi-square is 43.87 and the P-value is 0.0001. Some products have been added to the 'Other' group due to few observations.

The cells deviating from the hypothesis stating independence of the variables are the following: For holiday trips (a feel product, chosen on a blend of objective and subjective cues) consumers search the Internet more than expected. However, they do not necessarily use the Internet

more than other sources. Computers and Hardware, (choice based on objective cues) is, contrary to holiday trips, a type of product on which a larger part of consumers, who do use the Internet for information search, use the Internet much more than other sources. White goods and materials for rebuilding (choice based on objective cues) seem to be products for which consumers tend not to search the Internet for information before purchase, and when they do, it is only on a limited scale. This supports the notion that consumers search the Internet for information on products, for which they think they will find the information, they need.

In the Internet Search Model, belief in availability and credibility of the information on the Internet are hypothesized to be important determinants of perception of cost of finding the needed information on the Internet. The scales used for measuring these latent variables ranged from total agreement (5) to total disagreement (1) as described in chapter 3.6.

The first latent variable is 'Availability of information on the Internet'. The latent variable is measured by the four variables shown in table 4.13. As a large value indicates agreement, respondents are, on average, positive towards the availability of information on the Internet. They are more positive in their expectation towards there being information on prices than on products. The standard deviation is rather large compared to the values indicating that there are differences between individuals.

Table 4.13: Attitude towards availability of information

Item	Name	Mean (stddev)
There is much helpful information on the Internet	E51	3.53 (0.99)
It is easy to find information on the Internet	E52	3.68 (0.99)
There is good advice on products on the Internet	E53	3.24 (1.05)
There is good information on prices on the Internet	E54	3.68 (1.08)

1 indicates total disagreement, 5 indicates total agreement

Credibility of information on the Internet is also assumed to be an important factor in perception of usability of the information on the Internet. Mean values for the variables measuring this latent variable are displayed in table 4.14 below. On the three first variables, consumers, on average, are a little more positive than neutral. On the fourth variable measuring lack of predisposition consumers are on the negative side of the average value, 3, indicating that consumers believe that the source influences information given. In addition to this the measures are, on average, lower than measures of the belief in availability of the information on the Internet presented in table 4.13 indicating that consumers believe that the information is there, but that one has to be careful when using the information. Like for the variables measuring availability of information, the size of the standard deviation indicates differences between respondents, but on a lower level.

Table 4.14: Attitude towards credibility of information

Item	Name	Mean (stddev)
The information on the Internet is generally sincere	E55	3.29 (0.89)
The information on the Internet is generally honest	E56	3.11 (0.81)
The information on the Internet is generally dependable	E57	3.15 (0.81)
The information on the Internet is not predisposed	E58	2.70 (0.90)

1 indicates total disagreement, 5 indicates total agreement

Perceived cost of finding information is normally expected to influence use of information sources available to the consumer because the consumer is supposed to act rationally. The cost latent variable is measured by perceived time, money, and the effort which has to be put into the process, supplemented by two variables measuring overall evaluation of relative cost and benefit from using the Internet for information search along with weight variables for the cost elements. Results for the separate cost elements are shown in table 4.15 below.

Table 4.15: Perception of cost of using the Internet for information search

Item	Name	Mean (stddev)	Average importance
It is cheaper to search for information on the Internet than elsewhere	E59	3.28 (1.15)	3.97 (1.19)
It is faster to search for information on the Internet than elsewhere	E10*	3.87 (1.02)	4.24 (0.92)
It is easier to find information on the Internet than elsewhere	E511	3.75 (1.05)	4.62 (0.69)
How much more do you think you gain from searching the Internet for information compared to other places?	E3e	3.85 (0.85)	
How demanding do you find it is to search for information on the Internet than elsewhere? (scaled opposite)	E3f	3.53 (0.96)	

A high value is the same as low perceived cost, high importance, large gain, and not demanding

Table 4.15 shows that respondents generally agree that it is faster, cheaper and easier to find information on the Internet than off-line. However, perception of the level of each cost element is only one aspect in the attitude towards using a source for information. That a cost is perceived as high is irrelevant, if the level of the cost is unimportant to the consumer. In chapter 2.5.2, this issue was discussed in detail. To find the perceived cost to the respondents the elements are, therefore, when used in the structural equations, weighted by their respective importance to each consumer. The results in table 4.15 show that consumers, on average, place quite high importance on all cost aspects, as the mean value is around 4 on a scale from 1 to 5. Especially convenience seems to be

important, with quite a small difference among the individuals in the sample. Cost in money is seen as the least important issue with the largest difference among respondents. Detailed results regarding the two variables measuring cost and benefit on a general level are shown in appendix C.1. The response to the two questions shows that respondents in the sample, on average, find that they gain relatively much from searching the Internet compared to other places and also that the Internet is less demanding than other possible information sources indicating an, on average, positive attitude towards using the Internet for search for information.

The measures of perceived cost show that respondents in the sample on average find that using the Internet is less costly than using off-line sources. On the most important cost element, ease of finding the information, the Internet outperforms off-lines sources in respondents' minds. As there are differences among the individuals, the interesting issue then is whether these differences in perceived cost of using the Internet for information search can explain differences in use of the Internet for this purpose, as expected in the Internet Search Model.

4.1.7 Results on Facilitating Factors

The first facilitator in the Internet Search Model is Innovativeness in Shopping. The concept is believed to be positively correlated to perceived cost of using the Internet for information search. The concept is covered by two questions, and results are shown in the table below:

Table 4.16: Innovativeness in shopping

Item	Name	Mean (stddev)
When a new store opens, I am among the first to try it	E610	1.71 (0.86)
I like to try new places to shop	E611	2.97(1.18)

1 indicates total disagreement, 5 indicates total agreement

On average, respondents in the sample do not perceive themselves as innovative in shopping. They are more prone to try new places to shop

than to try new stores. This indicates that shopping behaviour may not be easily changed.

The second facilitator is 'In-home Shopper'. The hypothesis is that if the consumer is positive towards, and maybe used to, shopping from home, then the needed change in behaviour is smaller than if the consumer is an alien to home-shopping. Therefore, perceived cost of using the Internet for information search is expected to be lower, the more the consumer is inclined to in-home shopping.

Table 4.17: Inclination to in-home shopping

Item	Name	Mean (stddev)
Shopping at home saves me time and effort	E613	3.27 (1.27)
Shopping from home is more convenient	E614	2.81 (1.34)
I like to shop from home	E615*	2.83 (1.44)
It is saves money to shop from home	E616*	2.80 (1.33)
From home-shopping saves time	E617	3.53 (1.17)
I shop from home because I can't find what I want at local stores	E618	2.41 (1.42)
I love to browse through catalogues	E619	3.06 (1.28)

Note: s615 and s 616 was scaled opposite on the questionnaire.
In this table, 1 indicates total disagreement and 5 indicates total agreement

A glance at table 4.17 reveals that this scale may be measuring two separate aspects of attitude towards in-home shopping: perception of cost, and feeling or attitude. On average, respondents tend to agree that shopping from home saves time and effort (e613), and also agree to the other question, covering the time issue (e617). Respondents, on the other hand, tend to disagree on the 'feeling' statements: That they like to shop from home (e615), that it saves money to shop from home (e616), and that shopping from home is more convenient (e614). Furthermore,

on average, respondents disagree that they shop from home because of the selection in local stores (e618). There are quite large standard deviations indicating that there are differences among the consumers. Therefore, it is interesting to test whether these differences are explanatory in terms of use of the Internet for information search as proposed in the Internet Search Model.

The last variable expected to act as a facilitator, lowering the perceived cost of using the Internet for information search, is 'Role Overload', as perceived role overload makes use of the Internet more advantageous as explained in chapter 2.4.

Table 4.18: Role Overload

Item	Name	Mean stddev)
I often feel I have to do everything partly because of time pressure.	E620	2.90 (1.25)
I am way to busy in my everyday life	E623	3.22 (1.17)

1 indicates total disagreement, 5 indicates total agreement

As indicated in chapter 3.3 on selection of measurement scales, there was uncertainty whether the items in this scale were well suited for measuring 'Role Overload'. Therefore, a test of the relation between work hours, number of children, gender, and age, and felt role overload was carried out. As stated in chapter 3.3.8, the relations – maybe except age – should be positive if the measure is valid. A multiple regression showed a positive significant relation between 'Role Overload' measured on a summated scale, and work hours per week. There were also a close-to-significant positive relation between number of children and felt role overload. The relation between age and felt role overload was negative and significant at the 5% level. There were no differences for the gender variable. R-square was only 0.12 which indicates that there are many other factors explaining felt role overload than the ones found in this regression analysis. However, relations found were as expected, and therefore confirm the validity of the measurement scale used.

4.1.8 Attitudes towards Shopping

The final latent variable in the Internet Search Model is attitude towards shopping. Attitude towards shopping is hypothesized to have an off-setting effect on use of Internet for information search in relation to an upcoming purchase. This is because external search for information can be seen as an opportunity, or even an excuse, to go shopping. Therefore, search for information is, to some, entertainment and not a burden.

The first aspect of enjoyment in shopping is shopping seen as a social activity. This latent variable was measured using the seven items shown below in table 4.19. The variable is hypothesized to correlate negatively with 'Use of the Internet for Information Search'.

Table 4.19: Shopping enjoyment

Item	Name	Mean (stddev)
Shopping gives me a chance to get out and do something	E63	2.53 (1.31)
I like to go shopping with other people	E64	2.96(1.35)
Shopping is a topic of conversation	E65*	3.44(1.20)
I like to go to regional shopping centres	E66	3.61 (1.17)
I often combine shopping with lunch or dinner at a restaurant	E67	2.21 (1.17)
I get a psychological lift from shopping	E68	2.48 (1.16)
I am interested in shopping	E612	2.54 (1.27)

Note: e65 is scaled opposite on the questionnaire. Otherwise, 1 indicates total disagreement, 5 indicates total agreement

The respondents generally agree that they talk about shopping and that they like to go shopping at their regional shopping centre. On average, they are neutral towards the notion of shopping with other people. However, the standard deviation is rather large indicating that the individuals participating in the survey differ in this respect. Respondents disagree on average that shopping is something offering a chance to get out, give a psychological lift, and that shopping is interesting. Furthermore, they do not often combine shopping whit a meal on a restaurant. On all variables, the standard deviation is rather large indicating differences among

the individuals with respect to attitude to the variable. It is, therefore, interesting to see if these differences are determining use of the Internet for information search, as hypothesized in the Internet Search Model.

Like the measurement scale for inclination to in-home shopping which also included quite many items, this measurement scale seems to be measuring several different things: First, the social entertainment aspect, on which the level is quite similar (e63, e64, e67, e68 and e612) and second, the conversation and local aspect (e65, e66) on which the level is much higher than for the rest of the items. Especially the item 'I like to go to regional shopping centres' may prove to pose problems, because it may have been understood as if the regional part was the topic of interest. This is because malls and shopping centres are much less common in Denmark than in the USA, and the debate about retail structure in Denmark often centres on the ability to shop locally.

The second variable related to shopping enjoyment is importance of contact with sales staff. This concept is measured by the answers to the three items shown below in table 4.20.

Table 4.20: Importance of contact with sales personnel

Item	Name	Mean (stddev)
I like to shop where people know me	E61	3.69 (1.16)
I like to shop where clerks know my name	E62	2.85(1.34)
I like to get to know the clerks in the store where I shop	E69	2.26(1.18)

1 indicates total disagreement, 5 indicates total agreement

There is a quite high average value on the first variable indicating that, on average, respondents like to shop where the other persons know them. They are a lot less concerned whether clerks know their name. Actually, the average value indicates that quite a large number of respondents do not like clerks to know their name. The rather large standard deviation indicates that there are differences between respondents. The answers to the last item indicate whether the consumer likes to know clerks he or she deals with. The average value indicates that re-

spondents in the study do not try to get to know clerks where they shop. There are, however also differences in opinion on this question as on other questions. The hypothesis related to the Internet Search Model is that a high value on this variable is related to relative little use of the Internet.

4.1.9 Summing Up Initial Description and Validation

The purpose of this chapter was to investigate the data set to find out if the data are suitable for testing the hypotheses stated in chapter 2.6 on the average Internet user's use of the Internet for information search and to get an overview of the contents of the data set concerning the different parts of the Internet Search Model.

The first issue was to test the quality of the data set and to make the data suitable for statistical analysis. The response rate, calculated as percentage of persons meeting the screening criteria, is 43.6%, including all mails received, and 36.48%, including only usable responses. This percentage is far from 100% and lower than what is the ideal for a study of this kind. Nonetheless, the response rate is judged to be at an acceptable level for further analysis because it is at a level normally deemed acceptable as explained in chapter 4.1.1. The data set was tested for representativity for the population of Internet users. Tests for representativity showed that the data are representative for the population in respect of age, gender, geographic dispersion, and probably also workload. However, in the data set, well educated families with children are overrepresented. It is important to bear this in mind when concluding on the analyses carried out in the following chapters.

The descriptive measures of the data showed that there is variation in the data set on all relevant variables. Buying processes for all the products represented in the starting question in the questionnaire is represented in the sample, as is also search processes of different lengths. In the sample, different types of Internet users are also well represented. Especially worth noticing is that there is a fair amount of unskilled users, which makes this sample different from most of the samples, on which previous empirical research has been done.

The Internet users in the sample show different levels of Internet use, both in general and when searching for information in relation to an upcoming purchase. More than two thirds of the respondents in the sample used the Internet for information search in the selected buying process. The Internet was used the most when searching for information in relation to vacations and computers and equipment for computers, and the least for white goods and materials for rebuilding. This is not surprising, as the two products, for which the Internet was used much for search for information, are also products, on which the Internet contains much information.

When searching for information, consumers use the types of sources, they use at a regular basis for other purposes, as well. The consequence of this is that type of source used the most for information search is search engines. Consumers who visit producers' websites on a regular basis also use this type of source when searching for information in relation to a specific purchase. Retailers' websites are more commonly used for search for information in relation to an upcoming purchase than on a general level. This is natural, as some respondents probably visited some of these places because they wanted to make their purchase there. The data also showed that consumer information sites are not very popular, neither in general nor when searching for information with a specific purchase in mind. As expected, newsgroups were used more than chat-rooms for information search. Newsgroups had been used by 21.13% of the sample. Of these, about 22% had used the newsgroup much or some. This confirms that the newsgroup is interesting to study further, although it still is not a source, used by a substantial part of the ordinary Internet users.

When focusing on variables expected to explain use of the Internet for information search, descriptive measures showed that respondents, on average, find that it is convenient to find information on the Internet, and that the needed information is perceived as available and fairly credible. The respondents are, on average, not inclined to try new places to shop. They find that shopping from home saves time and effort, but not money. The consumers in the study have high internal locus of control and they often feel time pressure even though they do not agree to

be too busy in their everyday lives. These figures are not surprising remembering that Internet users tend to have a higher level of internal locus of control than do the average person, and in addition, this may also be true for the overrepresented busy working family with children, as is the feeling of role overload.

Inclination to shopping is supposed to be off-setting for use of the Internet for information search. The respondents in the sample like to go to regional shopping centres and to talk about shopping, but, on average, do not regard shopping as social entertainment giving a chance to get out. They do not wish for sales personnel to know their name but they like to shop where people know them. There were large differences in the sample with respect to the answers to these questions confirming that individuals differ with respect to the attitude towards shopping.

All in all, this initial analysis showed that the data are suitable for further analysis, as there are variations across respondents in the sample. Also, the analysis confirmed that newsgroups are used by a part of the sample large enough for this type of virtual community to be studied further with respect to use for information search. Before turning to this part of the study, however, there is a need to look further into the relations on the general level in the form of a test of the paths in the Internet Search Model. The prerequisites for such a test are discussed in the following chapter.

4.2 Validation of the Measurement Model

It is customary to test the measurement model before testing the structural equation model. The procedure followed in the testing of the model is described in detail in appendix D. Normally the test of the measurement model is done in two stages: An initial, partial test of each of the scales measuring the latent variables in the model followed by a confirmatory analysis testing the full measurement model. The first test ensures that the measurement scales are adequate for use in confirmatory factor analysis whereas the latter tests if there are any relations between the latent variables in the model and most of all that the observed variables load on one, and only one, unobserved – or latent - variable. In

the following, the additions ‘observed’ and ‘latent’ are used whenever there is possibility of confounding the type of variable.

4.2.1 Initial Validation of the Measurement Scales

In appendix C.1 the mean, standard deviation, and skewness for each variable is listed along with the variable name used in the data set. Also, reliability measures estimated by Cronbach’s Alpha are shown in the appendix for all scales.

The central variable in the Internet Search Model is Use of the Internet for Information Search. This latent variable is measured by the variables s20a (Internet in general) and s20astd (summated use of sources). The latter has been standardized to the same mean as s20a. S20astd has skewness on the large side (1.18) and is therefore a bit problematic. However, the skewness is in the same direction as for s20a and the variable is necessary for the identification of the model. The reliability of the scale is 0.63. This is not ideal, but acceptable. Therefore, this scale is used in the analysis in spite of the encountered problems.

The first variable explaining ‘Use of the Internet for Information Search’ is Importance of the Internet in the Person’s Life. This latent variable is measured by the variables e3d1, d2 and d3. None of the variables are skewed and reliability of the scale is 0.74, which is acceptable. Internal Locus of Control is hypothesized to enhance importance of the Internet in a person’s life, and is measured by the variables e621, e622 and e624. One variable (e621) has a skewness which is slightly too high. This skewness is in the same direction as the smaller skewness calculated for the other variables in the scale, and is therefore not considered a problem. However, reliability of the scale is 0.48. This is far too low for the latent variable measured by this scale to be included in the analysis. This is not considered a serious problem, as Internal Locus of Control is not considered central in understanding of use of the Internet for search for information.

The second group of variables in the model are variables measuring interest in and knowledge about the product and importance of the decision. These are the variables at the top of appendix C.1. The variables

measuring product interest and product knowledge do not pose any problem, as they show both low skewness and high reliability of the scales. Contrary to this, the scale measuring 'Importance of the Decision' is very problematic for use in structural equation modelling. One of the variables have very large skewness, the skewnesses of the variables are in different directions, and the reliability of the scale is low (0.38). Attempts were subsequently made to leave out one of the three variables at a time. The result of this was not uplifting, as the reliability measure was low in all cases. Therefore, 'Importance of the Decision' has to be left out of the analysis. This is not considered a serious problem for testing the Internet Search Model as such, 'Importance of the Decision' is not central in the analysis, and the variables measuring interest in and knowledge about the product capture at least part of the contents of 'Importance of the Decision'.

Availability and Credibility of Information on the Internet are measured by the variables e51-e54 and e55-e58, respectively. Neither of these scales poses any problems.

The scale measuring perceived cost of using the Internet has an Alpha value of 0.72 for the standardized scale. This is an acceptable level. All variables exhibit low values of skewness. Therefore, it is not considered an issue worth pursuing that skewness for the price variable is in opposite direction of skewnesses for the other variables.

The facilitating factors for use of the Internet for information search are innovativeness in shopping, being inclined to in-home shopping and felt role overload. 'Innovativeness in Shopping' is measured by variables e610 and e611, while 'In-home Shopping' is measured by variables e613-19. 'Role Overload' is measured by variables e620 and e623. None of the variables exhibit problematic sizes of skewness. Reliability of 'Innovativeness in Shopping' is 0.57, which is low. Therefore, it has to be considered if 'Innovativeness in Shopping' should be left out of the model in the further analysis. In favour of this speaks that the scale is, in reality, not a scale of its own, but was taken out of the scale measuring shopping enjoyment. Therefore, the foundation behind the variables in the scale is limited. Reliability of 'In-home Shopper' is 0.66,

which is on the low side, but acceptable, and reliability of 'Role Overload' is 0.79, which is fine. Still, as discussed in chapter 4.1.6, the scales seem to be measuring more than one thing. This is supported by skewnesses in different directions in same pattern, as was seen in chapter 4.1.6.

The final variable in the Internet Search Model is 'Shopping Enjoyment'. Shopping enjoyment is determined by two latent variables: 'Shopping Enjoyment' (e63, e64, e65, e66, e67, e68 and e612); and 'Enjoyment in Contact with Sales Staff' (e61, e62 and e69). Both scales have adequate levels of reliability, but as discussed in chapter 4.1.7, 'Shopping Enjoyment' seems to be measuring more than one thing. This is supported by the fact that skewnesses are in different directions in the same patterns as was seen in chapter 4.1.7, and adjustment of the scale has, therefore, to be considered. In consequence of this, reliability of the scale without e65 and e66 was calculated, resulting in a rise in reliability to 0.78.

In order to get ideas on what to do about problems encountered in the previous, the next step is to perform factor analysis to see if the variables belong to the scale they are expected to belong to and, therefore, also measure the latent variable they are supposed to measure.

As different parts of the data cover different, but interrelated, latent variables, factor analysis was done on each section separately. As sections 1, 2 and 3 cover general aspects, factor analysis on answers to the questions in these sections were performed in the same analysis.

The factor analyses were all performed using the mineigen criterion (mineigen=1) supplemented by a scree plot and varimax rotation. Loadings are considered large when they are numerically larger than 0.6 (Malhotra, 1999).

Factor analysis on background variables and variables on use of the Internet results in retention of 6 factors. The first factor, accounting for 21.5% of the variation in the data, contains the variables measuring product interest and product knowledge, indicating that there is not good

discrimination between these latent variables. The second factor, accounting for 13.43% of the variation in the data, contains the variables measuring the cost elements. The third factor, accounting for 9.65% of the variation in the data, contains the variables measuring importance of the Internet to the respondent (the third variable with a loading of 0.51). The fourth factor, accounting for 8.55% of the variation in the data, contains the two variables measuring 'Use of the Internet for Information Search'. The fifth factor, accounting for 7.22% of the variation in the data, contains two of the variables measuring importance of the decision. The last variable in this scale is in the sixth factor which accounts for 5.85% of the variation in the data, confirming that the scale measuring importance of the product/decision is problematic.

This factor analysis show that variables which are supposed to correlate, do correlate and mostly, they do not correlate with other variables indicating a rather good discrimination between the variables measuring the different latent variables. The exception is the scales measuring interest in and knowledge about the product. As a consequence of this, it is a possibility worth considering, putting variables measuring product knowledge and product interest into one variable, as they are in the same factor in this analysis. Also, the factor analysis confirmed that the scale measuring importance of the decision is problematic as the variables did not load on the same factor.

The second factor analysis was performed on the variables connected to the items in section 5 in the questionnaire on attitude towards the Internet. The analysis resulted in retention of 2 factors. The first factor accounts for 35.25% of the variation in the data, and contains the variables measuring credibility of information on the Internet. The second factor accounts for 31.08% of the variation in the data and contains variables measuring perceived availability of information on the Internet. The factor analysis produced the result which was expected in the sense that variables that are supposed to correlate, do correlate, and variables which should not correlate, do not. This means that both reliability and discriminate validity of the scales used seem to be in order.

The third factor analysis was performed on the variables connected to the items in section 6 in the questionnaire measuring various characteristics of the consumer. The results from the initial analysis indicated that there may be possibilities for improvement of the scales used in this section. There were 24 variables from this part of the questionnaire, resulting in the retention of 8 factors. The first factor accounts for 15.04% of the variation and contains variables e63, e64, e67, e68, e610, e611 and e612. Of these 7 variables, 5 are in the scale for enjoyment in shopping, whereas the last two variables are supposed to measure innovativeness in shopping. This indicates that respondents feel that part of the enjoyment in shopping is 'shopping for shops', as one might phrase it. These variables were, in fact, together in the original scale but were separated because they were believed to measure different latent variables. Finally, a word of caution is needed, as two of the variables have loading less than 0.6.

The second factor retained accounted for 10.29% of the variation in the data. This factor contains three of the variables measuring in-home shopping. The three variables all relate to the cost (in terms of time and lost convenience) items of 'In-home Shopping'. The third factor retained accounted for 9.83% of the variation in the data. This factor contains the variables measuring importance of contact with sales staff. The fourth factor retained accounted for 7.21% of the variation in the data. This factor contains the two variables measuring 'Role Overload'. The fifth factor retained accounted for 6.83% of the variation in the data. This factor contains two of the three variables measuring locus of control. The variables in the factor are the variables related to ability to plan one's life.

The sixth factor retained accounted for 5.63% of the variation in the data. This factor contains the two variables measuring general attitude towards in-home shopping and the perceived cost of in-home shopping. The seventh factor retained accounted for 5.21% of the variation in the data. This factor contains only the variable measuring whether the respondent likes to browse through catalogues. The eighth factor retained accounted for 4.96% of the variation in the data. This factor contains the two remaining variables from the scale measuring shopping enjoyment

(discussions on shopping and whether the respondent likes to shop locally); and one variable belonging to the scale measuring locus of control: Influence on how others behave towards me. The loadings on this factor are rather low, which is not surprising, as the variables measure different latent variables. As the variable measuring liking to browse through catalogues is alone in the factor analysis and is, therefore, not connected to the other measures, this variable is left out of the scale along with variable e616 which also posed problems. When these two variables are left out from the scale, the measure of reliability rises from 0.66 to 0.75. This is a very satisfying improvement.

A factor analysis of all the variables at one time shows almost exactly the same picture as was seen in the partial analyses indicating that there are no problematic correlations across the sections in the questionnaire.

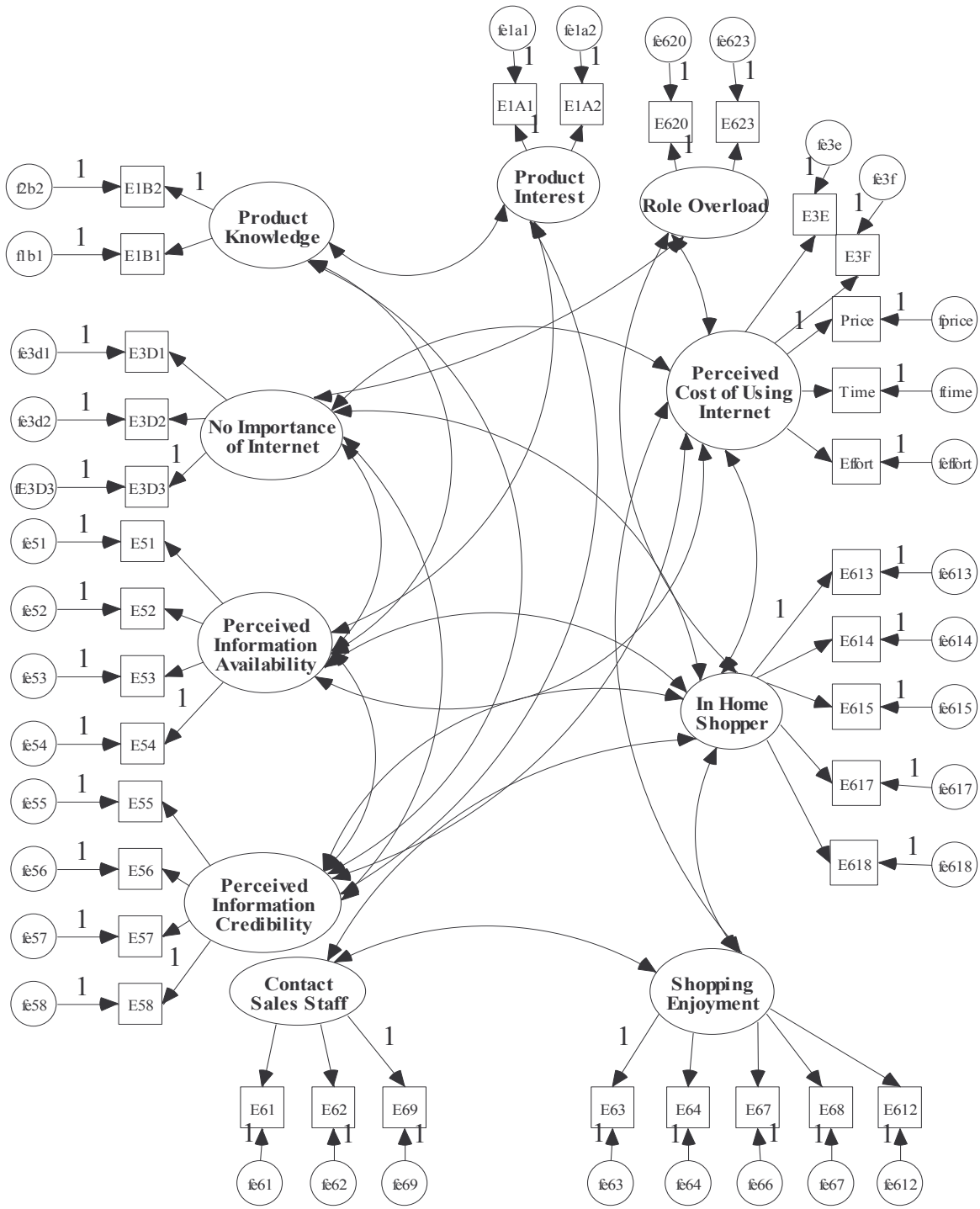
The conclusion to this partial analysis of the variables in the measurement model is that the latent variables 'Locus of Control', 'Importance of the Decision', and 'Innovativeness in Shopping' have to be left out of the model during the estimation of the paths of the mode. As these latent variables are not central in the model, this is not considered a large problem. In addition to leaving out these latent variables, two variables (e65 and e66) were left out of the scale measuring shopping enjoyment, and two variables (e616 and e619) were left out of the scale measuring inclination to in-home shopping. Both of these scales had many items and, therefore, it is not considered a problem to leave out a few items from the scales in the subsequent test of the measurement model in total and in the estimation of the equations in the structural model.

4.2.2 Validation of the Measurement Model in Total

The next step in Structural Equation Modelling is a test of the measurement model. The test focuses on two areas: The correlations between the latent variables and the corresponding observed variables, and also correlations among latent variables. The model is displayed in figure 4.1 below. The purpose of this test is to ensure that the measurement model is appropriate in the sense that observed variables load on latent variables they are supposed to measure, and not on other latent variables which they are not supposed to measure, and to test discriminant and

concurrent validity of the measures in the sense that the ‘right’ latent variables correlate. A problem in interpretation of the validity is that the model, in this case, is quite large and some of the latent variables are connected through other latent variables, even though they do not correlate directly.

Figure 4.1: Initial Measurement Model



The expectations in terms of correlations between the variables in the measurement model are, based on the hypotheses stated in chapter 2.6 and the discussions in chapter 3.3, as displayed in figure 4.2 below:

Figure 4.2 Expected Covariances in the Measurement Model

	Product knowledge	Product interest	Availability of information	Credibility of information	No Imp. of Internet in life	Role Overload	Cost of search	Inclination to in-home shopping	Shopping enjoyment	Contact with sales staff
Product knowledge		+	+	+						
Product interest	+		+	+						
Availability of information	+	+		+	-		+	+		
Credibility of information	+	+	+		-		+	+		
No Imp. of Internet in life			-	-		-	-	-		
Role Overload					-		+	+		
Cost of search			+	+	-	+		+	-	
Inclination to in-home shopping			+	+	-	+	+		-	-
Shopping enjoyment							-	-		+
Contact with sales staff								-	+	

Rem.: grey indicates no expected covariance; - a negative covariance; + a positive covariance.

As displayed in figure 4.2, product knowledge and product interest are thought to correlate with each other and also to correlate positively with belief in availability and credibility of the needed information on the Internet.

Lack of importance of the Internet in the person's life is assumed to correlate negatively with 'Role Overload', perceived (low) cost of using the Internet, with belief in availability and credibility of the needed information on the Internet and also with proneness to in-home shopping.

Belief that generally, there is information on the Internet and that Information on the Internet is credible are also assumed to be linked to each other, to perceived cost of using the Internet for information search and to proneness to in-home shopping, and negatively to lack of importance of the Internet in the person's life and interest in and knowledge about the product.

'Role Overload' is expected to be positively linked to perceived (low) cost of using the Internet, negatively to lack of importance of the Internet, and positively to inclination to in-home shopping.

Perceived (low) cost of using the Internet for information search is expected to correlate positively with 'Role Overload', negatively with lack of importance of the Internet, and positively to belief in availability of the needed information, and that the information is credible, and finally to inclination to shopping from home.

Inclination to in-home shopping is believed to be related to the variables already described, and also negatively to enjoyment in shopping and contact with the sales staff.

Shopping enjoyment is believed to correlate positively with enjoyment in contact with the sales staff and vice versa, and both variables are expected to correlate negatively with inclination to in-home shopping. Also, enjoyment in shopping is expected to correlate negatively with low perceived cost of using the Internet for information search.

The empirical test of the measurement model follows the procedure described in appendix D. Initially, the parameters were estimated using Maximum Likelihood estimation. Subsequently, the output (in detail printed in appendix E.1.1) was inspected for offending results as described in appendix D. The variance for the error term for one of the variables measuring 'Role Overload' was negative. Setting the variance to a very small value resulted in the other variable's error term having a negative variance. Negative variances point to a misspecification of the model. Because there was no theoretical reason for changing the model in terms of further correlations between 'Role Overload' and other variables, and because 'Role Overload' is not central in the model, 'Role Overload' was omitted from the model.

After the omission, the results on the chosen validation criteria are presented in table 4.21 (detailed output in appendix E.1.2):

Table 4.21: Fit indices, Measurement Model after Modification 1 Omission of 'Role Overload'

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
718.273	476	0.000	1.509	0.848	0.915	0.047	0.770

The p-value in table 4.21 is 0.00, indicating that there is zero probability that the model is true, given the data in the sample. However, as stated in appendix D, there are problems in using the p-value as the only indicator of level of fit**.** Therefore, the other fit indices are deemed more valid. All these measures are at an acceptable level except for GFI. The value for GFI is too low indicating possibilities for improvement of the model. Therefore, the output was inspected in detail to reveal possibilities for improvement. The inspection revealed that the error terms connected to the variables measuring product knowledge and product interest seem to be correlated. Furthermore, the variables interest in and knowledge about the product compared to friends seem to load on each other's latent variables. This problem was foreseen after the factor analysis, in which the 4 variables were allocated to the same factor. Be-

cause of these problems, it was decided to create a new latent variable, 'Product Involvement' measured by e1a2, e1b1 and e1b2. E1a1 was omitted from the model because there were very large crossloadings between the four variables measuring product involvement. The reason for choosing e1a1 for omission from the model is that omitting this variable has the smallest effect on alpha. Alpha for the new scale is 0.76 (with all four observed variables, it is 0.79).

After this alteration, the validation measures are as follows in table 4.22 (detailed output in appendix E.1.3):

Table 4.22: Fit indices after modification 2 of the Measurement Model

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
634.346	448	0.000	1.416	0.859	0.931	0.042	0.956

There is only a slight improvement in the measures and still, GFI is not at an acceptable level. The modification indices were therefore once again inspected for possibilities for improvement. There are two possibilities for improvement: Adding covariances between variables different from zero, or improving 'weak' scales by omitting observed variables which load on several latent variables. e3d1, e3d3, e615, e63 and e617 all load on several of the latent variables and are, therefore, candidates for removal from the model. However, removing both e3d1 and e3d3 from the model leaves us with only one variable measuring importance of the Internet. This poses a problem in the identification of the model. Therefore, an alternative model of measuring importance of the Internet was tried using two variables: The intensity in the use of the Internet (e3b) and 'Use of the Internet is important/not important' (e3d2). As the latter is scaled inversely to the first, the scale was recoded to its inverse values (and named e3d2inv). The latent variable is therefore now measuring 'Importance of the Internet' instead of 'No importance of the Internet'. Cronbach's alpha for this new scale is 0.55, which is rather low indicating that also this way of measuring importance of the Internet in the person's life is far from ideal. On the other hand, if using this scale redeems the problem with crossloading, it is

worth trying. Furthermore, a factor analysis on the data set with these new measurement variables puts them in the same factor with loadings above 0.7.

The result for the alternative measurement model is shown on table 4.23 below (detailed output in appendix E.1.4):

Table 4.23: Fit indices after modification 3 of the Measurement Model

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
549.852	418	0.000	1.315.	0.874	0.946	0.037	0.997

The fit indices for this model are better than for the previous model and, therefore, this way of measuring 'Importance of the Internet' is chosen for the subsequent analyses.

However, as the modification indices (especially GFI) show, there is still reason for trying omission of the variables e615, e63 and e617 as these variables load on several of the latent variables. The variables belong to the scales measuring shopping enjoyment and inclination to in-home shopping. Both of these scales posed problems in the partial analysis of the scales in the form of skewnesses in different directions. Both scales also include quite many variables. Therefore, removing a few variables does not provide any problems in terms of the model not being identified.

Results for the omissions were as follows in table 4.24. (Detailed output in appendices D.1.5, 6 and 7):

Table 4.24: Fit indices after modifications 3, 4 and 5 of the Measurement Model

Omitting	(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	P-CLOSE
E615	487.646	389	0.000	1.254	0.883	0.959	0.033	1.000
E63	437.613	361	0.004	1.212	0.890	0.966	0.030	1.000
E617	387.076	334	0.024	1.159	0.897	0.975	0.026	1.000

The results in table 4.24 are very satisfying and show improvement after each modification. The next step then is to inspect regression coefficients and standard errors for offending estimates. All regression coefficients are significantly different from zero indicating that all variables load on the latent variable they are supposed to measure. None of the regression coefficients are close to zero or 1. None of the remaining modification indices are large. Therefore, no further modifications of the model are made. The next step is to find out whether any of the covariances found are not significant. If they are insignificant they should be left out of the model in order reach the simplest model possible.

The estimated covariances are shown in table 4.25 below.

Table 4.25: Estimated Covariances in the Measurement Model

Covariance between		Standardized Estimate	Critical ratio	p-value
Perceived Availability of Information	Product Involvement	0.026	0.410	0.68
Perceived Credibility of Information	Product Involvement	-0.055	-.800	0.42
Perceived Availability of Information	Importance of Internet	0.509	4.507	<0.001
Perceived Availability of Information	Perceived Credibility of Information	0.477	4.629	<0.001
Perceived Credibility of Information	Importance of Internet	0.146	1.607	0.54
Perceived Availability of Information	In-home Shopper	0.362	4.014	<0.001
Perceived Credibility of Information	In-home Shopper	0.291	3.476	<0.001
Importance of Internet	In-home Shopper	0.188	1.978	0.05
Perceived Cost of Acquiring information	Importance of Internet	0.450	3.538	<0.001
Shopping Enjoyment	In-home Shopper	- 0.107	-1.373	0.17
Contact to Sales Staff	Shopping Enjoyment	0.233	2.694	0.01
Contact to Sales Staff	In-home Shopper	-0.055	-0.741	0.46
Perceived Cost of Acquiring information	In-home Shopper	0.370	3.489	<0.001
Shopping Enjoyment	Perceived Cost of Acquiring information	0.085	1.179	0.24
Perceived Availability of Information	Perceived Cost of Acquiring information	0.676	4.398	<0.001
Perceived Credibility of Information	Perceived Cost of Acquiring information	0.337	3.264	<0.001

A glance at critical ratios for the covariances displayed in table 4.25 shows that far from all covariances are significantly different from zero. The covariance between 'Belief in Availability of Information' and importance of the Internet and the covariance between 'Belief in Availability of Information' and belief in credibility of the information on the Internet are both significantly different from zero and positive as expected according to the change from lack of importance to importance of the Internet in the person's life. Furthermore, belief in credibility and availability of information both correlate positively with inclination to in-home shopping and with perceived cost of using the Internet for information search.

None of the covariances linked to product involvement are significant. Contact with sales staff correlate significantly only with shopping enjoyment. The covariance between shopping enjoyment and inclination to in-home shopping is not significantly different from zero. Still, the sign of the estimated covariance is negative, as expected.

As expected, inclination to in-home shopping correlates significantly with both belief in availability and 'Belief in Credibility of Information' on the Internet, and with perceived (low) cost of using the Internet and importance of the Internet to the respondent.

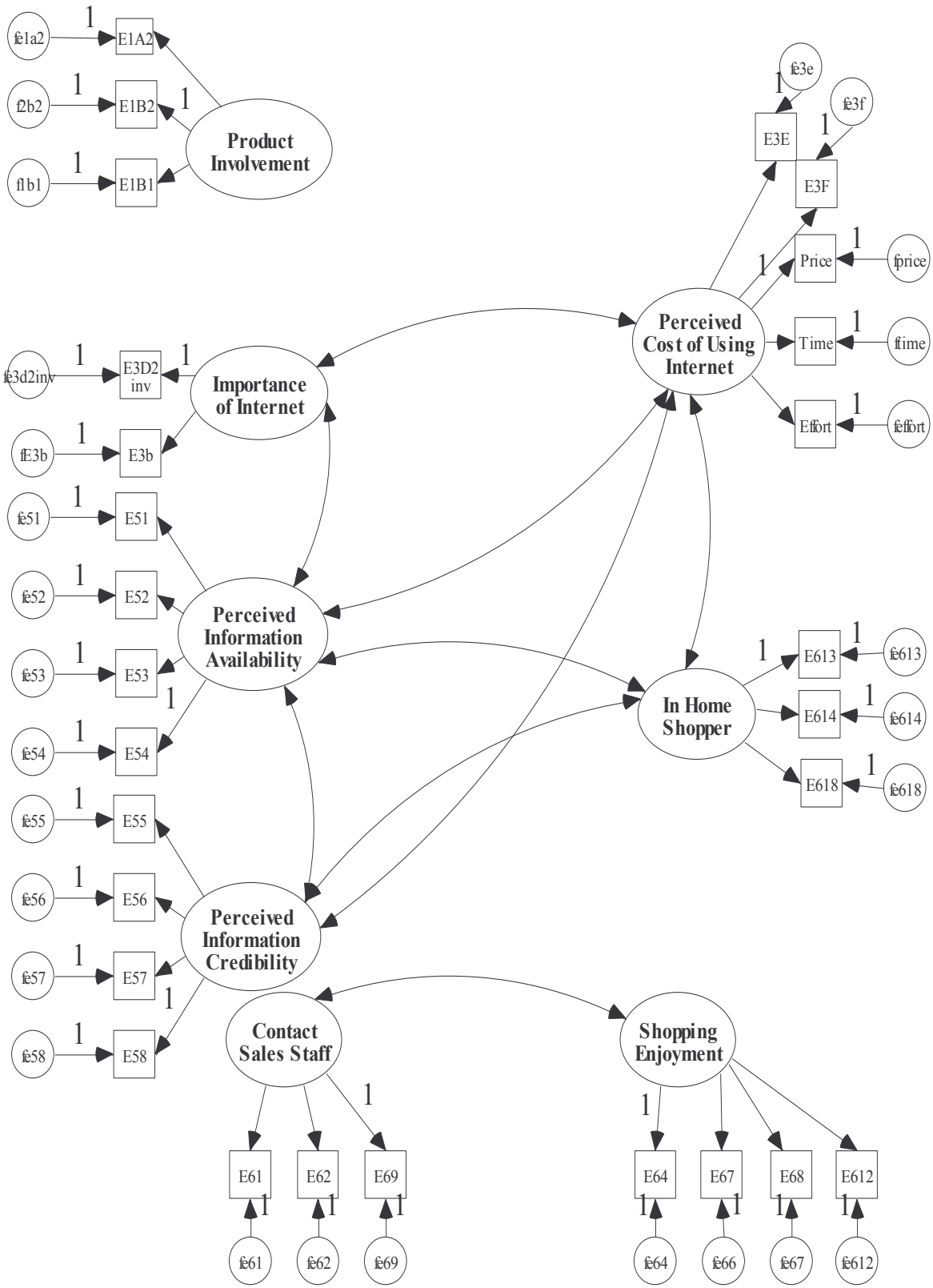
In addition to the covariances, significantly different from zero already accounted for, (low) 'Perceived Cost of Using the Internet' correlates significantly with importance of the Internet to the respondent.

The results are mainly in accordance with the hypotheses concerning relations in the Internet Search Model and not contradictory to the expectations put forward in the beginning of this chapter even though not all relations have proven to be significant: 'Contact with Sales Staff' and Shopping Enjoyment' are closely related to each other but not significantly to inclination to in-home shopping. However, the sign of the non-significant relation is negative, as expected. 'In-home shopping', 'Belief in Availability and Credibility of Information', 'Importance of

the Internet' and 'Perceived Cost of Using the Internet' are related, as expected.

Due to a wish for parsimony in the model, the covariances that were not significantly different from zero were successively omitted. Detailed output from this process is reproduced in appendices D1.8 and D.1.9. As the latent variables 'Contact to Sales Staff', 'Shopping Enjoyment', and product involvement are linked only to 'Use of the Internet for Information Search' in the Internet Search Model, these latent variables which do not correlate with the rest of the variables are kept in the model. The final measurement model is reproduced in figure 4.3 below.

Figure 4.3: Final Measurement Model



The fit indices for the final measurement model are as follows in table 4.26:

Table 4.26: Fit indices, Measurement Model

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
397.115	341	0.019	1.165	0.896	0.974	0.027	1.000

All fit indices are at a level showing a good fit. There are no offending residuals or estimates and the scales behave as expected in relation to each other, confirming their validity.

Table 4.27: Reliability Measures for Scales used in Estimation of Structural Equations

Latent Variable	Reliability measured by Chronbach's Alpha
Product Involvement	0.76
Importance of Internet	0.55
Belief in Availability of Information	0.79
Belief in Credibility of Information	0.85
Perceived Cost of Using the Internet for Information Search	0.72
In-home Shopper	0.60
Contact to Sales Staff	0.83
Shopping Enjoyment	0.73
Use of the Internet for Information Search	0.63

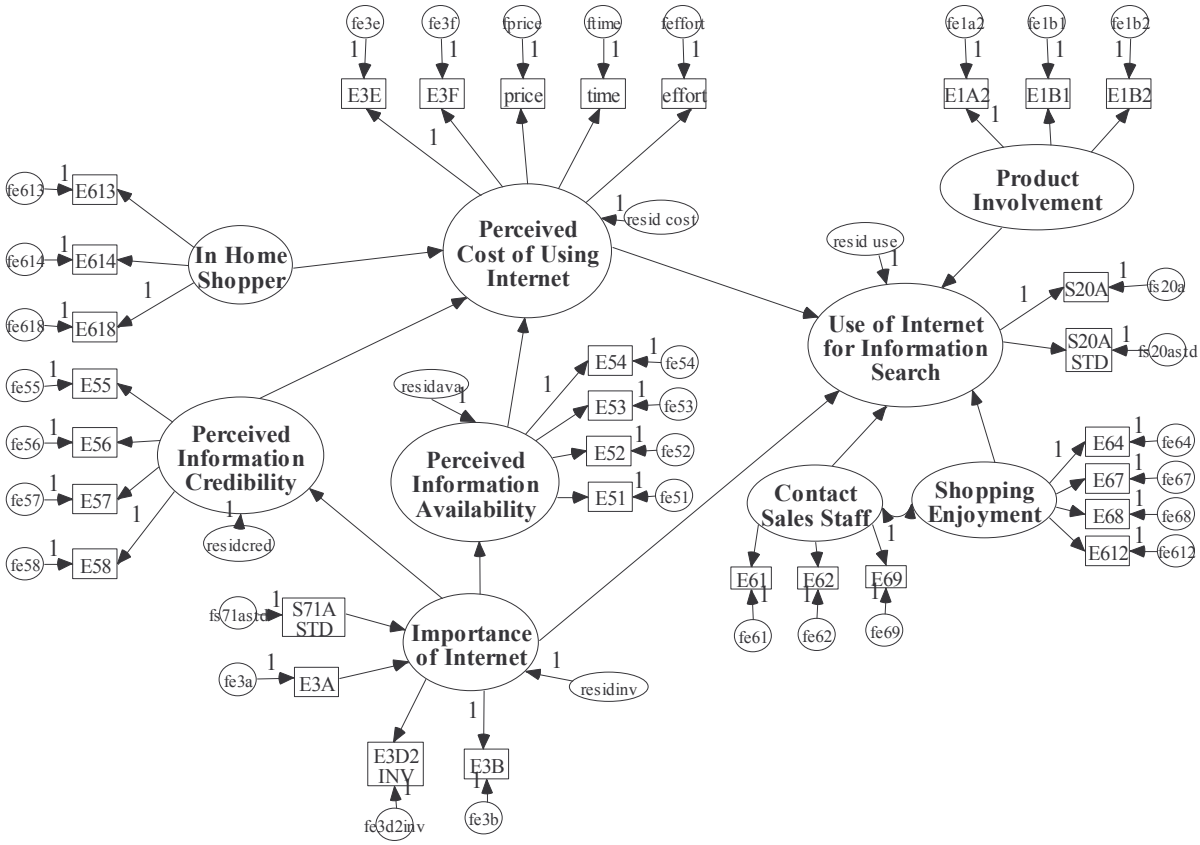
In table 4.27 above, final reliability measures of the scales are presented including the scale measuring Use of the Internet for Information Search, as this scale will also be used in the following. For two of the scales, the reliability level is lower than the often used level of acceptability (0.7) and for one of the measures, "Importance of the Internet in the person's life", the level is even lower than the, by others recommended "desirable" level of "about 0.6" (Bagozzi and Yi, 1988). The low reliability can produce problems in the later estimation in the form

of non-significant paths and is in any case a sign that the measurement of the latent variable in question in problematic. However, the factor analysis carried out in this chapter showed that the two variables measuring Importance of the Internet is in the same factor and indicates that these constructs do, in fact measure the same thing. Therefore, in spite of the problems in measuring importance of the Internet in the person's life, the measurement model displayed in figure 4.3 is used in the analysis of the hypothesized structural equations.

4.3 Analysis of Structural Equations

The proposed Internet Search Model with observed variables is shown in figure 4.4 below. The difference between this model and the measurement model is that the causal structure is introduced. As a consequence of this some covariances are replaced by causal paths. Also, the latent variable 'Use of the Internet for Information Search' which is explained by the other variables is introduced in the model.

Figure 4.4: Internet Search Model – Initial Model



The model is close to the model developed in chapters 2 and 3 but has been modified according to the results in the previous chapter. In short, the modifications carried out are deletion of 'Role Overload' and a few of the variables from two of the measurement scales.

The hypotheses regarding the paths in the model are the same as stated in chapter 2.6. To recapitulate, the expected relations are: 'Use of the Internet for Information Search' is dependent on the relative perceived cost of using the Internet, the involvement in the product, importance of the Internet and enjoyment in shopping. More involvement in the product results in more use of the Internet for search for product information. Smaller perceived cost of using the Internet for information search compared to other sources will result in the consumer using the Internet more for search for information. Also, the larger a place, the Internet occupies in the consumer's life, the more he is prone to using the Internet for information search purposes. In contrast to this, if the consumer likes to go shopping as a social activity, involving either sales staff or

co-shoppers, this will be off-setting the positive influence from the other factors.

Perceived cost of using the Internet for information search is augmented by a general preference for shopping from home plus a belief that the information is there, and that the information available is credible. Belief in availability and credibility of information, in turn is dependent on importance of the Internet to the consumer. Inclination to shop from home, and belief in availability and credibility of the needed information, move the barrier constituted by the need to bee physically 'there', while importance of the Internet in the person's life makes the adaptation of the use of the Internet for information purposes easier. Importance of the Internet is influenced negatively by age and positively by the time, the person have had access to the Internet.

The parameters in the model were estimated using Maximum Likelihood estimation as proposed by Byrne (2001). The results on the fit indices are presented in table 4.28 below (detailed output in appendix E.2.1):

Table 4.28: Fit Indices, Initial Structural Model

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
647.299	453	0.000	1.429	0.854	0.914	0.043	0.93

There are no offending results in the output and the fit indices are all, except for GFI, at an acceptable level. However, there are some quite large values among the standardized residuals, and also some rather large values among the modification indices, indicating misfit of the model. The misfit is related to 'nonexistent' covariance between 'Availability of Information' and 'Credibility of Information' on the Internet. Furthermore, several of the observed variables linked to other latent variables load on the credibility variable. A possible solution to this problem could be to add a non-zero covariance between the two latent variables to the model. However, an inspection of the regression coefficients shown in table 4.29 reveals that the path from 'Perceived Credi-

bility of the Information on the Internet' to perceived cost of using the Internet is close to zero, and the path from 'Importance of the Internet' to 'Perceived Credibility' is far from significantly different from zero[19]. Therefore, another possible strategy to follow is to eliminate 'Perceived Credibility of the Information' from the model, remembering in the following that there is a close connection between belief in availability and belief in credibility of the relevant information on the Internet.

[19] A rule of thumb is that the critical ratio, based on a significance level of 0.05 has to be larger than 19.6 or smaller than -1.96 in order to be significantly different from 0 (Byrne, 2001; 76).

Table 4.29: Regression weights, paths in the model

Regression Weights Path to	Regression Weights Path from	Standardized estimate	Critical ratio	p-value
Use of Internet for Information search	Perceived Cost of Acquiring information	0.109	1.105	0.269
Use of Internet for Information search	Product Involvement	0.257	2.915	0.004
Use of Internet for Information search	Importance of Internet	0.448	3.697	<0.001
Use of Internet for Information search	Shopping Enjoyment	0.041	0.468	0.640
Use of Internet for Information search	Contact with Sales Staff	-0.076	-0.891	0.373
Perceived Credibility of Information	Importance of Internet	0.164	1.873	0.061
Perceived Cost of Acquiring information	Perceived Credibility of Information	0.023	0.325	0.745
Perceived Cost of Acquiring information	Perceived Availability of Information	0.634	5.168	<0.001
Perceived Cost of Acquiring information	In-home Shopper	0.185	2.267	0.023
Perceived Availability of Information	Importance of Internet	0.503	4.410	<0.001
Importance of Internet	Time used Internet (e3a)	0.441	5.090	< 0.001
Importance of Internet	Age (s71astd)	-0.255	-3.196	0.001

Due to a wish for parsimony, the second strategy was chosen, and a second Maximum Likelihood estimation was made. The results for the fit indices are presented in table 4.30 below (detailed output in appendix E.2.2):

Table 4.30: Fit indices, Modification 1 – omitting perceived credibility of the information from the model

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
456.205	341	0.000	1.338	0.881	0.933	0.038	0.984

The fit indices for this model have all improved in comparison with the fit indices from the former model presented in table 4.28. However, an inspection of the fit indices indicates that there are still problems: The age ('s71astd') variable also loads on 'Shopping Enjoyment'. One solution to this would be to add a path from age ('s71astd') to 'Shopping Enjoyment'. However, the regression paths displayed in table 4.29 indicate that the paths, from 'Shopping Enjoyment' and enjoyment in contact to the sales staff to 'Use of the Internet for Information Search', are close to zero and far from being significantly different from zero. As a consequence of this, it seems to be a better solution to eliminate these two variables from the model, even though it means that one of the hypotheses, stated in chapter 2.6, will fall.

The results from the following estimation are displayed in appendix E.2.3 and the fit indices are shown below in table 4.31 (the variables were removed one at a time):

Table 4.31: Fit indices, Modification 2 - omitting 'Shopping Enjoyment' and 'Contact to Sales Staff' from the model

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
222.454	183	0.025	1.216	0.916	0.967	0.031	0.993

The fit indices have all improved remarkably except RMSEA, which has not changed very much. The small change in RMSEA shows that, taking the fewer variables into account, the improvement is there but is not very big. In spite of the good results in terms of the values of the fit indices, there is still need for improvement of the model. This is because an inspection of standardized residuals and modification indices reveals that there is probably a covariance which is significantly different from zero between inclination to in-home shopping and 'Belief in Availability of Information'. The interpretation of this is that individuals who have preference for in-home shopping also tend to believe that there is relevant information on the Internet. This may very well be the case and, therefore, the covariance between these two variables was freed and replaced by a path from In-home shopper to perceived availability of information. The resulting fit indices from the estimation were as follows in table 4.32 below (detailed output in appendix E.2.4):

Table 4.32: Fit indices, Modification 3 – adding a path in-home shopping to perceived availability of information

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
205.754	182	0.109	1.131	0.921	0.980	0.024	0.999

The fit indices in table 4.32 all indicate a very good fit and are all improved by the alteration of the model. $\Delta\chi^2_{(1)}$ is 16.7 and, therefore, the improvement in fit is significant. There is but one slightly problematic standardized residual covariance and, therefore, further fitting of the model may be problematic. Still, one problem is left: The variable, product interest in general (e1b1), loads on several of the latent variables. Therefore, this variable was removed from the model. The fit indices from the following estimation are presented in table 4.33 below (detailed output in appendix E.2.5):

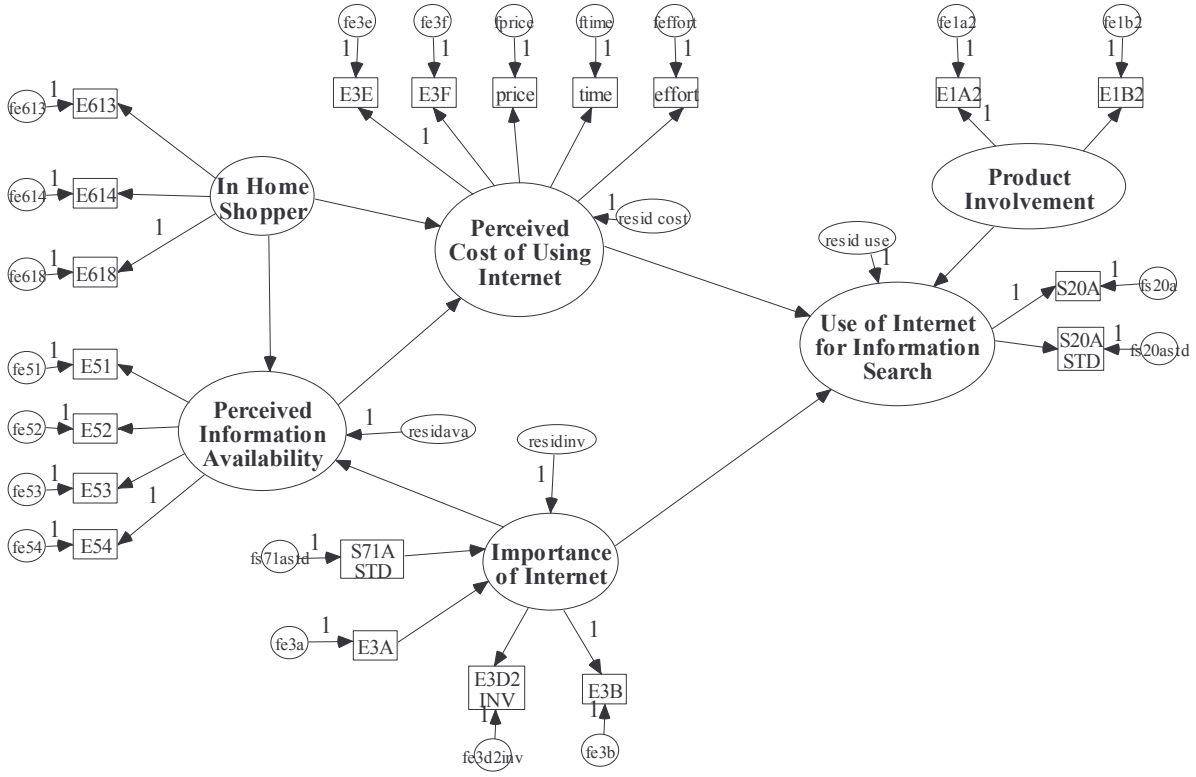
Table 4.33: Fit indices, Modification 4 – omission of e1b1 from the model.

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
187.258	163	0.094	1.149	0.924	0.975	0.026	0.998

The fit indices are slightly deteriorated but still indicate excellent fit. This model is, therefore, chosen as the model that best fits the data.

To check validity of the model, the estimation procedure was run using bootstrapping because a few of the variables deviate from normality. The results from the bootstrapping procedure show that the bias caused by this deviation is very small and has no significance regarding size and significance of the estimates. The Bollen-Stine bootstrapped p-value testing the hypothesis that the model is correct is as high as 0.358 indicating very good fit (detailed results in appendix E.2.5). To further test the validity of the model the parameters were estimated using a measure of use of the Internet including only the respondents who used the Internet for information search. This was to see, if there was a confusing difference between users and non-users of the Internet for information search. The Maximum Likelihood estimation resulted in a small deterioration in the fit indices (RMSEA=0.037), but the paths in the model did not change significantly. This again supported the validity of the model.
The final model is shown in figure 4.5 below.

Figure 4.5: Final model of determinants of use of the Internet for Consumer information search.



There are 47 parameters to be estimated in the model. This is an adequate number compared to the sample size. There are no offending estimates among the error variances or the regression estimates and the fit indices indicate excellent fit.

The regression paths (loadings) of the variables are as follows in table 4.34:

Table 4.34: Estimated regression paths for the Internet Search Model

Regression Weights Path to	Regression Weights Path from	Standardized estimate	Critical ratio	p-value
Use of Internet for Information search	Perceived Cost of Acquiring Information	0.152	1.574	0.116
Use of Internet for Information search	Product Involvement	0.290	2.600	0.009
Use of Internet for Information search	Importance of Internet	0.452	3.786	<0.001
Perceived Cost of Acquiring information	Perceived Availability of Information	0.624	5.110	<0.001
Perceived Cost of Acquiring information	In-home Shopper	0.136	1.656	0.098
Perceived Availability of Information	Importance of Internet	0.421	3.319	<0.001
Importance of Internet	Time used Internet (e3a)	0.475	5.601	<0.001
Importance of Internet	Age (s71astd)	-0.254	-3.237	0.001

The relations in the model are mostly as hypothesized: There is a positive relation between a lower perceived cost of using the Internet and the actual use of the Internet for search for consumer information in relation to an upcoming purchase. This path is not significantly different from zero. However, this can be because of the relatively small sample (Byrne, 2001). There also is a positive relation between the importance of the Internet to the consumer and use of the Internet for search for information in relation to an upcoming purchase. This latter relation is much stronger than the indirect relation which goes through availability and cost of searching for information indicating that the choice to use the Internet for information search is mostly dependent on a general lik-

ing of using the Internet, and not so much on a cognitive decision based on an evaluation of which source will provide the needed information at the lowest perceived cost.

The Internet is, as hypothesized, generally used relatively more by people who are involved in the product, for which they search information. Importance of the Internet is augmented with the length of time that the person has used the Internet. Also, younger persons are more involved in the Internet than older person are when they have been online for the same amount of time. We do not know if this has to do with maturation in the sense that more mature persons use the Internet less in general, or if it is linked to the earlier use of Internet in life and, therefore a more habitual use of the Internet analogue to the diffusion of many other electronic devices in society like the video recorder, mobile phone etc.

At the end of chapter 4.2, we saw that the reliability of the scale measuring importance of the Internet in the person's life was rather low (0.55). As this can cause problems in the estimation of the paths connected to this latent variable, experiments were made using different ways of incorporating the variables connected to this latent variable in the form of substitution of the latent variable by one of the measurement variables at a time. These experiments showed the stronger of the two variables in explaining use of the Internet for information search is frequency in the use of the Internet and not the indication of importance of the Internet as an answer to a question about this. The fit indices for this model are slightly lower than for the model in figure 4.5. However, the path from 'Perceived cost of using the Internet' to 'Use of the Internet for information Search' becomes highly significant (p-value 0.007). This is because the variable which was removed (e3d2inv) correlated also partly with the cost variable and thereby, the picture is confounded. However, the results from this experiment did not alter the conclusions as to relations in the model.

Perceived cost of using the Internet for information search plays a role in 'Use of the Internet for Information Search'. Perceived cost is, in turn, dependent on perceived availability of relevant information, and also on a general preference for home-shopping. The latter path is not

significantly different from zero, but the critical ratio is not far from the significance level. Because of this and the relatively small size of the sample, the path is kept in the model even though the relation is weak. As put forward in the above, the weak path is probably due to the poor measurement of importance of the Inter in the person's life.

The final latent variable in the model is 'Perceived Availability of Information on the Internet'. A general inclination to home-shopping and importance of the Internet in the person's life augments belief in availability of information on the Internet. The first relation was not expected in the starting hypotheses, but seem plausible, as persons who have a general preference for home-shopping are used to, or at least accept, relying on information they can retrieve at home from catalogues etc.

Even though most of the paths in model were as expected from the hypotheses in chapter 2.6, there were also paths which proved to be far from significant in the empirical test of the Internet Search Model. These are the relations which are related to enjoyment in shopping and/or personal contact with sales staff. A positive attitude towards shopping and contact with sales staff was expected to be offsetting to search for information on the Internet. However, this does not seem to be the case. Out of precaution, I tried to link these variables directly to the variable measuring importance of the Internet, but these paths were very far from significant.

Table 4.35 below shows the squared multiple correlations for the final model. In the table is presented only the part of the output which is relevant for the variables which are explained by the Internet Search Model. As the focus in this project is on use of the Internet for information search, the squared multiple correlation for this latent variable is, of course, of special interest.

Consulting table 4.35, we see that the Internet Search Model explains 34.7% of the variations in the answers, concerning use of the Internet for information search. There are no strict rules regarding acceptable sizes of this value, but consulting with the sources used in appendix D, the level of explanation of use of the Internet for information search

seems to be at a satisfactory level. However, the level also indicates that there are determinants of use of the Internet for information search which are not incorporated in the model.

Table 4.35: Squared multiple correlations for the Internet Search Model

Squared Multiple Correlations	Estimate
Importance of Internet	0.290
Perceived availability of information	0.274
Perceived cost	0.461
Use of Internet for information search	0.347

The other latent endogenous variables are also explained fairly well, as the numbers in table 4.35 show us. The squared multiple correlation for Perceived Cost (of using the Internet for search for information) is 0.461, while a little less than 30% of the variation in the data for the variables importance of the Internet and perceived availability of the needed information on the Internet is explained by the model.

As there may be gender differences in the determinants of information search behaviour, I tested whether it is safe to hypothesize that the two genders share the same paths in the model. Detailed results from the estimations can be found in appendices D.2.6 and D.2.7. The χ^{2-} value from a model with the constraint that the paths are the same for the two groups is 392.957 with 335 degrees of freedom. To evaluate this number, it is compared to the value for the same model with no constraints (Byrne, 2001). The value from this model is 383.309 with 326 degrees of freedom. This means that $\Delta\chi^2_{(9)} = 9.648$, and the model with no constraints, therefore, does not fit the data significantly better than the model with the constraint that the paths are the same for men and for women. The conclusion to this is that it is safe to assume that the paths in the model are the same for women and men, meaning that there are no gender differences in the explanations for use of the Internet for information search.

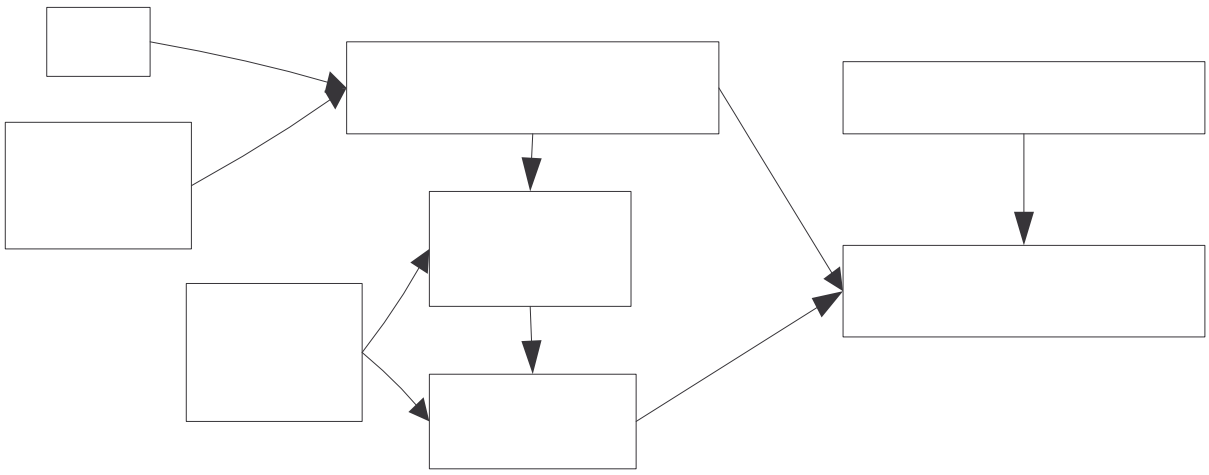
The model was estimated using all respondents in the sample and thus, not using the Internet was considered to be a level of Internet use. However, some respondents stated not to have searched at all, even though they later stated to have used the Internet for information search, and others stated not to have used the Internet, for search for information. Because of these problems, I estimated the parameters in the model without 'non-searchers' and without non-internet searchers. This did not alter the estimations significantly, although the relations, due to the fewer observations, were weaker.

4.4 Determinants of Use of the Internet for Information Search

In chapter 2.6, I stated a number of hypotheses in relation to the Internet Search Model. The empirical test of the model mostly supported most of the hypotheses but not all. Especially, it was surprising that the hypothesis on the importance of shopping enjoyment and contact to sales staff was not supported.

The final model is as shown in figure 4.6 below.

Figure 4.6: Final Internet Search Model



In numerical order, the conclusions on the tests of the hypotheses are as follows:

Hypothesis 1: There is a positive relation between involvement in the product, measured as interest in and knowledge about the product, and

use of the Internet for search for information. The original hypothesis did not expect the two variables to be so closely related that they act as one variable in the model.

Hypothesis 2: There is a positive relation between a perceived low cost relative to the benefit of finding information on the Internet, and use of the Internet for search for information. However, this relation was only close to being statistically significant.

Hypothesis 3: There is a rather strong relation between importance of the Internet and use of the Internet for information search purposes, and also between importance of the Internet and perceived cost of using the Internet through the belief in the availability of the needed information on the Internet.

Hypothesis 4: There was support to the hypothesis that consumers, to whom the Internet is important, believe that the information is there. There was no direct support to the hypotheses that consumers, to whom the Internet is important, are more prone to believing that the available information is credible. There was, however, a strong covariance between belief in availability and belief in credibility of the needed information on the Internet. But the conclusion, nonetheless, is that consumers in general think no more, no less of the information on the Internet than they think of the information they can acquire off-line.

Hypothesis 5: The analyses carried out in chapter 4.1.6 gave support to the hypothesis that consumers use the types of sources for information search that they also use in general for other purposes.

Hypothesis 6: The test of the model showed a positive relation between an inclination to in-home shopping and perceived low cost of using the Internet as an information source. However, the relation was only close to being significantly different from zero. Due to scaling problems, it was not possible to test the impact of the latent variables 'Innovativeness in Shopping' and/or 'Role Overload' as facilitators in the change in search behaviour.

Hypothesis 7: There was no support to the hypothesis that shopping enjoyment and contact with sales staff would work counter to the use of the Internet for information search. The estimated path from shopping enjoyment to use of the Internet for information search was very close to zero, while the estimated path from enjoyment in contact to the sales staff was larger with a negative sign, as expected in the hypothesis. However the path was very far from being significant, and was therefore deleted from the model. The results on this question may be connected to the choice of products for the study.

Hypotheses 8 and 9: There was strong support to the two hypotheses on determinants of importance of the Internet in the consumer's life. The longer a person has been on-line, the more he finds the Internet important and spends time on-line. This is, in part, off-set by the person's age. The older the person is, the less the person finds that the Internet is important to him, and the less time, he spends on-line.

Hypothesis 10: The relations in the model have not been found to be dependent on gender.

Hypothesis 11: There was support to the hypothesis that consumers use newsgroups more than chat-rooms when searching for product information. We shall return to this hypothesis in further detail in the following chapter on analysis of the data gathered in newsgroups on the Internet.

All in all, the hypotheses regarding the paths between the elements in the Internet Search Model which was developed in chapter 2 were supported. There was, however, one important exception to this, as the significance of the enjoyment in the social aspects of shopping could not be shown to have an off-setting effect on the use of the Internet for information search in relation to an upcoming purchase. This may be connected to the choice of products in the survey. Also, the connection between perceived relative cost of using the Internet for information search and use of the Internet for information search in relation to an upcoming purchase was only close to being significant. This may be due to the size of the sample or poor measurement of the importance of the Internet in the person's life but, along with the strength of the other

paths, indicate that the cost factor is less important than the perceived benefit from using the Internet in general.

5. Results from the Newsgroup Study

This chapter contains results from the empirical studies in the chosen newsgroups. The results are based on two related studies: Observation of communication in the selected newsgroups and a survey in the same newsgroups.

The chapter starts with a short description of the sample and evaluation of the validity of the sample. This leads to a description of the respondents in the newsgroups and their relations to the group and to the Internet as such.

After this introduction to the selected newsgroups, we are ready to turn to the purpose of collecting data from the newsgroups: A test of the hypotheses proposed in chapter 2.6. The first hypothesis (11) stated that consumers generally will use newsgroups more than chat-rooms for information search purposes. This hypothesis was confirmed by the analysis performed in chapter 4.1.6. Results from the study among Internet users in general confirmed that newsgroups are, indeed, used for search for consumer information in relation to an upcoming purchase. In the sample, 21.13% had used newsgroups for information search in relation to the purchase they made. Among these 21.13%, 77.52% had used newsgroups a little whereas 10.21% had used newsgroup much or very much (4 and 5 on scale from 1 to 5). Newsgroups had been used for information search in relation to all products groups in the questionnaire expect for vacuum cleaners. There was found no clear connection between age and use of newsgroups for information search. Results from the sample among Internet users in general confirm that the use of newsgroups is not yet widely spread among Internet users but not uncommon, either. In addition to this, virtual communities are the really new kind of source of consumer information on the Internet, as stated in chapter 2.4. All in all, the results confirm that it is interesting to learn more about consumer information search in newsgroups through an empirical test of the hypotheses concerning this issue.

Before answering the main question whether information search in newsgroups has replaced formerly used marketer- and non-marketer

dominated sources, it is necessary to know more about the selected newsgroups, the different types of consumers visiting the groups, and the relation between the two. This is because theory tells us that there is a relation between the tie a person feels to a source and the extent, to which the source is used, at least when the information needed is not straightforward to give. This issue was discussed theoretically in chapter 2.5.1.

The first hypothesis to be tested is the hypothesis stating that there is a connection between the subject of the newsgroup and the ties within the group (hypothesis 12). Groups centring on types of products with high self-centrality to the consumer like animals and hobby issues are expected to be more tightly knit than groups centring on non-personal issues. As a consequence of this, groups which centre on issues with high self-centrality to the consumer also should contain more Insiders, than do other groups[20]. This hypothesis is tested in chapter 5.3 below.

The second issue centres on communication within the group. In newsgroups, communication takes place as postings in the groups. Users of the newsgroups have to be motivated to enter the group and communicate in the newsgroup; they do not enter the newsgroup because they are put there by some force other than their own will. In addition to this, in order for the newsgroup to be used for exchange of consumer information, there has to be consumers in the group who are motivated to enter the group by the possibility to acquire information, and others who are capable of and also willing to give this information as explained in chapter 2.4.1. In order to find out if the selected newsgroups work this way, it is necessary to clarify the motivations of the members of the newsgroups and the connection between motivation and asking for and giving advice.

Kozinets (1999) have stated that the motivation to enter consumption-oriented newsgroups can be described by 4 factors. The factors are connected to the opportunity to get: Entertainment, Social Relations, Giving

[20] Insiders are characterized by the product having high self-centrality and by a strong tie to the group as defined by Kozinets (1999). See chapter 2.4.1 for more on this framework.

Advice, and Getting Advice. Hypothesis 13 states that the motivation to enter a newsgroup can be described by a model including these factors. As the model, to my knowledge has not been tested empirically, and in any case not with the measurement scales proposed in chapter 3.4, it necessary to test the model before testing relations to other latent variables. This is the topic of chapter 5.4.1.

After the measurement model has been tested, it is time to test hypothesis 14 which states that individuals belonging to different segments (as defined by Kozinets) have different motives for entering consumption oriented newsgroups.

As stated before it is not the intention to dig further into the sociological aspects of the communication in the newsgroup. More interesting in the current study is to clarify whether the motives for giving and asking for advice in consumer oriented newsgroups are linked to the person being opinion leader or opinion seeker, indicating that the motive to enter the newsgroup is, in fact related to personality traits in addition to the acute need for information and interest in the product. If this is true, it must encourage exchange of consumer information in the newsgroups, as these are then good places for the two types of persons to meet. Therefore, the topic of chapter 5.5 is to test whether the level of communication in the newsgroups can be explained by different motives to visit the newsgroup which fit each other well, and if the motives are governed by different types of personality in terms of inclination to seek and/or to give advice to others (hypothesis 15).

After this look into the newsgroups, the types of members they have, and the members' different motives for entering the newsgroups, and thereby having established that newsgroups can be regarded as 'communities', we are ready to try and answer the main question in this chapter: Have newsgroups replaced off-line sources formerly used for information search? Hypotheses 16 states that 'The use of newsgroups primarily replaces marketer-dominated sources, and this is more prevalent for Tourists than for the other segments' while hypothesis 17 states that 'The use of newsgroups has also replaced the use of other personal advice and this is more prevalent for Insiders than for the other segments.'

The replacement is expected to be larger in the newsgroups centring on products with high self-centrality to the consumer. This is because these newsgroups are expected to be tighter knit than the other newsgroups. Hypotheses 16 and 17 are tested in chapter 5.6.

The test of the hypotheses follows the procedures and uses the validation criteria described in chapter 3.1.2 and appendix D.

5.1 Size of and Communication within the Newsgroups

The newsgroups we are about to look into in the following were originally chosen for this study on the basis of data on the number of postings in the newsgroups during 2000 (Tønnesen, 2001). The data from Tønnesen were subsequently combined with observations in several newsgroups concentrating on the achievement of a better understanding of which topics are discussed in the various newsgroups on USENET. The newsgroups chosen for sampling are shown in table 5.1. The choice of these 4 newsgroups was based on three criteria: The first criterion was that the discussions in the newsgroups should be consumption relevant. The second criterion was that the newsgroup should be active in terms of a sufficient number of postings per time unit. The third criterion for selection was that the newsgroups should represent both products which are mostly chosen on objective criteria and products which are mostly chosen on subjective (feel) criteria. The newsgroups on products which are judged to possess high self-centrality to consumers are ‘Car’ and ‘Dogs’ – and the two newsgroups, in which the products discussed are expected to be chosen on objective criteria and, in addition to this, are not central to the consumer’s identity are ‘House and Garden’ and ‘Consumer Issues’.

Table 5.1: Activity in and size of the chosen newsgroups

Newsgroup:	Number of persons posting in November	Percentage with less than 4 postings	Average number of postings/person	Average number/posting pr. thread (discussion length)	Average number of threads/ person	Average postings pr. thread/person
dk.forbruger ('Consumer Issues')	187	80.21	2.72	7.82	1.74	1.50
dk.fritid.hus-og-have ('House and Garden')	150	84.93	3.16	5.92	2.11	1.41
dk.fritid.bil ('Car')	213	76.42	3.44	6.96	2.34	1.37
dk.fritid.dyr.hund ('Dog')	134	68.42	4.96	9.57	2.54	1.61

Table 5.1 shows the traffic in the selected newsgroups in terms of postings in total as well as statistics on member level. The distribution of size of the groups is the same as in the USENET statistics (Tønnesen, 2001). It is important to remember that the number of users registered is measured as users who have posted in the newsgroup during the observation period. This number is probably smaller than the total number of users visiting the newsgroups because persons who are just reading what others write to the newsgroup are not registered anywhere.

The newsgroup 'Car' is the largest newsgroup followed by the newsgroup 'Consumer Issues', which is a bit larger than the newsgroup 'House and Garden', whereas the smallest newsgroup is the newsgroup 'Dogs'. In all of the selected newsgroups, most of the members account for less than 4 postings during the month of November 2001. In each newsgroup, a few persons have many postings – as much as 30 or more in a month. The newsgroup 'Dogs' has a larger percentage of members with a relatively high number of postings compared to the other newsgroups. This is also reflected in the higher number of postings per person, on average.

The discussions in newsgroups run in so-called threads. The average number of postings per thread range from 5.92 in 'House and Garden', and up to 9.57 in the 'Dog' newsgroup, reflecting the larger number of persons participating in each discussion in this newsgroup. This is also reflected in the fact that the selected newsgroups show almost the same number of postings per person per thread, indicating that the larger number of postings per person in the 'Dog' newsgroup is the result of the same persons participating in more discussions and more different individuals participating in each discussion.

To investigate this issue further, I counted the number of individuals in each newsgroup participating in discussions with each other at least once, twice and so on during the month of November 2001. The results are presented in table 5.2 below. The table is read in the following manner: In the newsgroup 'Dogs', 82 of the individuals have participated in a discussion with each other once during the month of November 2001. 24 individuals 'met' the same individuals twice, 9 'met' the same individuals three times and 12 individuals 'met' with the same individuals on at least 4 occasions.

Table 5.2: Number of meetings with the same person

Newsgroup Number of meetings	Dogs	Consumer	'Car'	House and Garden
1	82	156	169	106
2	24	18	22	23
3	9	7	15	6
4+	12	2	6	5

From the numbers in table 5.2, we see that, by far, the most individuals have participated in the same discussion only once, indicating that persons visiting the selected newsgroups do not get to know each other by many direct encounters in discussions. This does not prevent that they can feel that they 'know' each other, because they may have read each others postings. In all newsgroups, around 20 persons have met other persons 2 times in discussions. The table shows that there is a clear dif-

ference between the newsgroups in terms of numbers of ‘meetings’ and a chi-square test confirms that the difference is significant. The largest deviation can be found to the ‘Dog’ newsgroup, where much more of the members than in the other three newsgroups have met in at least 4 threads. Also, several of the individuals in the ‘Car’ newsgroup have met in as much as three threads. This supports the thesis that the newsgroups ‘Consumer Issues’ and ‘House and Garden’ are looser, more information oriented newsgroups than the other two newsgroups.

The numbers indicate that the newsgroups ‘Car’ and ‘Dogs’ which are newsgroups discussing subjects more central to the persons involved, are newsgroups with a relatively larger share of the members being active, and the same members participating in several discussions. This indicates that these newsgroups, as expected from hypothesis 12, are tighter knit than the other two selected newsgroups. Especially the newsgroup ‘Consumer Issues’ is characterized by a large number of discussions, almost all taking place between different people.

5.2 Evaluation of the Sample

The number of respondents in the sample taken in the four chosen newsgroups was in total 106. However, 3 of the respondents had more than 10 unanswered items, and were therefore deleted using the same criteria as in chapter 4.1 on evaluation of the quality of the sample among Internet users in general. This omission of 3 respondents resulted in a sample consisting of 103 respondents. Of the 103 respondents, 91 had answered all questions, 6 respondents had one missing answer, and 4 respondents had missing answers to 2 items. The responses were distributed on the selected newsgroups as follows: 32 of the responses came from ‘House and Garden’, 14 from ‘Consumer Issues’; another 32 came from ‘Car’ and 25 from ‘Dog’.

As we do not know anything about the profile of the newsgroup user in general, it is not possible to evaluate the sample in terms of representativity in general. As we are in possession of observation data from the newsgroups, it is, however, possible to evaluate the sample in terms of activity level as compared to the newsgroup in total. It is also possible to

see whether persons who answered the questionnaire were the same persons, who posted in the newsgroup during the month of November 2001 (called the observation list in the following), and thereby judge the response rate. However, it is quite possible that some of the respondents have used a different e-mail address when responding to the questionnaire than the one used in the newsgroup. Therefore, there may be more persons in the sample who are on the observation list than it is possible to see in a comparison of e-mail addresses from the sample with the e-mail addresses from the observation list.

For each newsgroup the sample was checked for how many of the respondents it was possible to locate on the observation list, to see the overlap between the sample and the observations. It is important to remember that the observation list contains only members who have posted in the newsgroup and not all members of the newsgroup. Because three of the samples were drawn using self-selection, there is reason to fear that the persons responding are the persons who are more active in the newsgroup. Therefore, the respondents, whom it was possible to locate on the observation list, were compared to all individuals on the observation list in terms of number of postings, number of threads participated in and number of threads for each newsgroup. The comparisons were made on average numbers. As the distributions are often skewed, it was checked whether it would change the results if the median was used instead. This was not the case.

The first newsgroup which was evaluated was the newsgroup ‘House and Garden’. The numbers for the respondents in the newsgroup compared to the values for all observed postings in the newsgroup are displayed in table 5.3 below.

Table 5.3: Sample from Hus-og-Have ('House and Garden')

	Number of persons posting in November		Average number of postings/person	Average number of threads/person	Average posting pr. thread/person
Registration, November	150		3.16	2.11	1.41
In sample	On observation list	33	2.86	2.22	1.49
	Not on observation list	2			

Table 5.3 shows that 35 respondents came from the newsgroup 'House and Garden'. Of these, 33 were identified on the observation list. The relatively high number compared to what we are going to see in the other newsgroups is connected to the way the sample was collected: Through unsolicited e-mail to addresses harvested in the newsgroup. Therefore, there are no 'lurkers' between these respondents. The persons who answered the questionnaire are quite representative for persons posting in the newsgroups measured in terms of postings and number of threads participated in. 10 of the 33 identified persons posted at least one question to the newsgroup, compared to 75 of the 150 registered users in the newsgroup. In percentages, this is 30% in the sample, compared to 50% of the posters in total. This means that the respondents in the sample are less active than members of the newsgroup as such.

Table 5.4: Sample from dk.fritid.bil ('Car')

	Number of persons posting in November		Average number of posting/person	Average number of threads/person	Average posting pr. thread/person
Registration, November	213		3.44	2.34	1.37
In sample	On observation list	5	9.2	6.4	1.44
	Not on list	30			

The next newsgroup for evaluation is the newsgroup 'Car'. Table 5.4 shows that 35 usable responses came from the newsgroup 'Car'. Of these, only 5 were identified on the observation list. The persons who answered to the questionnaire and who are also identifiable on the list are not representative for persons posting in the newsgroup, as they have made far more postings and participated in far more threads than the average poster in the newsgroup. 2 of the 5 identified persons posted at least one question into the newsgroup as compared to 88 of the 213 registered users in the newsgroup. In percentages, this is 40 percent compared to 41%. This indicates that the persons who responded to the questionnaire are not more active in terms of asking questions than are members of the group in general. However, the number of identified persons is very low, and thus, the comparison is weak. The lack of match between the sample and the observation list indicate that either this is a newsgroup with a relatively large number of lurkers, or a large number of persons are using an alias in the newsgroup. The high level of activity in terms postings of the respondents who are also on the observation list, compared to the average poster in the newsgroup, indicates that the sample is probably skewed towards the more active users of the newsgroup. This was expected to be the consequence of using self-selection of respondents but poses, nonetheless, a problem in inferring from the data to the population in question.

The next newsgroup to be evaluated is the newsgroup 'Dogs'. The numbers for this newsgroup are displayed in table 5.5 below.

Table 5.5: Result for sample from dk.fritid.dyr.hund (Dogs)

	Number of persons posting in November		Average number of posting/person	Average number of threads/person	Average postings pr. thread/person
Registration, November	134		4.96	2.54	1.62
In sample	On observation list	10	13.22	7	1.89
	Not on list	15			

Table 5.5 shows that 25 usable respondents came from the newsgroup 'Dogs' and of these, only 10 were identified on the observation list. The persons who answered the questionnaire are, like we saw in the same figures for the preceding newsgroup, not representative for persons posting in the newsgroup, as they have posted far more often than the average poster have done. 4 of the 8 identified persons posted at least one question into the newsgroup, as compared to 49 of the 116 registered users in the newsgroup. In percentages, this is 50%, compared to 42%, indicating no large deviation from the average user when it comes to asking questions in the newsgroup. However, the numbers are, once again, quite small for this kind of comparison. It seems that more persons in this newsgroup use their own e-mail address than in the newsgroup 'Cars' or that the number of lurkers in this newsgroup is smaller. Both reasons indicate that this newsgroup is tighter knit than the newsgroup on cars. It may also play a role in the different results from these two newsgroups that the respondents from this newsgroup are all female, and the subject is dogs, whereas the car newsgroup is an almost exclusively male newsgroup, discussing a 'dead' object like cars. I do not know anything about that, though. The thought presented is based on mere prejudice (and a bit on observations of the tone in the newsgroups).

Table 5.6: Sample from dk.forbruger ('Consumer Issues')

	Number of persons posting in November		Average number of posting/person	Average number of threads/person	Average posting pr. thread/person
Registration, November	187		2.72	1.74	1.50
In sample	On observation list	5	3.4	2.4	1.47
	Not on list	9			

Table 5.6 shows the number of persons posting in November 2001 in the newsgroup 'Consumer Issues'. The result is much like the result from the other two newsgroups in which self-selection was used: Only 5 of the 14 respondents were identified as posters in November. This again indicates that there must be quite a large number of lurkers in the newsgroups. The respondents identified are quite representative for the newsgroup, as they have the same level of postings and threads as the average poster in the newsgroup. Of the 5 respondents identified on the list, 2 posted questions into the newsgroup, compared to 59 of the 187 participants in the newsgroup. In percentages, this number is about the same, again indicating that the sample is probably representative for the level of activity of the average user of the newsgroup even though the numbers are, once again, quite small for this kind of comparison.

The above partial validations of the samples drawn from the 4 selected newsgroups show that the overlap between persons on the observation list and persons in the sample varies quite much between the newsgroups. The lack of overlap can be due to two different factors: First that the sample was taken in April 2002, whereas the observation list is based on the communication in the newsgroups in November 2001. Second, it is possible that, in some of the newsgroups, the number of lurkers is quite high. Our problem in evaluating the sample in terms of distribution between lurkers and posters is that we do not know anything about the true distribution. In spite of these reservations, it seems quite

clear that the respondents who could be identified on the list are, on average, more active in the newsgroup than the average users on the observation list. As this is mostly the case in the part of the sample which was taken by self-selection of the respondents, this result confirms that a sample taken in such a manner is often skewed towards the more involved individuals.

Concluding on this evaluation of the sample taken in the newsgroups, I judge that the sample is usable with some caution. The judgement is based on the following assumptions: It is inevitable that there are some respondents in the sample who are not on the observation list as this list contains only individuals who have posted in the newsgroup. There may also be respondents who are actually on the list, but have not been registered as such because they have used a different e-mail address in the survey than the one they use in the newsgroup. At least, some respondents in the survey did not give me their e-mail address at all and may, therefore, be on the observation list. Probably, some of the respondents who were on the list from November were no longer users of the newsgroup in April the following year, but have been replaced by other 'short term users' and are, therefore, represented in the sample by these other short-term users. With the known numbers on activity level, it is, however, important to remember in the following that the persons in the sample are probably more involved in the newsgroup than is the average user.

After this initial evaluation of the usability of the sample, we are ready to look into the data to learn more about newsgroup members and their use of the newsgroups for information search. This is the topic of the chapters to follow.

5.3 Description of the Newsgroup Users

The first step in the analysis of the data from the newsgroup sample is an initial description of the newsgroup user in terms of demographic and personality variables; involvement in the product in terms of interest in and knowledge about the product; internet use; and finally, use of the newsgroups.

After the initial description it is time to investigate whether there are differences between members of newsgroups and the average Internet user and whether members of the 4 selected newsgroups differ. The current sample was compared to the sample among Internet users in general and the newsgroups were also compared with each other. Due to the very different sample sizes, the comparisons were done on two levels: Newsgroups contra average Internet users and newsgroups with each other. In the comparison of the newsgroups and the average Internet user, all the data were used. In the analyses of the differences between newsgroups, the data from the newsgroup 'Consumer Issues' were excluded because the number of observations from this newsgroup was too small for separate analysis.

Comparisons were performed using stepwise discriminant analysis supplemented by Tukey's test for comparison of means if the groups were found to differ for the variables which are metrically scaled, and chi-square analysis of for the variables which are nominally scaled. These statistical procedures are commonly known and, therefore, selection of estimation procedures and validation criteria are not described in detail. In general, the default procedures and significance levels are used.

5.3.1 Description of the Sample on Demographic Variables

The average age in the newsgroups is 33.85 years. The age level is significantly lower than the average age among Internet users in general, but still sufficiently high to show that the use of newsgroups is not exclusively for young people. There are differences between the newsgroups; the members of the newsgroup 'House and Garden' are significantly older on average than the members of the other newsgroups (41.78 years compared to 28.65 and 30.65 in 'Dog' and 'Car', respectively). Compared to all Internet users in Denmark, there are too many respondents in the age group 20-39 and too few in all other age groups.

In the sample, 23% are single. This is not significantly different from the sample among Internet users. The sample thus also contains too many not-singles compared to the population of Danish Internet users. 31 respondents or around one third of the sample are women. The

women in the sample are almost exclusively from the newsgroup 'Dogs' which makes this newsgroup different from the other newsgroups, as these are male-dominated. The respondents from the newsgroups tend to come more from east of Great Belt than the respondents in the general sample do and, therefore, also more than the average Internet user.

There were found no differences between the newsgroups in terms of geographic dispersion. There were also found no interesting differences in level of income or level of education neither between the sample from the newsgroups and the sample among Internet users in general nor between the 3 newsgroups.

It is not possible to evaluate the newsgroup sample in terms of representativeness compared to the population of newsgroups users, but only to compare the sample of newsgroups users with the sample among Internet users in general. The comparison in terms of demographic profile shows that the members of the newsgroups in the sample are, on average younger, than the average Internet user, and tend to live more on Zeeland than the average Internet user. The demographic data also show that the members of the 4 selected newsgroups differ quite much in terms of age and gender, and it seems that off-line differences in interests between gender and age newsgroups are reflected in the newsgroups. The members of newsgroups do not differ significantly from each other or from the general Internet user on marital status, level of income, or education. This shows that the use of newsgroups is not a phenomenon restricted to certain socio-economic segments.

5.3.2 Interest in and Knowledge about the Product

The fact that newsgroups users are not special in terms of most socio-economic measures does not hinder that newsgroup users can differ from the average Internet user in other ways. One of these ways could be the level of interest in the product discussed in the newsgroup, as the newsgroup gives the consumer an opportunity to meet others sharing the same interest across large distances in time and space. The numbers from the sample (reproduced in appendix C.2) show that the members of the selected newsgroups consider themselves more interested in and knowledgeable about the product class, which is the topic of the news-

group than the ‘average person’. At the same time, they consider themselves at the same knowledge and interest level as the other members of the newsgroup. Discriminant analysis using groups as classification variable revealed that there are differences between the newsgroups with respect to interest in the product. The difference is due to a very high interest in the two newsgroups ‘Car’ and ‘Dog’ and a close-to-average interest in the newsgroup ‘House and Garden’. This is as expected from the differences in the type of product, on which the newsgroups centre. When it comes to knowledge about the product, there is a large significant difference between the members of the ‘Dog’ newsgroup and the ‘Car’ newsgroup. A closer look at the numbers reveal that the members of the ‘‘Dog’ newsgroup on average feel that they know less than the other members of the newsgroup, whereas the members of the ‘Car’ newsgroup on average feel that they know more than the other members of the newsgroup. Maybe this picture is related to the difference in gender dominance in the two newsgroups. Again, this is based on subjective prejudice.

Results from the samples taken in the newsgroups confirm that members of the newsgroups, on average, perceive themselves as more interested in and knowledgeable about the product than does the average Internet user. At this point, we do not know if this is related to the motivation to visit the newsgroup or to search for information in the newsgroup.

5.3.3 Importance of the Internet in the Consumer’s Life

In the former chapter it was established that respondents in the newsgroup sample are more involved in the product, on which the newsgroup centres, than is the average Internet user. Chi-square analysis shows that the respondents in the newsgroups sample have also used the Internet longer than the respondents in the sample among Internet users in general. As could then be expected from the Internet Search Model, the respondents in the sample use the Internet significantly more than does the average user of the Internet and perceive higher benefit from searching the Internet for information. There were found no significant differences in the relative use for work or private purposes. There were also found no significant differences between the groups on this issue.

5.3.4 Relation to the Selected Newsgroups

Table 5.7 shows the period of time persons in the sample have used the newsgroups in question. The table shows that most of the respondents have used the newsgroup for more than a year, and at the same time, quite a large part of the sample has used the newsgroup for less than a year. This confirms that there are persons in the newsgroups, who are long-time users, while others float in and out of the newsgroups. There were found no significant differences between the newsgroups in this study in terms of length of membership in the newsgroup.

Table 5.7: Length of membership in the newsgroup

	Percentage	Number of respondents
Less than one month	1.91	3
Between one month and six months	21.36	22
Between 6 months and 1 year	17.48	18
More than a year	57.28	59

As it is possible to be a member of the newsgroup without actually using the newsgroup, tie to the newsgroup is probably more truly measured in terms of how often the respondent logs on to the newsgroup or how much time, he spends there, once logged on, as discussed in chapter 3.4.2. Table 5.8 below shows how often the respondents log on the newsgroup. The table shows that about two thirds of the respondents log on to the newsgroup at least once every day, and most of the respondents use the newsgroup each week. This shows that the newsgroup plays a significant part in most of the respondents' life. This number may be estimated too high because of the sampling method, resulting in too many active members of the newsgroups in the sample. If we, in spite of this reservation, look closer at the data, we find that there are significant differences between the newsgroups in terms of how often the members log on to the newsgroups: The users of the tighter knit newsgroups on products with high self-centrality to the consumer, 'Dogs' and 'Car', log on to their newsgroups more often than the members of the newsgroup 'House and Garden' do. This again confirms hy-

pothesis 12 that newsgroups on products which have higher self-centrality are tighter knit than newsgroups on products with low self-centrality.

Table 5.8: Frequency of Log-on to the newsgroup

	Percentage	Number of respondents
Once a month and less	4.95	5
About once every 2 weeks	2.91	3
About once a week	7.77	8
Several times each week	27.18	28
Once every day	32.04	33
More times every day	22.33	23

There is no connection between length of membership and frequency of logging on to the newsgroup. The explanation for this can be that, even though some members are only members for a short period of time, they may visit the newsgroup quite often, once they are members.

Frequency in logging on to the newsgroup is but one dimension in use of the newsgroup. Therefore, the respondents were also asked to rate how much time they spend in the newsgroup, once they have logged on. Table 5.9 below shows the distribution of the answers to this question. The table shows that, by far, the most common time spent there is between 5 and 15 minutes. No significant differences were found between how often the respondent logs on and the time spent, once logged on. There were differences between the newsgroups, as the respondents in the newsgroup ‘House and Garden’ were logged on for a shorter time span than the members in the newsgroup ‘Car’.

Table 5.9: Time spent pr. Logon

	Percentage	Number of respon-dents
Less than 5 minutes	17.48	18
Between 5 and 15 minutes	46.60	48
Between 15 and 30 minutes	16.50	17
Between 30 minutes and 1 hour	7.77	8
More than 1 hour	8.75	9

The above shows that the members of the different newsgroups are not equally attached to the newsgroup they are a member of, neither in terms of length of membership nor in time spent in the newsgroup. There seems to be a tendency that the members of the newsgroups on products with high self-centrality to the consumer, use the newsgroup on a more regular basis than the members of the newsgroup 'House and Garden', confirming hypothesis 12.

5.3.5 Membership Segments in the Sample

According to Kozinets (1999), members of newsgroups can be segmented into four groups based on the self-centrality of the consumption activity to the consumer and felt tie to the newsgroup. The respondents in the sample were assigned to the 4 four segments as follows: The two axes (see figure 2.7) were self-centrality of the product - in this survey measured by the two questions on interest in the product class – and tie to the newsgroup – in this survey measured by how often the respondent logs on to the newsgroup. A high value is defined as a value above average in the sample in total. The distribution of the segments in the newsgroups is shown in 5.10 below.

Table 5.10: Number of members in the segments in each newsgroup

	House and Garden	Dogs	Car	Consumer
Insider	5	19	19	3
Mingler	11	0	4	6
Tourist	8	1	5	4
Devotee	7	5	4	1

The newsgroups differ, as expected from the results presented in the previous chapters, significantly in composition[21] in terms of member segments. In the newsgroups 'Dogs' and 'Car' there is a strong overrepresentation of Insiders compared to the other newsgroups, whereas Tourists and especially Minglers are overrepresented in the newsgroups 'Consumer Issues' and 'House and Garden'. This result is in accordance with the results from the previous chapters.

5.3.6 Summary of the Description of the Newsgroups

In comparison to the average Internet user, users of the selected newsgroups are skilled Internet users who are interested in the product, on which the newsgroup is centred. The newsgroup members, on the other hand, are not different from the average Internet user when it comes to income level and education. Therefore, there is reason to believe that newsgroups will be used by more consumers as time goes by and the younger persons who, as we learnt from the Internet Search Model, use the Internet more, get older, and as more and more users get to be experienced Internet users.

The newsgroups were found do be composed differently. As expected, members of the newsgroups centred on products with high self-centrality to the consumer were more interested in the product, and also used the newsgroup more than the members of 'House and Garden'. There were also quite a large gender difference in the composition of the newsgroups, as one newsgroup seems to be female dominated newsgroup, whereas the 'Cars' newsgroup is male dominated.

[21] The test was made both with and without the consumer group as this group contained too few respondents for a valid test.

Hypothesis 12 stating that newsgroups centred on products with high self-centrality to the consumer are tighter knit than other newsgroups was supported by the results from the sample, as it was by the results from the observation data. Members of the 'Dog' and the 'Cars' newsgroup both use the newsgroup more often and log on the newsgroup for a longer period of time on each visit than do members of the newsgroup 'House and Garden'.

As the following chapters test the hypotheses on motivation for and use of the newsgroup, which different segments will exhibit, respondents were assigned to segments in chapter 5.3.5. The distribution of the segments in the selected segments was as expected from the previous results. In the newsgroups on products with high self-centrality to the consumer, there was an overrepresentation of 'Insiders'.

5.4 Motivation for Entering the Newsgroup

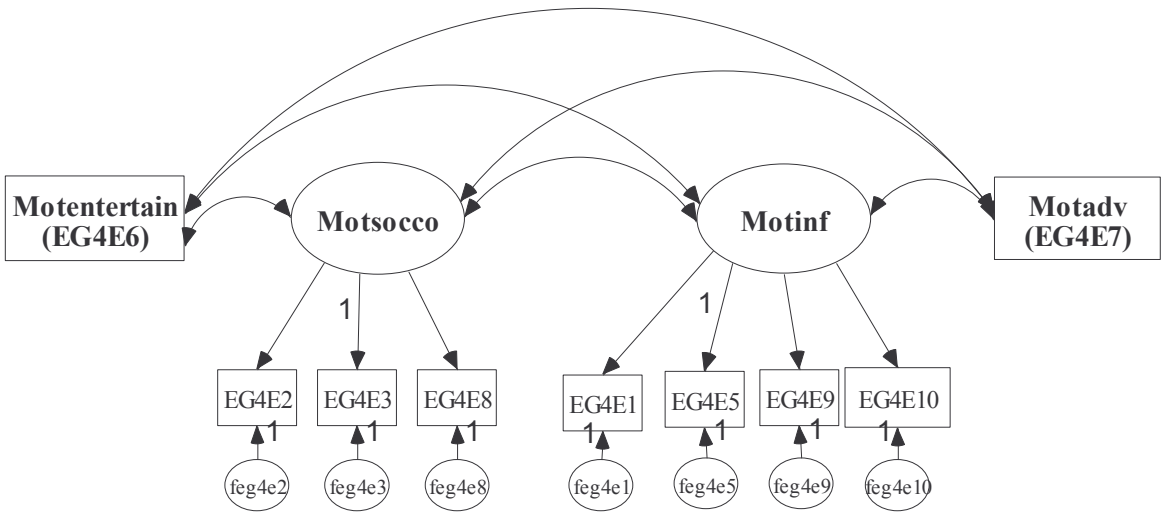
The previous chapter has confirmed that members of newsgroups differing in terms of type of product centred on use these newsgroups differently. The pattern suits the expectations based on the differences in type of product. However, different consumers in the same newsgroup may still have different motives for visiting the newsgroup. Hypothesis 13 stated that consumers enter newsgroups for 4 different reasons, and hypothesis 14 in chapter 2.6 stated that these reasons are related to the tie the consumer feels to the newsgroup and to the tie the consumer feels to the type of product.

In order to investigate this, it is necessary first to test whether the 4-factor model proposed by Kozinets and put into measures by me is a model well-suited for explaining the variation in the data. This is the topic of chapter 5.4.1. After this initial test of the measurement model, we are ready to test whether correlations between the motivation factors can be explained by variations in perceived self-centrality of the product and tie to the newsgroup, measured as frequency in using the newsgroup.

5.4.1 Model of Motivation for Entering the newsgroup

The 4-factor model of motivation for entering newsgroups is shown in figure 5.1. The four factors in the model are entertainment (motentertain), socializing (motsocco), getting information (motinf), and giving advice to others (motadv). The elements are all expected to correlate positively, except getting information and giving advice which are expected to be oppositely correlated.

Figure 5.1: Model of motivation



The parameters in the model were tested using AMOS. Before this estimation, the scales were evaluated for usability in the analysis. The values calculated for Cronbach's alpha are reported in appendix C.2. As one of the variables measuring information gathering was quite skewed, this variable was kept out of the analysis (eg4e). Another variable was also skewed. However, leaving this variable out of the scale has a very deteriorating effect on the size of Cronbach's alpha (from 0.69 to 0.57) and, therefore, it was decided to keep the variable in the scale.

After this preliminary analysis, the estimation was made. Detailed results are reproduced in appendix F.1. The result of the empirical test of the model was a χ^2-value of 35.518 with 23 degrees of freedom producing a p-value of 0.058. This means that the model fits the data quite well. There are no offending values in terms of negative variances or

large standardized residual covariances. However, RMSEA is 0.07, which is just acceptable, and the modification indices show a possibility for modification. It seems that eg4e1 may load on more than one latent variable. An examination of answers to the items in this scale measuring motivation by need for information shows that the reliability of the scale is about the same whether the problematic variable is in the scale or not. Therefore, the next step was to try and leave this variable out of the scale. The alpha values for 'Motivated by Possibility to Socialize' and 'Motivation by Possibility to get Information' are hereafter 0.84 and 0.69, respectively. Both are at an acceptable level.

The modification resulted in a χ^2-value of 20.991 with 16 degrees of freedom corresponding with a p-value of 0.179. After the alteration of the model, the p-value is thus rather large**.** GFI has augmented to 0.95 (compared to =. 925), CFI to .981 (compared to 0.959), and RMSEA to 0.055. These results indicate acceptable fit, close to what is normally regarded as values for good fit.

As the number of estimated parameters is 20, the sample size is adequate, and there are still no offending estimates. Therefore, this measurement model is accepted as a description of the factors explaining motivations for entering the consumer oriented newsgroups.

The correlations between the latent variables in the model are shown in table 5.11 below:

Table 5.11: Correlations from the confirmatory factor analysis of model of motivation factors.

Correlation between		Standardized Estimate	Critical ratio	p-value
Motivated by opportunity to get information (motinf)	Motivated by opportunity to socialize (motsocco)	0.329	2.296	0.02
Motivated by opportunity for entertainment (motentertain)	Motivated by opportunity to socialize (motsocco)	0.529	4.405	<0.001
Motivated by opportunity for entertainment (motentertain)	Motivated by opportunity to get advice (motinf)	0.273	2.098	0.04
Motivated by opportunity to give advice (motvadv)	Motivated by opportunity to get advice (motinf)	0.374	2.639	0.01
Motivated by opportunity to give advice (motadv)	Motivated by opportunity to socialize (motsocco)	0.409	3.579	<0.001
Motivated by opportunity for entertainment (motentertain)	Motivated by opportunity to give advice (motadv)	0.427	3.969	<0.001

All correlations are significantly different from zero. They are all positive, meaning that the motivation factors are all positively correlated. The strongest correlation is between the entertainment motive and the social motives. These motivation factors thus seem to be closely related. Large is also the correlation between the advice motive and the entertainment motive, and the social and the advice motive. The information search motive is also positively related to the other motives but much less strong than they are. The correlations are as expected, except for the motivation by needing advice and the motivation to give advice, which are positively correlated, indicating that giving and asking for advice are both means of communicating about products in the newsgroup and are

not perceived as opposites by the consumers in the newsgroup. The good fit of the model confirms hypothesis 13 that the model can be used as a measurement model for motives for entering newsgroups.

5.4.2 Explaining Differences in Motivation for Entering

After confirming the adequacy of the measurement model, it is time to test hypothesis 14 which states that the distribution on the motivation factors for entering the newsgroups differs across the member segments defined by Kozinets (1999). As the motives are positively correlated, they are not contradictory, but merely different aspects of motivation.

Following the thoughts in Kozinets (1999), the details in the expectations from hypothesis 14 are as follows:

- Insiders are very interested in both the product and the newsgroup, and should, therefore be more motivated by the opportunity to give advice than all members of all other segments, and more by the opportunity of social contact in the newsgroup than Tourists and Devotees.

- Tourists are primarily interested in getting information from the newsgroup, and should, therefore, stand out from the rest of the segments on this variable.

- Devotees are mostly interested in the product and, therefore, Devotees are expected to enter the newsgroup primarily for the entertainment of reading about the product.

- Minglers are primarily interested in newsgroup interaction and should, therefore, be motivated by the ability to socialize more than the other segments are.

In order to investigate whether these hypotheses hold true, I ran a discriminant analysis with the summated scales acting as explanatory variables for segment membership to see whether the motivation elements discriminate between the newsgroups as expected from Kozinet's

framework. Because of the limited size of the data set, it was decided to do the analysis this way, using the standardized summated score on the variables on the relevant questions as independent variables. The model thus consisted of the 4 motivation factors discriminating between the 4 segments. The result of this test was that only one variable discriminated significantly between the newsgroups: Social contact. On this motivation factor, the Insider segment had a significantly higher value than both the Mingler and the Tourist segment, whereas the Devotee segment displayed a value close to the Insider but significantly larger than the Tourist. A further look into the variables showing non-significant differences between the segments revealed that on all motivation variables persons belonging to the Insider segment showed the highest average value while persons, belonging to the Tourist segment, showed the lowest value whereas the two remaining segments lie between these two extremes. This shows that the Insider segment has a higher level of motivation for entering the newsgroup than all other segments, and that the opposite is true for the Tourist segment. This confirms that the segments differ in motivation for entering the newsgroup, and that the more devoted, a person is to the newsgroup and the product, the more motivated he feels for entering the newsgroup.

As the members of the newsgroups seem to driven by a different level of motivation and we are also interested in the relative weight of the 4 motivation factors, I investigated whether there was a difference between the segments in terms of the relative weight, the members of the segments place on each motivation factor. The weight is calculated as the sum of the standardized, summated scores on each motivation factor. These weights are shown in table 5.12 below:

Table 5.12: Relative weight of motivation factors for the 4 segments

Motivation factor / Segment %	Getting information	Socializing	Entertainment	Giving Advice
Devotee	0.27	0.24	0.25	0.22
Insider	0.27	0.22	0.28	0.23
Mingler	0.28	0.19	0.28	0.25
Tourist	0.32	0.17	0.27	0.24

Overall, the 4 motivation factors have about equal weight, although especially the socializing, but also the giving advice dimension, has smaller weight than the getting information and entertainment factors. The numbers in table 5.12 show that, with respect to the motivation factors 'Entertainment' and 'Giving Advice', there is almost no difference between the segments. The largest difference is found on the 'Socializing' factor, on which the Tourist segment, as expected, places less emphasis compared to the members of the Devotee and Insider segments. On the motivation factor 'Getting information', we see almost the opposite picture. The Mingler segment has a relatively low value on the 'Socialization' factor. This is opposite to the expected, as members of this newsgroup spend a relative large amount of time in the newsgroup, but are not very interested in the product. Maybe the explanation lies in the wording of the variables measuring this motivation factor, as two of the items emphasized the chance to be with others who share the same interest. Applying analysis of variance using the generalized least squares method for estimation to the relative weight of each factor, using segment as classification variable, reveals that the difference in the importance of the motivation factors is significant for two of the motivation factors (Output is in appendix F.1.3): 'Getting Information' and 'Socializing'. The Tourist segment is more motivated by the possibility of getting information than the other segments, and less motivated by the opportunity of getting in contact with other people than both Insiders and Devotees. This again confirms hypothesis 14 which states that different segments are differently motivated for entering the newsgroup.

As women and men may be motivated differently, a two-sided Analysis of Variance with gender and segments as classification variables was carried out on each of the motivation factors' weights. This analysis showed the same significant differences between the segments as the one-sided analysis of variance, plus significant differences for women and men on two motivation factors: Socializing and Entertainment. Women are more motivated by the possibility of contact and men are more motivated by the possibility of getting entertained. There is a significant interaction between gender and segment on the entertainment factor. A look into the data reveals that the larger difference between the two genders is in the Mingler and the Tourist segments. These are the segments not very interested in the product. An interpretation of this could be that women are less motivated to enter a newsgroup for entertainment if they are not interested in the product than are men. However, as most of the women belong to one newsgroup with a product with an, on average, high self-centrality to the consumers, it is difficult to know if the result is a gender difference or a difference stemming from differences in the product. On the other hand, the men came from three newsgroups which differ in terms of type of product and, therefore, there is reason to believe that the gender difference is truly there.

The same analysis as the above performed for newsgroups instead of segments show that members of the 'Dogs' and 'Car' newsgroups are generally more motivated to enter the newsgroups than are members of the newsgroups 'House and Garden' and 'Consumer Issues'. This is not surprising, as we earlier saw that these newsgroups contain more members who are Insiders than do the other newsgroups. Also, members of the newsgroup 'Consumer Issues' single out as not motivated, compared to the other groups. On the relative level, only the socializing factor singles out as bearing significantly different weight in the newsgroups with the socializing factor being more important in the newsgroup 'Dogs' than in the newsgroups 'House and garden' and 'Consumer Issues'. Still, remembering that the respondents from the 'Dogs' newsgroup were all women, it may be a gender factor more than a newsgroup factor, we see in these results.

5.4.3 Different Motives for Different Segments?

The analyses carried out in this chapter were targeted to test hypotheses 13 and 14 on the adequacy of the proposed measurement model for motivation factors to enter newsgroups and that different segments in the newsgroups are motivated differently to enter the newsgroups.

The confirmatory factor analysis did not reject hypotheses 13 and, therefore, the 4-factor model was deemed adequate for measuring motivation to enter newsgroups on the Internet related to exchange of consumer information.

The analysis showed that members of the Insider segment generally are the most motivated to enter the newsgroup, whereas members of the Tourist segment generally are the least motivated to enter the newsgroups, of which they are a member.

On the relative level differences were found between the segments on two of the four motivation factors: The possibility of getting information and the possibility of getting in contact with other people who share the same interest. The Tourist segment was significantly higher motivated by the opportunity to get information than the other segments and significantly lower motivated by the contact factor than the Insider and the Devotee segment. In the analysis, the Devotee and the Mingler segments seem to have switched places on the contact factor. This may be because of the wording of the variables measuring the contact dimension as two of these emphasized the shared interest in the product.

Additional analysis bringing gender into the analysis did not change this picture, and additionally revealed that women, when controlling for segment, are more motivated by the opportunity to socialize and less by the entertainment factor than men are. Especially, men in the Mingler and Tourist segments are more motivated by the opportunity to get entertained than the women in these segments are.

The above all in all confirms hypothesis 14 stating that different segments are motivated differently to enter the newsgroups and also supports the expectation that the Insider segment would be generally more

motivated to enter the newsgroup and that members of the Tourist segments mostly visit the newsgroup in order to get information they need.

In order for the newsgroup to keep being interesting to visit to exchange consumer information, there has to be consumers who are willing and able to share the needed information in the newsgroup with the members who visit the newsgroup mainly to acquire this information. This is the topic of the next chapter.

5.5 Opinion Leadership and Communication

We have already seen both in the analysis of the data from the sample among Internet users in general and in the preceding section in this chapter that there are individuals who visit newsgroups because they seek consumer information there. We have also seen that in the newsgroups, there are individuals, who perceive themselves as knowledgeable about the type of product in question. What we do not know is whether these knowledgeable members of the newsgroups are motivated to give the needed information to the persons, seeking this information – in other words, if they are what are traditionally termed 'Opinion Leaders'[22]. Hypothesis 15 states that the communication in the newsgroup can be explained by different motives to visit the newsgroup, fitting each other well. The motives are governed by different types of personality in terms of seeking and wanting to give advice. The aim of this chapter is to test this hypothesis.

In order to test the hypothesis, it is necessary that the data set contains both opinion leaders and opinion seekers. As opinion leadership may be a desirable trait to possess, it will strengthen the conclusion if the self-designation of opinion leadership can be confirmed by others naming these persons as opinion leaders. Furthermore, it is interesting to see if the persons who are named opinion leaders in the selected newsgroups, are more active in the newsgroups than the average member is. Therefore, this chapter starts out by a description of the sample and the observation data in terms of opinion leadership.

[22] The theory behind this was discussed in chapter 2.4.1

After this initial validation and description of the data in terms of opinion leadership, it is almost time to test the proposition in hypothesis 15 in the form of a test of a structural model containing the elements, and relations between them, as expected from hypothesis 15. Before the test of this model, is it, however, necessary to test the measurement model for opinion leadership and opinion seeking proposed in chapter 3.4.6 on measurement of these latent variables. The tests are carried out following the procedure described in appendix D.

5.5.1 Opinion Leaders in the Study

Opinion leadership was measured in two ways: By self-designation of opinion leadership using a multi-item measurement scale and by the respondents denominating other members of the newsgroups as influential to them. In order to be able to evaluate whether the persons who, themselves, indicate that they are influential actually are influential, a sufficient number of the persons named as opinion leaders should also appear in the sample. The validation criterion then is that respondents considering that they are influential should be named as influential by others.

55 different persons were designated to be opinion leaders by others. As can be seen from table 5.13 below, 16 of these 55 persons are represented in the sample, and 33 are on the observation list. There are large differences between the newsgroups in terms of overlap between named opinion leaders and respondents in the sample. In the newsgroup ‘House and Garden’, 9 persons were named opinion leaders. Among these 9 persons, only one was identified in the sample. Almost the same picture applies to the newsgroup ‘Car’. In the remaining two newsgroups, the sample is more valid from the perspective of the ability to check whether persons, who are named opinion leaders by others, also considered themselves to be so, and vice versa.

Table 5.13: Index of activity of named opinion leaders, all news-groups

Index number for named opinion leaders:	Number of persons named	% of news-group members named	Number of persons on the observation list	Number of persons named act as respondents	Number of posting/person relative to news-group	Average number of threads/person relative to news-group	Average posting pr. thread /person relative to news-group
Consumer Issues	4	0.02	3	3	7.60	5.36	1.43
House and Garden	9	0.06	8	1	2.33	2.31	0.86
Car	25	0.12	11	1	3.59	4.13	1.17
Dogs	17	0.12	11	11	5.92	3.75	1.40

The validation of opinion leadership is difficult because of the lack of overlap between the two samples. In the sample, 20 individuals (about 20 percent of the sample), are above average on self-designated opinion leadership. Of these, only 4 are mentioned by other respondents in the sample. Also, 10 of the individuals, who have been named opinion leaders, are below the average on opinion leadership measured by self-designation. These results indicate that it may be difficult for respondents to know whether they are opinion leaders. Another explanation can be that some individuals may be opinion leaders on the product class in general, but not in the newsgroup. This, for instance, is the case, if the person on the general level posses much product knowledge, but that the knowledge, relative to the newsgroup, is on average. A third, quite plausible, explanation for the lack of overlap is the wording in the questionnaire, as the question was asked about other persons having influence on the respondent, and not in the newsgroup as such. Nonetheless, no matter what the explanation is for this discrepancy between self-designated opinion leadership and other-designated opinion leadership,

it is important to remember that the only thing measured in the questionnaire is the perceived level of opinion leadership, and not an objective measure of opinion leadership. This means that there is a danger that persons who give a lot of advice in the newsgroup and perceive themselves as opinion leaders are, in fact, trusted by no-one in the newsgroup.

Returning to the numbers in table 5.13, we see that the individuals whom others have named as opinion leaders, exhibit a much higher rate of communication in the newsgroup than the average user. This high level of communication is mostly due to a participation in far more discussions (threads) than the average users. The number of discussions participated in is between 2.31 and 5.36 times the number for the average user. This indicates that the newsgroups, at least in part, is kept going by a small number of very active members, who are willing to answer to the questions others post to the newsgroup. This supports hypothesis 15 even though it is difficult to conclude if the other members of the newsgroups follow the advice given by these people.

Returning, once again, to the numbers in table 5.13 we see that, in the tighter knit newsgroups, a larger percentage is named as opinion leader. It seems plausible that in these tighter knit newsgroups, individuals feel that they know the persons who advice them, even though they sometimes only know these persons by alias. How influential are, then, the 55 persons, who are designated as influential to other persons? The most common level of influence is 'some' (51%). However, in 12% or 12 of the cases, the named person has had decisive influence on the choice made. 8 of these 12 cases were in the 'Dogs' newsgroup, 1 in the newsgroup 'Car' and 3 in the 'House and Garden' newsgroup. As expected, the influence is largest in the tighter knit newsgroup 'Dogs. Except for the 'Dogs' newsgroup, there has been no additional contact to the opinion leaders neither by telephone nor in person. In the 'Dogs' newsgroup, 7 individuals have met the opinion leader they refer to in person. One third of the persons naming opinion leaders have been in direct contact with the opinion leader via e-mail. These results confirm that the life of the newsgroup is in the communication in the newsgroup via the newsserver, and not via bilateral contacts outside the newsgroup.

In the data from the sample from the selected newsgroups, opinion leadership is, as generally assumed, linked to product knowledge. As product knowledge is closely related to product interest, it was not surprising to find out that the score on opinion leadership was higher in the Insider segment than in the other segments. This means that the consumers, who are visible in the newsgroup because they often answer questions, are also the individuals who are frequent users of the newsgroup.

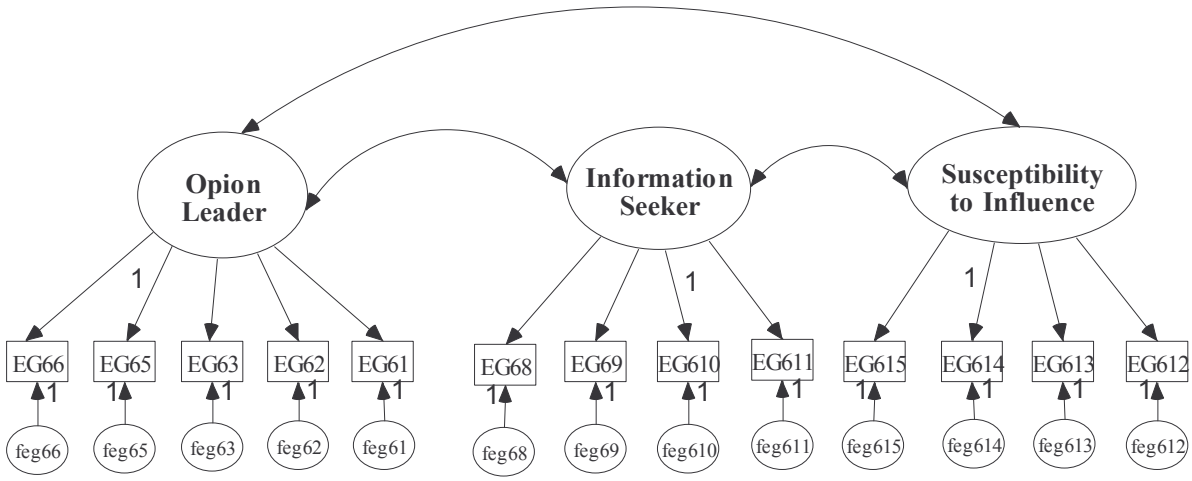
In total, this initial description of opinion leadership in the newsgroup has confirmed that there are individuals in each newsgroup, whom others think have or are able to influence them in their choice of product and that these persons are far more active in the newsgroup in question than the average user. The data also showed that there seems to be a larger percentage of such persons in the tighter knit newsgroups than in the looser knit newsgroups. This, once again, confirms the picture of the different segments in the newsgroups and their different roles. However, a word of caution is needed, as the validation of the measurement of opinion leadership by self-designation showed that the only thing which we are sure of measuring is perceived level of opinion leadership, and not true opinion leadership.

5.5.2 Test of the Measurement Model

After this initial description of opinion leadership in newsgroups and initial test of the validity of the measures, the time is ready to test the validity of the measurement scale on the operational level.

The initial measurement model for opinion leadership as described in chapter 3.4.6 is shown in figure 5.2 below:

Figure 5.2: Measurement model, Opinion leadership



According to the measurement model, 3 factors are assumed to capture the concept of opinion leadership. Opinion leaders are supposed not to be susceptible to influence (negative or no covariance) from others, while opinion seekers are (positive covariance) susceptible to influence from others. The theory behind the model indicate no expectation in regard of the relation between being opinion leader and opinion seeker, as each individual can be neither, one, or both according to this theory.

The parameters in the model were tested using AMOS. Before the estimation was made using Maximum Likelihood estimation, the scales were evaluated for usability in the analysis, as described in appendix D. The values calculated for Cronbach's Alpha are reported in appendix C.2. The variables measuring opinion leadership all meet the criteria stated in appendix D. The alpha value for the scale is 0.68. This is a little lower than the preferable level. Because of this, and because the number of data points are quite high compared to the size of the sample, the output from the calculation of Cronbach's Alpha was inspected in further detail. This inspection revealed that omitting the two variables for which the questions were asked in the opposite direction of the other variables (eg64 and eg67) will result in an Alpha value of 0.85. This is a very satisfactory value. Therefore, the measurement model was tested without these variables. The variables measuring opinion seekership all meet the criteria and the scale has an Alpha value of 0.72, which is acceptable for the scale to be used in the test of the measurement model.

The variables measuring susceptibility to influence pose more problems, as they are all skewed to the right. Two of the variables (eg616 and eg617) are more severely skewed than the others. These variables were, therefore, left out of the analysis. The alpha value was 0.78 after this omission. As this is a satisfactory value and the variables are all skewed in the same direction, the latent variable is used in the test of the measurement model but the results have to be interpreted with caution.

The test of the initial measurement model after adjustment of the scales resulted in a χ^2-value of 118.275 with 62 degrees of freedom and an RMSEA of 0.094 (detailed output in appendix F.2). This is not an acceptable level of the fit indices. An inspection of the modification indices revealed that two of the observed variables (eg65 and eg66) loaded on more than one of the latent variables. Therefore, these two variables were successively omitted from the model. The reliability of the scale is 0.86 after this omission. As the covariance between opinion leadership and susceptibility to influence was not significantly different from 0 and this is in accordance with the theory behind the model, this path was left out of the model.

The results from the following Maximum Likelihood estimation are shown in table 5.14 below:

Table 5.14: Fit indices for model measuring Opinion Leadership and –Seekership

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
47.159	42	0.27	1.123	0.924	0.987	0.035	0.670

All measures indicate excellent fit. The correlation between susceptibility to influence and opinion seeker is 0.417, indicating that at opinion seekers are much more easily influenced than opinion leaders are. The correlation between opinion seekership and opinion leadership is 0.433, indicating that the two traits are not contradictory, but supplementary, as proposed by Flynn et al. (1996). Both correlations are significantly different from 0. As the measurement model after these initial adjustments

seems to fit the data quite well, the next step is to investigate the hypothesized relation between opinion leader-/seekership, motivation, and exchange of consumer information in the newsgroup.

5.5.3 Explaining Interaction in the Newsgroup

As put forward in the beginning of this chapter, the target is to test hypothesis 15 which states that: 'The communication in the newsgroup can be explained by different motives to visit the newsgroup which fit each other well'. The motives are governed by different types of personality in terms of seeking and wanting to give advice. After testing the validity of the data and the measurement scales we are now ready to test a model containing the latent variables and relations put forward in hypothesis 15. As the sample size is small, the connection between level of opinion leadership, motivation to give advice, and answering to questions posted is first tested separately, as is the relations behind asking questions in the newsgroup. Before testing the models, it was ensured that the 'new' variables (eg4d1, 2 and 3) were not severely skewed and that the alpha level of the two-item scale, measuring inclination to answer questions, is at a sufficient level. The relevant data for this can be found in appendix C.2. After the partial tests, the full model is tested.

Before testing the models, it is interesting to establish the level of asking and answering questions in the newsgroups. Table 5.15 below contains the percentages of the sample which have posed and/or answered questions in the newsgroup when this has been relevant.

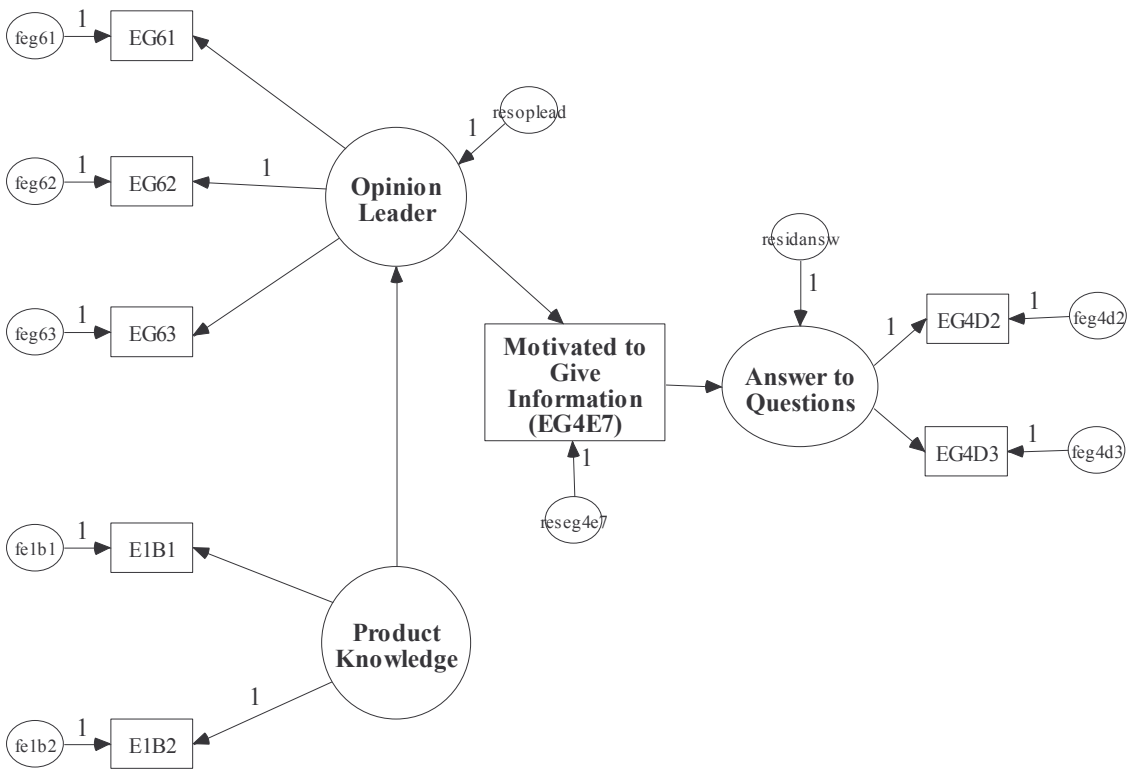
Table 5.15: How often questions are posed to and answered in the newsgroup by the respondents

Percentage	How often do you pose questions to the newsgroup regarding purchase advice if you have a relevant problem?	How often do you answer questions regarding purchase advice when you have the knowledge?	How often do you answer questions regarding purchase advice when you have experience with the subject?
Never	10.89	2.97	3.96
Seldom	33.66	13.86	11.88
Not often	12.87	11.88	6.93
Sometimes	26.73	41.58	34.65
Mostly	11.88	21.78	33.66
Every time	3.96	7.92	8.91

It is easily seen from table 5.15 that the respondents more often answer to questions than pose them to the newsgroup. This is the same picture as is seen in the observation data. About 15% of the respondents pose questions to the newsgroup mostly or always if they have a purchase decision to make which is related to the newsgroup, and as much as 26.73% pose questions sometimes. The percentages also show that the willingness to answer questions, posed in the newsgroup, is quite high as only around 25% state that they answer questions less than sometimes, provided that they possess the relevant knowledge. There was found no significant differences between the segments on this matter. However, there was a close to significant (0.06) difference in the tendency to answer questions when possessing knowledge about the issue. Insiders seem to be more inclined to answer questions than are Tourists and Devotees. There were no differences between the newsgroups on this matter. As we have now established that the major part of the respondents in the sample do seek or give consumer information in the newsgroups, we are ready to test the relations discussed at the beginning of this chapter.

In figure 5.3 below, a model assuming that the opinion leader's motivation to answer questions in the newsgroup explains his answering to questions in the newsgroup is shown. Perceived knowledge about the product is added to the model to confirm that persons who see themselves as opinion leaders also see themselves as knowledgeable about the product. That is, they believe that they have something to offer when giving advice to others. The model was developed based on the theoretical contributions put forward in chapter 2.4.1 on opinion-leadership, especially Flynn (1996). This theory states that opinion leaders tend to be knowledgeable about the product and that opinion leaders like to show that they are knowledgeable about the product. They can do this by answering questions posed by others, as assumed in figure 5.3.

Figure 5.3: Answering questions as a result of level of opinion leadership

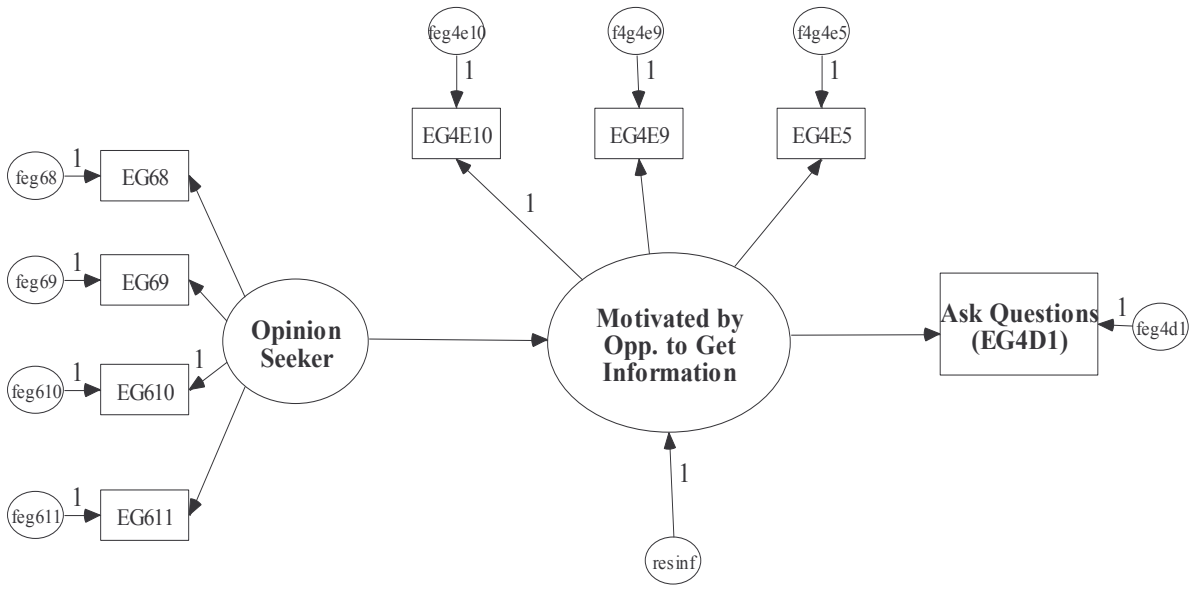


The model was tested using AMOS and proved to fit the data very well (detailed output in appendix F.3.1). χ^2 was 26.108 with 18 degrees of freedom, resulting in a p-value of 0.097. All other fit indices also indicated good fit and there were no offending estimates. All paths were

positive and significantly different from zero with positive loadings. The squared multiple correlation for opinion leadership was 0.285 showing that although a higher level of product knowledge leads to a higher level of perceived opinion leadership, many additional things influence perceived level of opinion leadership. The relation between being opinion leader and being motivated to enter the newsgroup to give advice is weaker, and the squared multiple correlation for giving advice is a mere 0.045. The squared multiple correlation for answering questions in the newsgroup was 0.27. This means that answering questions in the newsgroup can be explained, but only partly, by opinion leadership.

Like opinion leaders are expected to be motivated by the chance to give advice, opinion seekers are supposed to be motivated to join the newsgroup by the chance to get advice and, therefore, are expected to ask questions in the newsgroup. The expected relations are shown below in figure 5.4.

Figure 5.4: Asking questions as a result of possibility to get information



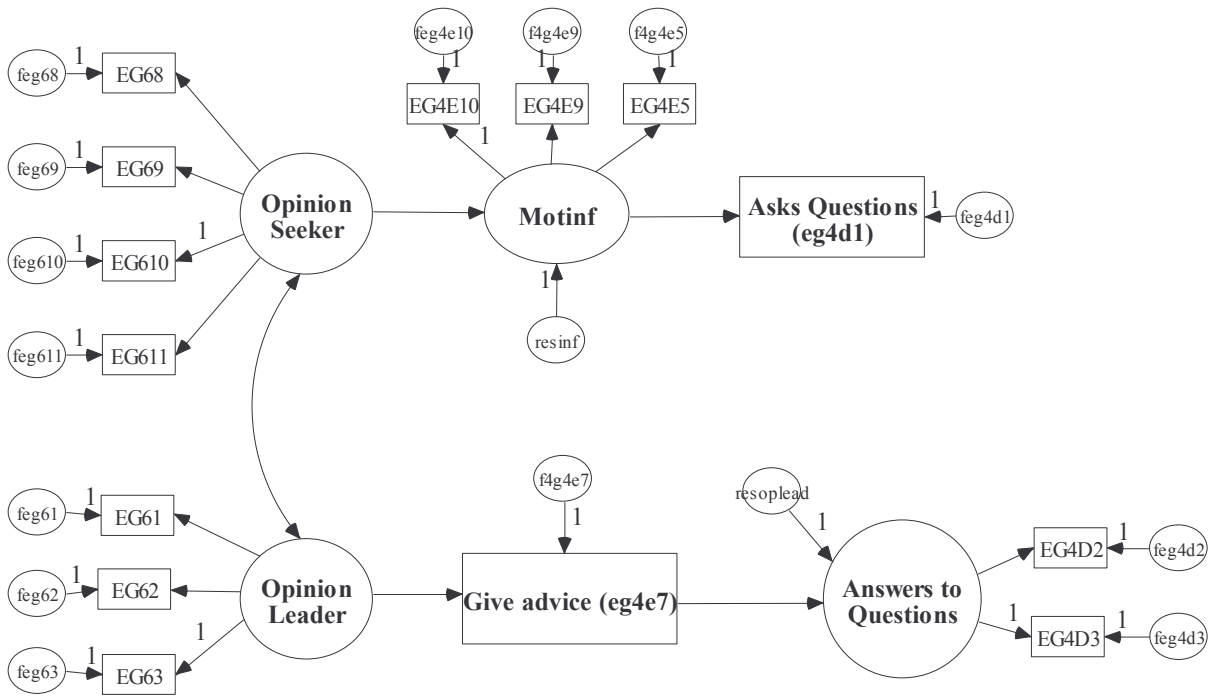
The model was tested using AMOS. The test resulted in a χ^2-value of 27.292 with 19 degrees of freedom which corresponds to a p-value of 0.098 (detailed output is in appendix F.3.2). All other fit indices also indicated good fit and the paths were positive, as expected. The squared

multiple correlation for motivation by opportunity to get information (motinf) was 0.16, and the squared multiple correlation for level of answering questions was 0.12, meaning that only 12% of the variation in answering questions in the newsgroup is explained by the model.

The test of the measurement model for opinion leader- and –seekership showed a positive relation between the two. As opinion leadership and opinion seekership is correlated as are the variables measuring inclination to answer and pose questions, the next step is to try and put the two models together to produce a fuller picture of information exchange in newsgroup explained by fitting motivations to ask and to answer questions, respectively as expected from hypothesis 15. When interpreting the data one has to be cautious as the number of parameters to be estimated is quite large compared to the sample size. Also, it is important to remember that the measures of the latent variables are subjective and not objective as was discussed in chapter 5.5.1. There are about 3 observations per estimated parameter and the target is 5 observations per estimated parameter. As a consequence of this, product knowledge was kept out of this final model.

The proposed model is shown below in figure 5.5.

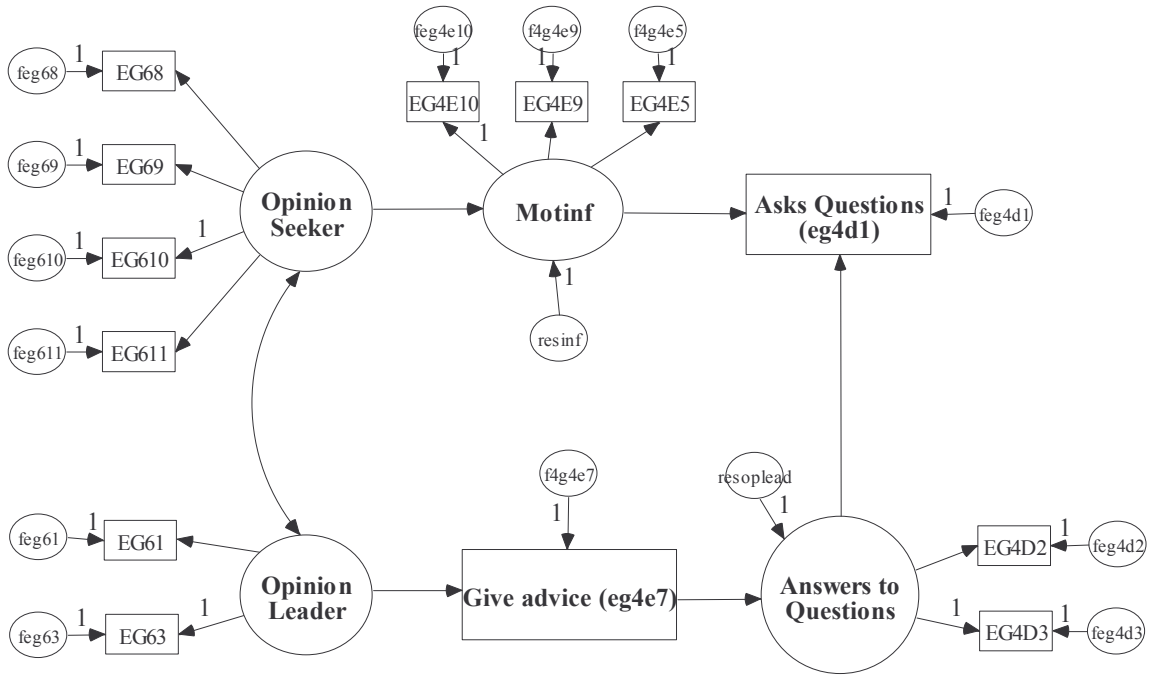
Figure 5.5: Model of newsgroup interaction as a result of opinion leadership



In the model, it is assumed that opinion leaders are motivated to enter newsgroups because they want to give advice to others and are, therefore, also inclined to answer questions whenever they feel capable of doing so. Opinion seekers, on the other hand, are motivated primarily by the opportunity to get information and, therefore, are more motivated to ask questions in the newsgroup.

A test of the model on the data from the newsgroups indicated that a path should be added from inclination to answer questions to the inclination to ask questions and that one of the observed variables loaded on more than one of the latent variables and should, therefore, be removed. After these adjustments, the model was as follows (output in appendix F.3.3 and F.3.3.1):

Figure 5.6: Model of newsgroup interaction after adjustments



The final reliability measures for this model are as follows in table 5.16:

Table 5.16: Final reliability measures, model of newsgroup interaction

Latent variable	Cronbach's Alpha
Opinion Seeker	0.72
Opinion Leader	0.74
Motivated by Opportunity to get Information	0.69
Answers Questions	0.84

The fit indices from the Maximum Likelihood estimation are as follows in table 5.17:

Table 5.17: Fit indices for Model of Newsgroup Interaction

(χ^2)	df	p-value	cmin/df	GFI	CFI	RMSEA	PCLOSE
86.062	61	0.019	1.411	0.894	0.931	0.063	0.232

All fit indices indicate good fit. In table 5.18 below, regression weights are displayed.

Table 5.18: Estimated regression paths for the Model of Newsgroup Interaction

Regression Weights Path to	Regression Weights Path from	Standardized estimate	Critical ratio	p-value (app.)
Answers	Motivated by Possibility to Give Advice (eg4e7)	0.503	4.979	<0.001
Ask Questions (eg4d1)	Answers	0.565	5.880	<0.001
Ask Questions (eg4d1)	Motivated by Possibility to Get Information	0.283	2.991	0.003
Motivated by Possibility to Give Advice (eg4e7)	Opinion Leader	0.320	2.738	0.006
Motivated by Possibility to Get Information	Opinion Seeker	0.396	2.705	0.006

All paths are, as expected, with a positive sign and all are significantly different from zero. The squared multiple correlation for the inclination to answer is 0.253 while the squared multiple correlation for inclination to ask questions is 0.408, indicating that the model captures only a part of the variation in inclination to ask and to answer questions in the newsgroup, respectively. There are no offending estimates.

The results show that in newsgroups, individuals who are in general opinion leaders enter the newsgroup motivated as expected from the general theory on opinion leadership: To give advice; and opinion seekers enter the newsgroup to get advice, also in accordance with the theory. The results also show that behaviour in the newsgroup is as expected from the theory on opinion seekership and opinion leadership: Opinion seekers tend to ask questions and opinion leaders tend to answer questions. Also, it seems that answering questions support the desire to ask questions in the newsgroup. In the newsgroup there are, however, 15% of the individuals or as much as 80% of the individuals who have higher than average score on opinion leadership who exhibit a high score on both opinion seekership and opinion leadership. These individuals thus must be very active in the newsgroup, asking and answering questions.

5.5.4 Conclusions on Communication Flow in Newsgroups

The analysis in the above shows that in newsgroups, as in the real world, individuals (members) have different roles in the spreading of consumer information: Some give information, some ask for advice, some do both and still others do neither. It seems that in the tighter knit newsgroups, there are a larger percentage of the members who are considered opinion leaders by the other members of the newsgroup than is the case in the loose-knit newsgroups. This adds to the expectation that the influence from these newsgroups and therefore also the substitution for sources formerly used, is larger than is the case in newsgroups, to which the members feel less attached. The newsgroups with a high level of attachment are the newsgroups centring on products with high self-centrality to the consumer.

The analysis also showed that the role in the communication flow between respondents which the consumer has outside the newsgroup is also reflected in the motivation to enter the newsgroup and the behaviour in the newsgroup. Respondents who perceive themselves as opinion leaders concerning the product class on a general level enter the newsgroup more than others, to give advice, and also answer questions more

than opinion seekers who enter the newsgroup to get advice and ask questions in the newsgroup more frequently.

In total, the analyses carried out in chapter 5.5 has supported hypothesis 15 stating that the level of communication in newsgroups can be explained by different motives to visit the newsgroup which fit each other well. The motives are governed by different types of personality in terms of seeking and wanting to give advice. Therefore, there is reason to expect that the analyses in the following chapter 5.6 are also going to support the two final hypotheses on influence in the newsgroups and substitution for other sources. As the analysis in the current chapter have confirmed that the communication in the newsgroups are a good example of fitting demand and supply of information, there is also reason to believe that the importance of newsgroups on the general level will grow, rather than diminish, in the future.

5.6 Replacement of Other Sources of Consumer Information

The preceding sections in this chapter presenting the results from the empirical study in the four selected newsgroups were all partial studies mainly aimed at underpinning the conclusions which it is the goal of this chapter to draw. The topic of this chapter is to test hypothesis 16 which states: 'Use of newsgroups primarily replaces marketer-dominated sources. This is more so prevalent for Tourists than for the other segments', and hypothesis 17 which states: 'Use of newsgroups has also replaced use of advice from primary reference groups. This replacement is larger for Insiders than for the other segments and larger in newsgroups on products with high self-centrality to the consumer than in other newsgroups'.

In the preceding sections in this chapter, we have learnt that the selected newsgroups are differently composed and that the composition is connected to the type of product, the newsgroup centres on. It is clear from the analyses that newsgroups, in which the topic is products with high self-centrality to the consumer, possess more members, to whom the newsgroup has high importance. This means that such newsgroups have more members from the Insider segment.

We have also learnt that Insiders are, on the general level, more motivated to enter the newsgroup, in which they are a member, than are the other segments. This group is opposite to the segment, Tourists, of which members are the lowest motivated on all factors. On the relative level, members of the Tourist segment are more motivated to enter the newsgroup by the opportunity to get information than are members of the other segments.

From the analysis of the premises for exchange of information in the newsgroup, we have learnt that, in the newsgroup, there are consumers who consider themselves knowledgeable on the type of product in question, who perceive themselves as opinion leaders on the type of product, and who are willing to give advice to others. These newsgroup members often are Insiders and, therefore, are at the same time active members of the newsgroups, keeping them alive by their conversation. We have also learnt that there are members of newsgroups, who are, in general, opinion seekers on the type of product in question and who are motivated to visit the newsgroup by the opportunity to get information As a consequence of this, these members are inclined to ask questions in the newsgroup and thereby participate in keeping the group alive.

Concluding to this, the partial analyses have shown that there is reason to believe that the following analysis will confirm hypotheses 16 and 17. Still, before looking at the rate of substitution of the sources, formerly used, by information search in newsgroups, it is relevant to find out whether the information from the newsgroups is considered to have any influence by consumers using newsgroups, as regards the choice of product, they finally make.

In chapter 5.5, we saw that the chosen newsgroups are, indeed, used for taking and giving advice in relation to purchase decisions, and for some of the respondents, the newsgroup is a source, often used. The question at this point is whether the advice is actually taken seriously. The answer to this question can be found in table 5.19 and 5.20 below.

Table 5.19: Chosen based on advice from news-group

	Percentage
Never	21.00
A few times	35.00
On several occa-sions	17.00
It hasn't been relevant	27.00

Table 5.20: Will choose based on advice from newsgroup

	Percentage
Never	12.00
Sometimes	45.00
Mostly	22.00
Certainly	19.00

The two tables containing prior and intended use of advice from the newsgroup show that as much as 17% of the respondents have chosen a product based on advice given in the newsgroup on several occasions and 35% have chosen a product based on advice given in the newsgroup a few times. This is quite a large percentage. The table containing intended behaviour (table 5.20) shows that the respondents are even more in favour of choosing based on advice coming from the newsgroup in the future than they have been in the past, as only 12% of the respondents indicate that they will never base a decision on advice from the newsgroup and as much as 22% indicate that they will mostly make their choice relying on information they have acquired in the newsgroup in question.

For some individuals, the advice given in newsgroups thus plays an important role in the choice of which product to buy. There seems to be a connection between the impact of the information from the newsgroup and the tie to the newsgroup. The percentage of respondents who indicate that they, on more occasions, have chosen a product according to recommendations from the newsgroup is twice as high for Insiders as for the other segments. Likewise, the percentage of Insiders indicating that they most certainly will choose products according to advice given in the newsgroup is about 3 times as high as for the other segments.

The above again supports that newsgroups, to some consumers, act as a resource of consumer information and, for some, information from the

newsgroup is even decisive as regards which product or brand to buy. The next question to ask then is which sources the newsgroup mainly replaces and if the larger impact of the advice from the newsgroup on members of the Insider segment is reflected in the rate of replacement of the types of information sources used in the past.

Table 5.21 shows the use and replacement of consumer information sources for the individuals in the chosen newsgroup. The table shows that before use of the newsgroup, respondents mostly used friends they saw often and sales personnel for consumer advice. Also family which the respondent often sees was a source of information used by many of the respondents. The table also shows that the newsgroup, on average, mostly has replaced marketer-dominated sources. This is as expected from hypothesis 16 and 17. The source which was formerly used the most was sales personnel. This type of source is also the type of source which has been replaced the most by information from the newsgroup. This may be because respondents, being more interested in and knowledgeable about the type of product in question, formerly have turned to sales personnel for information because these were the only source available possessing an adequate level of information. As the newsgroup seems to be able to provide information on a fairly high level and as this source is not marketer-dominated, it is not surprising that this replacement has taken place. However, when interpreting the marketing implications of this rather large replacement, we still have to remember that the percentage of the population of Internet users using newsgroups on a regular basis is quite small. Therefore, the impact on retailing is probably quite small at this point in time, but may very well be larger in the future.

Table 5.21: Use and replacement of consumer information sources

Source	Percentage used much or some	Percentage of these where the source is replaced at least partly
Friends, you see often	63.00	13.85
Friends, you see seldom	22.33	8.70
Family, you see often	51.46	16.98
Family, you see seldom	14.56	20.00
Colleagues at work	33.00	15.71
Sales personnel	71.84	63.16
Brochures	34.10	42.86
Sales Material received through mail etc.	30.10	35.50

rem: Almost none indicated full replacement.

When it comes to non-marketer dominated sources, the replacement is, as expected, smaller than for marketer-dominated sources. The source which was formerly used the most was friends, seen often, closely followed by family, seen often. This is in accordance with traditional theory, as these are informal primary reference groups. Colleagues at work, which is a formal primary reference group, have also been used quite much but, as was expected, less than the two other groups.

The next question to pose, when testing hypotheses 16 and 17, is whether the replacement is equal in the four segments defined in the framework proposed by Kozinets (1999). The following tables are used to investigate this question. As the sources friends and family, seen seldom, have formerly been used only a little, there is no separate analysis for these two groups. Also, there were found only minor differences in the use and replacement of sales personnel and colleagues at work as information sources. Therefore, the tables showing differences between the segments with respect to the replacement of these newsgroups are not displayed here.

Hypothesis 16 states that information search in newsgroups have replaced search using marketer dominated sources more for Tourists than for other segments. The next tables are designated to test this.

Table 5.22: Use and replacement of brochures as an information source

Segment	Percentage used brochures much or some	Percentage of these where brochures are replaced at least partly
Insider	30.43	35.71
Mingler	38.10	37.50
Tourist	31.58	83.33
Devotee	41.18	28.57

rem: Almost none indicated full replacement.

The numbers in table 5.22 show that prior to the use of newsgroups all segments used brochures about the same. The segments, however, differ in the rate of replacement. For Tourists, the use of brochures is almost fully replaced by the use of newsgroups whereas the replacement is about equal but at a much lower level for the other three segments.

Table 5.23: Use and replacement of mail distributed advertisements as an information source

Segment	Percentage used advertisements much or some	Percentage of these where advertise-ments is replaced at least partly
Insider	28.26	53.85
Mingler	28.57	16.67
Tourist	26.32	20.00
Devotee	42.18	28.57

rem: Almost none indicated full replacement.

Looking at the numbers in table 5.23 we see that, prior to the use of newsgroups, all segments used advertisements at about the same level except for Devotees who used advertisements much more than the other segments. The biggest replacement we find in the Insider segment, in which newsgroups, for half of the respondents, partially have replaced information search in advertisements.

The numbers in table 5.22 and 5.23 support the hypothesis that information search using the newsgroup has replaced active information search for Tourists more than for the other segments. However, this kind of information search has not replaced the more passive information acquisition gained from mail distributed advertisements. This, on the contrary, seems to be the case for the Insider segment.

Hypothesis 17 states that information search in newsgroups have replaced search using non-marketer dominated sources more for members of the Insider segment than it has for other segments. The next tables are designated to test this hypothesis.

Table 5.24: Use and replacement of friends seen often as an information source

Segment	Percentage used friends much or some	Percentage of these where the source friends have been replaced at least partly
Insider	69.56	6.25
Mingler	47.62	40.00
Tourist	63.16	8.33
Devotee	64.71	18.18

rem: Almost none indicated full replacement.

In table 5.24 above is displayed replacement of friends, seen often, as an information source when searching for information in relation to an upcoming purchase. The numbers in the table show that Minglers differ from the other segments in this respect. Formerly, Minglers used friends as advisors less than the other segments and at present, the newsgroup has to a great extent replaced this source of information. For members of the Devotees segment which by definition is interested in the product, but not much in the newsgroup, the newsgroup has replaced friends far more than it has for members of the Insider and Tourist segments. This only confirms hypothesis 17 in part, as it was expected from this hypothesis that friends as an information source would primarily be replaced for Insiders, and maybe Minglers, as these two newsgroups have the strongest ties to the newsgroup.

Table 5.25: Use and replacement of family seen often as an information source

Segment	Percentage used family much or some	Percentage of these where the source family have been replaced at least partly
Insider	47.83	18.18
Mingler	66.67	21.43
Tourist	42.11	12.50
Devotee	17.64	0.00

rem: Almost none indicated full replacement.

In table 5.25, the numbers for former use and replacement of family as an information source are presented. The use of family as an information source in relation to an ongoing decision process is at a much lower level than the use of friends is, expect for the Mingler segment, for which the shift is in the opposite direction. Again, we see that the replacement has been at the highest level in the Mingler segment, but at a lower rate than the replacement of the use of friends. Insiders are almost at the same level of replacement as Minglers, whereas replacement in the other two segments are at a much lower level, as expected from hypothesis 17.

Looking at table 5.24 and 5.25 in total, we are not able to find strong support to the hypothesis that the newsgroup has replaced the strong-tie sources more for Insiders than for the other segments. However, for the second segment whose members often visit the newsgroup, the Mingler segment, the newsgroup has replaced friends and, to some extent, family as a source more than it has for the other newsgroups. Furthermore, for members of the Insider segment, family has been replaced more than for Devotees and Tourists. Therefore, hypothesis 17 has been partly confirmed in the sense that the empirical material supported that the use of newsgroups have replaced some of the use of non-marketer dominated sources and that the replacement is clearly largest in one of the segments, in which the persons have a strong tie to the newsgroup. Also, the replacement is rather large, compared to the two segments with a weak tie to the newsgroup, in the second segment with strong ties to the newsgroup, again confirming hypothesis 17. However, the replacement

of friends as an information source has been very limited for the Insider segment which is contrary to hypothesis 17. Therefore, the hypothesis is only supported in part.

The analysis in the above shows that information search in newsgroups has, to some extent, substituted sources of information formerly used. The newsgroup mostly acts as a supplement to the information formerly gathered from both personal and impersonal marketer-dominated sources, except for the use of sales personnel, which is almost completely replaced by use of the newsgroup. For the members of the segments with stronger tie to the newsgroup, there is a tendency that replacement is larger than for the other segments. There is, however, one exception to the rule: The use of brochures has almost been replaced by information search in the newsgroup in the Tourist segment.

5.7 Conclusion on Findings on Consumer Search in Newsgroups

The analysis of the data from the sample among Internet users in general showed that 21.13% of the Internet users use newsgroups in their search for information before they make a purchase. The data confirmed the hypothesis that newsgroups are used more than chat-rooms for consumer information search.

The newsgroups selected for the study differ in terms of number of participants and the frequency of contact within the newsgroups. The newsgroups on products with high self-centrality to the consumer, ‘Dogs’ and ‘Car’, were more tightly knit than the two other newsgroups (‘House and Garden’ and ‘Consumer Issues’).

The comparison of the newsgroup sample with the sample among Internet users in general showed fewer differences than expected between the two. Still, there are differences: Members of the newsgroups are more interested in the product class than the average person, they are younger, more of them are males, and the Internet plays a larger part in their lives.

The first hypothesis, tested after this initial description and validation of the sample was whether the motivation to enter a newsgroup can be described by a 4-factor model as proposed by Kozinets (1999). This hypothesis was supported by the data. After this, hypothesis 14 stating that consumers belonging to different user segments are differently motivated to visit the newsgroup was tested. This hypothesis was supported both in terms of level of motivation and relative weight of the motives. Members of the Insider segment are more motivated than members of the other segments on all 4 factors, and the Tourist segment is less motivated to enter the newsgroup than all other segments are. The analysis further showed that, on the relative level, the Tourist segment was significantly higher motivated by the opportunity to get information than the other segments, and significantly lower motivated by the opportunity to socialize than the Insider and the Devotee segments who, to a larger extent, visit newsgroups to get in contact with others sharing the same interest.

The test of the hypothesis concerning newsgroup interaction (hypothesis 15) was supported. The interaction in the newsgroup is thus encouraged by the composition of the newsgroups. In newsgroups, there are members seeking information, and others, who are able and willing to provide this information. Insiders are often also opinion leaders. One might say that participation in the newsgroup gives them a chance to live out their urge to give advice to others and this encourages them to use the newsgroup.

The information from the newsgroup has proven to be quite influential. Quite a large percentage of the respondents indicate that he or she has, and/or in the future will, make their choice of product based on advice given in the newsgroup. The percentage of respondents who indicate that they, on more occasions, have chosen a product according to recommendations from the newsgroup is twice as high for Insiders as it is in the other segments. The percentage of Insiders indicating that they, most certainly, will choose products according to advice given in the newsgroup is about 3 times as high as for the other segments. This evidence confirms that bonding to the newsgroup is an important factor in

relying on information sources in newsgroups, as it is in the physical world.

The analysis of the data from the selected newsgroups confirmed the hypothesis that the use of newsgroups primarily has replaced the use of marketer-dominated sources. There was also support to the hypothesis that the use of newsgroups has replaced the use of personal advice but on lower level than for marketer-dominated sources. For the segments with stronger tie to the newsgroup there is a tendency that the replacement is larger than for the other segments with one exception: The use of brochures has almost been replaced by information search in the newsgroup for members of the Tourist segment.

In total, the analysis of the data from the newsgroups showed that these newsgroups are important places of exchange of consumer information for members of the newsgroups. The information exchange is influenced by the role, the consumer plays in the newsgroup. Some members primarily use the newsgroup for information gathering while others play an active part in the community. These last individuals influence others and are, themselves, more influenced by the information sought or received in the newsgroup than are the other types of users.

The importance of newsgroups in consumer exchange of information is expected to rise in the future, as more Internet users get to be experienced users, and because the members of newsgroups are rewarded when using the newsgroups for information search. This is because members of newsgroups are a mix of people, looking for information, and others, capable of and willing to give this information.

6. Limitations and Issues for Further Research

This almost final chapter of the thesis treats two topics: Limitations of the study and issues for further research. The chapter begins discussing concerns about the chosen angle, from which the main question in the thesis has been seen, and the consequences drawn from my own getting wiser, while working on the thesis. After this, I try to pinpoint the main limitations of the study in terms of questions, not answered, and conclusions drawn on a less firm empirical basis than the ideal for scientific research.

After this moment of looking in the rear-window mirror, it is time to look forward and, therefore, the next part of the chapter is devoted to the discussion of which questions not answered in this study, it seems interesting to study further, both theoretically and empirically.

My work with this thesis started out asking a rather general question about the consumer's use of the Internet for search for consumer information in relation to an upcoming purchase. At this point in time, at the end of the work with the thesis, this is still the main question, I am trying to answer.

However, during the course of the work with the thesis, I came to ask additional questions, had to give up on answering some of them, and answered the rest more or less thoroughly. This process, I will try to explain in more detail in the following.

First, a problem in the choice of which theoretical elements to treat and in which sequence to treat them is that the question of consumer search for purchase-related information on the Internet in fact deals with two decisions: 1) The decision on which product/brand to buy and 2) the decision on which medium to use as a means of getting the necessary information. Widespread consumer decision models (*e.g.* Sheth et al., 1999; Engel et al., 2001) imply that a consumer will only seek for information if he or she experiences difficulties in making the decision based on the amount of information, he or she possess in advance. Assuming this is correct, consumer search, and antecedents for search to

happen, must be the more important of the two types of decisions in the above. This is because the consumer only chooses between media after realizing that he needs information in order to make a choice between alternative solutions to the problem at hand. However, it is possible that the order of importance in the two types of decisions is the opposite: What if the case is that the consumer decides to shop without a specific purchase in mind? The main decision then concerns where to go, and maybe who with. This decision scenario would make the choice criteria quite different from the traditional scenario because the best medium for shopping might not be the best medium for entertainment while getting the information. Although this angle is very interesting, I decided to follow the 'traditional' path starting with the decision to search for product related information, mainly because most of the theoretical developments have followed this path.

Related to this discussion is that the more I worked with consumer search for information, and the consumer's use of the Internet in general and for information search purposes, the more it became obvious to me that the Internet is not one source of information, but rather several, quite different types of sources of information. Therefore, for some time, I wondered if, in fact, it was necessary and possible to cover each source individually, as all sources differ from one another. I also thought of a redefinition of the research question, focusing on one type of source, in order to get a deep understanding of the consumers' use of this information source.

After thorough consideration, I came to the conclusion that the general level is interesting as research object in spite of the above reservations. This conclusion is mainly based on two arguments: First, in order to use the different sources on the Internet, the consumer has to turn on the computer. Once the computer is turned on and on-line, it is easy to switch between sources. It is thus plausible that the consumer actively decides to use the Internet for information search on the general level and, once on-line, uses different kinds of sources, depending on what he experiences while searching. This means that it is, in fact, interesting to understand more about what the determinants are for the consumer to use the Internet for information search in the first place. The second ar-

gument for the interest in use of the Internet for information search on the general level is more pragmatic. In order to learn which types of sources on the Internet are more interesting to investigate further, it is important to know which sources consumers use for information search and the allocation of time between the various sources. As this information does not exist as available secondary data, it is necessary carry out a study on a representative sample of Internet users to establish this knowledge.

In spite of the above described arguments for interest in choice of using the Internet on the general level, and the choice to keep this the main question of the Ph.D. project, I became increasingly convinced that the really new type of information source available to the consumer on the Internet, is virtual communities. This conviction was the result of my theoretical and explorative empirical studies of communication on the Internet. Due to this insight, I felt that it was important to discover whether virtual communities play an, in any way, significant role as a consumer information source. In pursuing this issue, I knew my own limitation: I am not a sociologist and, therefore, would have to do the research into this matter on a rather superficial level, leaving interpretation of communication in the virtual communities to those who are skilled to do so. Consequently, the hypotheses and empirical research into consumer communication in newsgroups relate to rather straightforward issues, and not to issues such as discovery of consumers' deeper motivations for using the groups in the first place. The acknowledgment of my limited skills in sociological studies also encouraged the decision to study only newsgroup communication, because newsgroups are more straightforward to study than other types of virtual communities. Still, the main reason for choosing newsgroups for the empirical study is not this practical limitation, but the fact that newsgroups seem to be the most common virtual community, in which consumers seek information. Nonetheless, it is a limitation of the current study that the study does not cover virtual communities as such, but only one type of community.

Another limitation of the study is that, although the thesis has given some insight into *who* uses the Internet for information search, it has not really brought any explanation to *why* these individuals use the Internet.

The answer to this question is implicit in the Internet Search Model. Looking back, a study using more qualitative methods might have gained better insight into this matter. On the other hand, spending time and resources on qualitative research into different parts of the model would probably not have brought much new knowledge, as the Internet Search Model is built on established theories about the different parts of the model. The limited resources for this study were, therefore, better spent working with advanced statistical methods and sampling, in order to be able to test the model in total on a representative sample of Internet users.

The above discussed limitations in the questions asked and answered in the thesis. There are, however, also some limitations to the inferences, it is possible to draw from the empirical studies carried out in the thesis. The limitations are mainly attributable to discrepancies between the ideal empirical study and the empirical study it was possible to carry out in the real world. Below, I shall point to the major shortcomings of the data from the empirical studies:

The data used to test the hypotheses about the relations between the elements of the Internet Search Model pose several problems. First and foremost, the sample turned out not to be representative for the population, which is 'Internet users in Denmark'. The sample is skewed towards well-educated families with children. The conclusions on the relations in the Internet Search Model, therefore, may not hold for Internet users who belong to the working class. This is important to remember when drawing conclusions from the data. Secondly, the resulting sample was on the small side for some of the analyses. Especially, the sample was too small to do in-dept subgroup analysis on, for instance, the different products. The research design was developed in order to avoid these problems, foremost the non-representative problem. Still, it turned out that, even with all precautions made, it was not possible to get the not well educated appropriately represented. This is probably a problem in surveys in general, and on the Internet especially, and therefore points to a field for further research. Also, the number of households, selected for the screening procedure, was expected to result in a higher number of usable respondents. As described in chapter 4.1, part of the blame for

the small sample size can be put on technical problems. A remedy for the rather small sample would have been to recruit additional respondents. However, I had already spent rather a large amount of time gathering the 233 usable responses and, therefore, I decided to settle for the sample as it was.

In relation to the hypotheses regarding use of the Internet on the general level, the empirical studies were done on rather solid ground. In contrast to this, the empirical research into the newsgroups was based on far more unstable premises. This is because the hypotheses in relation to the use of this source of information were less well underpinned than the hypotheses on the general level, and because the literature on how to do research into these groups is very limited. On top of this, my own skills in doing research in groups are much more limited than my skills in research on the general level. Finally, the sample from newsgroups had to be drawn, using methods which are not good from a research point of view. The consequence of these problems is that there are several limitations to the conclusions based on the data from the newsgroup sample:

First of all, the sample in the newsgroups was mainly drawn by self-selection of the respondents. Therefore, respondents are probably more involved in the research question than is the average newsgroup user. Second, the sample was taken three months later than the download of the observation data. This produced problems in the comparison of the individuals in the two data sets. Third, the sample taken in the newsgroups was rather small meaning that the statistical analyses had to use the summated scales rather than the original scales. Finally, it is, to some extent, problematic to use the whole sample at one time in the analyses, as the respondents were not recruited in the same way in all 4 newsgroups. However, because the universe is quite small in Denmark, there was not really any possibility to draw additional respondents in additional newsgroups. Therefore, I decided to use the data and adjust the selection of statistical tools, instead of trying to improve the size of the sample.

Another problem, encountered in the newsgroup study, was that for several of the latent variables in the hypotheses, there were no pre-designed

and tested scales available for measurement of the latent variables. Therefore, scales were developed for the study. Looking back, it is my opinion that I should have spent more time testing the scales, before using them, as they could have proven not to be usable in the statistical analyses. The reason for not doing pretests of the scales on a larger sample was, again, mainly the limited universe in which the research was done. Fortunately, the scales seemed to work mostly as expected and, therefore, this lack of thorough pretesting did not prove to be a serious problem although the analysis of the data pointed to some possible problems in wording of some of the items. The conclusion to the discussion of the problems in this part of the empirical research is that there, no doubt, is room for further research into consumer information search in newsgroups in order to get an understanding of this topic which is based on firmer ground than are the conclusions in this thesis.

Returning to the general level of the thesis, it is my opinion that the strengths of this thesis are, in many ways, also the source of its weaknesses. The thesis is, in some ways, broad, as it covers a rather broad research question, discusses and uses a 'new' way of gathering data, and uses advanced statistical methods for analysis of the data. This means that I, in my education as a researcher have had the opportunity to work with and learn about several tools which will be of use to me in my further career as a researcher. On the other hand, looking back, I may have spent too much time learning about these subjects, instead of discovering really new insights in consumer behaviour,

Some researchers may also criticise the way of doing research altogether, as there, as discussed earlier, is a chance that more qualitative research would have produced a deeper understanding. There is a chance that more time spent doing qualitative research on the general level could have been a better choice than to look deeper into one type of source on the Internet. However, it was my choice to choose the quantitative path, and thus, we shall never know, if the results of qualitative research would have produced more valuable knowledge than my research into newsgroups did. What I do know is that the research into the newsgroups produced interesting results, and also has shown many issues of further research.

Even though the current study has brought a number of new insights into consumer use of the Internet as an information source on the general level and in newsgroups in particular, there are still many interesting questions to ask in order to gain a better understanding of this research area.

The first additional question is related to the hypothesis which, surprisingly, could not be confirmed: Liking to shop and contact with sales staff is not off-setting use of the Internet for information search purposes. It would be interesting to find explanations for this lack of relation. Possible reasons could lie in the selection of products for the study, as the products involved might not all have qualified as shopping items. It is, therefore, interesting to pursue the relation between shopping, understood as browsing, and using the Internet further, while focusing on other products. In relation to shopping in general, the current research has also identified a need for a deeper understanding of in-home shopping and the explanatory factors for using this kind of shopping. Is the explanation to be found in the type of product, in the consumer's micro-environment in form of time pressure, opening hours etc., or in psychological factors like internal or external locus of control as hypothesized by Wolfinbarger and Gilly (2000)?

Along the same line of reasoning is the question on the decision sequence in the choice of information source, as discussed in the chapter on limitations of this study: Is the Internet chosen as an information source before or after the choice of type of source? This is, of course, important as the answer to this question determines whether presence on the Internet is equally important for all companies.

When it comes to the newsgroup study there are numerous prospects for further research as the study was rather limited in many ways, but large enough to show that the newsgroup is an important place for exchange of consumer information. First of all, it would be interesting to know more about the motivations for entering newsgroups, as the knowledge provided in this study is just scratching the surface. Secondly, a longitudinal study following the selected newsgroups through observation and

additional surveys among the same respondents in order to follow the development of the communication going on the newsgroups would be interesting.

Thirdly, this was a study strictly limited to Denmark. As Denmark is a small country with a limited number of individuals speaking the same language, it would be interesting to spread the study not only across time but also across nations or cultures, in order to investigate whether newsgroups which are centred on consumer subjects are used in the same manner across cultures. As the groups selected for the Danish study are centred on products/issues which are quite common, at least in the western world, it should be possible to find equivalent groups in other countries like for instance France and Germany. Of course, such a study would have to take language barriers into account.

In relation to a subject, a little at the outskirts of the project, but not of my traditional interest areas, I would like to dig further into the pros and cons of electronic surveying, as the studies into and experiences with this kind of survey while creating the research design has shown that there are still many questions, not answered, in this field. Areas which I find interesting and neither theoretically nor empirically well investigated, are the problems faced when surveying the fast growing segments on the Internet: The elderly and the non- or low-educated. In my study, consumers with little education were underrepresented. Probably, they also tend to be underrepresented when using other survey methods. The question is whether the Internet, to them, seems to be an easier or a harder medium to use, when answering to questionnaires. The same question can be asked in relation to the elderly even though this group is quite well represented in my study.

Finally, on the general level, the elderly are of interest as a consumer group using the Internet. The number of respondents in the present study was not large enough in order to be able to perform separate analyses for this group. The elderly Internet users is a growing group for two reasons. Firstly, more and more elderly go on-line. Secondly, as the population grows older, more and more persons who have been on-line when they were younger also become elderly, who are on-line. I find this

segment of special interest because some of the elderly have problems getting around and some are in addition to this afraid to go out. Therefore, the perceived advantages of the Internet in terms of lower search costs should be higher for this group than for the average consumer.

The above has shown that there certainly are limitations to this study, both theoretically and empirically. Nonetheless, the limitations are not so severe that it is not possible to draw a main conclusion from the research carried out in this thesis. This is the topic of the next and final chapter.

7. Conclusion and Implications for Marketing

The goal of this thesis was to add to the understanding of consumers' use of the Internet for information search in relation to an upcoming purchase. The Internet can be seen as an information source on two levels: On the general level, the Internet is seen as one source. Under this general level, there are various types of information sources on the Internet, as there is off-line. The choice of using the Internet is a prerequisite for choosing among the sources on the Internet. Therefore, it is vital to gain knowledge about the determinants of the use of the Internet as information source per se, to know if the Internet has any importance as a consumer information source.

The various sources on the Internet must have a different level of usage and a different level of impact on the consumer. In order to add to the knowledge about the impact of the Internet, it was necessary to move to the level of the various types of sources on the Internet. It was not possible to investigate all types of sources and, therefore, a choice of source had to be made. The type of information source on the Internet which deviate the most from the traditional, off-line sources, is the virtual community. Virtual communities are individualized, non-marketer dominated sources and, therefore, information gathered in a virtual community may be very influential to the consumer. As a consequence of this, virtual communities were selected as the type of source, of which the influence on the consumer was studied. This leads to the research questions for the thesis:

1: What are potential determinants of consumer use of the Internet for in formation search in general and

2: What is the significance of virtual communities in consumer search for information?

In order to investigate these two topics, it was necessary to clarify how and under what circumstances the consumer can be expected to search for information and what type of source, he uses when searching for in-

formation. This investigation was carried out in order to be able to make hypotheses which could be tested in a subsequent empirical study.

The thesis thus presented theory on the consumer's search for information in relation to an upcoming purchase. This theory related to information search using the Internet in general and in virtual communities in particular. The theory presented resulted in statement of several hypotheses related to a model which was named the 'Internet Search Model', and additional hypotheses in regards of use of and influence from newsgroups centred on consumption related topics when searching for information. Newsgroups were selected as an example of a virtual community because newsgroups seem to be the kind of virtual community used the most for exchange of consumer information.

The hypotheses were tested by a survey using electronic questionnaires tailored to the two different research objects, and to the different newsgroups which were selected for the study. The electronic survey was chosen as collection method after consideration of pros and cons for using this method of collecting data. The empirical studies, targeted to test the hypotheses on the relations in the Internet Search Model were thus separate, but related studies. As a consequence of this, the analyses of the data provided were also performed separately. Below, the presentation of the conclusions drawn from the two empirical studies follows the sequence of the hypotheses stated in chapter 2.6. This means that first, conclusions in relation to the 10 hypotheses about the relations in the Internet Search Model are presented and next conclusions to the 7 hypotheses on the communication in and influence from the newsgroups are presented.

The target of the analysis of consumer search behaviour on the Internet in general was to be able to conclude on the search behaviour for the average Internet user. Therefore, much effort was put into the selection of a sample which would be representative of this average Internet user. The resulting sample was representative in terms of age, but well-educated families with children were overrepresented in the sample. The consequence of this is that conclusions drawn in the following are more

valid for Internet users belonging to this segment, than it is for the poorly educated, Internet user living alone.

The main conclusion on the empirical test of the Internet Search Model is that the hypotheses mostly survived the test. The final model was presented in figure 4.5. In detail, the conclusion is that use of the Internet for search for information is determined by interest in and knowledge about the product; importance of the Internet in the person's life in terms of use of the Internet; and perceived cost of using the Internet in terms of time, money and effort. Perceived cost of using the Internet is, in turn, determined by belief in availability of the needed information which, in turn, depends on importance of the Internet and a positive attitude towards shopping from home. There was also support to the hypotheses that the model is the same for women and men.

Surprising in the empirical test of the Internet Search Model was that the expected off-setting effect of a positive attitude towards shopping on the use of the Internet for information search purposes was very far from a significant level. Also perceived cost of using the Internet for information search was a weaker factor in the choice of source than what was expected based on the theories on Internet use, based on traditional economic theory. This supports the researchers, who emphasize the augmented benefits, rather than the reduced costs, as an explanation for consumer use of the Internet. On the other hand, the weak impact of perceived cost may also be due to poor measurement of importance of the Internet in the person's life as altering the measure of this latent variable made the cost factor stronger in the model.

As importance of the Internet in the consumer's life turned out to be an important factor in the choice to use the Internet for search for consumer information, it was also of interest to discover what determines perceived importance of the Internet. The empirical material gave strong support to the two hypotheses on determinants of importance the Internet. The longer a person has been on-line the more he perceives that the Internet is important to him. This is, in part, off-set by the person's age. The older the person is, the less important the Internet is to the person. This result holds regardless if importance of the Internet is measured as

frequency in use of the Internet, stated importance of the Internet, or a combination of both.

When using the Internet for information search purposes, the consumer can select from various types of sources. The empirical data showed that consumers tend to use the type of source they are familiar with because they also use them for other purposes. This was expected, as these sources are the ones which are the least costly to use in terms of psychological effort and probably also time.

During the study of the theoretical foundations for expectations of consumer search in general and on the Internet in particular, it became increasingly clear that the really new source of information available at the Internet, is the virtual community. As newsgroups seem to be the type of community best suited for exchange of consumer information, newsgroups were selected as a special type of source which was interesting to look further into.

First, a hypothesis stating that consumers use newsgroups more than chat-rooms when searching for product information was supported. This confirms that the newsgroup is an interesting type of virtual community when it comes to exchange of consumer information. Second, the analysis of the data from the sample among Internet users in general showed that quite a large percentage (21.13) of the consumers use newsgroups in their search for information before they make a purchase.

The comparison of the newsgroup sample and the sample among Internet users in general showed only few differences. The members of the newsgroups are more interested in the product class than the average person, they are younger, and more of them are males. They have also used the Internet longer than the average Internet user, and they use the Internet more than do the respondents in the sample among Internet users in general. Finally, newsgroup users also find it more beneficial to search the Internet for information, than the respondents in the sample among Internet users in general do.

Hypothesis 12 stating that groups centred on products of which the consumption has higher self-centrality to the consumer are tighter knit was confirmed. This supports the notion that newsgroups differ from each other depending on the type of product discussed in the group.

The first hypothesis tested after this initial description of the newsgroups and validation of the sample was whether motivation to enter a newsgroup can be described by a 4-factor model. This hypothesis was supported by the data. After this, hypothesis 14 was tested. This hypothesis states that consumers belonging to different user segments are differently motivated to visit the group. There was support to this hypothesis on both the level of motivation and the relative weight of the motives. The Insider is more motivated than the other segments on all 4 factors, and the Tourist segment is less motivated to enter the group than all other segments. The analysis further showed that on the relative level, the Tourist segment was significantly higher motivated by the information factor than the other segments, and significantly lower motivated by the contact factor than the Insider and the Devotee segments which, to a larger extent, visit newsgroups to get in contact with others who share the same interest.

The test of the hypothesis concerning group interaction (hypothesis 15) was supported. This means that interaction in newsgroups is encouraged by the composition of the newsgroups. In newsgroups, there are members seeking information and others who are able and willing to provide this information. Insiders are often also opinion leaders. One might say that participation in the newsgroup gives them a chance to live out their urge to give advice to others and this encourages them to use the newsgroup.

The information from the newsgroup proved to be quite influential. Quite a large percentage indicates that he or she has and/or in the future will make their choice of product based on advice given in the newsgroup. The percentage of respondents who indicate that they, on more occasions, have chosen product according to recommendations from the group is twice as high for insiders as it is in the other segments. The percentage of Insiders indicating that they, most certainly, will choose

products according to advice given in the group is about 3 times as high as for the other segments. This shows that bonding to the group is an important factor in dependence on information sources in newsgroups, as it is in real life.

Finally, analysis of the data from the selected newsgroups confirmed the hypothesis that use of newsgroups primarily has replaced use of marketer-dominated sources. There was also support to the hypothesis that use of newsgroups has replaced use of personal advice but on lower level than for marketer-dominated sources. For segments with a stronger tie to the newsgroup, there is a tendency that the replacement generally is larger than for the other segments.

In total, the analysis of the data from the newsgroups showed that these groups are important places of exchange of consumer information for the members of the newsgroups. Information exchange in the groups is influenced by the role, the consumer plays in the group. Some members primarily use the newsgroup for information gathering, while others play an active part in the community. These last individuals influence others and are, themselves, more influenced by the information sought or received in the group than the former group is. Importance of newsgroups is expected to rise in the future, as more Internet users get to be experienced users and because members of the newsgroups are rewarded when using the newsgroups for information search because newsgroups are a mix of people, looking for information, and others, capable of and willing to give this information.

In total, the above shows that the Internet is, indeed, used by the average Internet user, when searching for information before making a decision of what to purchase. There was also evidence that quite a large percentage of consumers use virtual communities, and especially newsgroups, when searching for information, and that they ask questions in newsgroup which they would otherwise direct to sales personnel in retailing. Therefore, the newsgroup acts more as an information source than a reference group for most newsgroup members.

Conclusions drawn in this thesis are interesting news in relation to marketing practice in three ways. First, this study has confirmed that presence on the Internet is important for at least two reasons: The first reason is that consumers, apparently, like to visit marketer-dominated websites in general, and also search for information on objective cues on these websites. Therefore, providing this information to consumers on the Internet is important. The second reason is that consumers in general, and when searching for decision relevant information, use search agents to help find relevant sites to visit. This means that it is important for companies to be found by the search agents because these agents act as filters between companies and consumers unless the company is known to the consumer in advance and has a website domain easily guessed by the consumer (like www.sony.dk).

Secondly, the research into newsgroups showed that these are important information sources to some consumers. As newsgroups are credible sources to the consumers, because they are not marketer sponsored, it is not a possibility to market one-self in the groups directly. Out of ethical reasons it is, of course, also not possible to do this by use of a stand-in person answering questions in the group. This leaves two ways of using the newsgroups in marketing practice: In some groups, it seems to be quite common that companies answer questions openly and this is, of course, an opportunity open to every company. There is an obvious danger in doing this, though: The credibility of the group is undermined and with this, the importance of the group as a credible information source. Far more interesting is, therefore, the other possibility: To use the newsgroup as a sort of on-going focus group on important problems, solutions, and information needs concerning the relevant product class. This can be done at a low cost by observation in the relevant groups at suitable time intervals.

Finally, the study has shown that information on the Internet and especially in newsgroups has supplemented or replaced mostly information formerly acquired from retail personnel. This point to two things: Some consumers, in the future, will use retailing mainly as a distribution channel and a place where it is possible to inspect the product before purchase. However, they will not be willing to pay a higher price for the

product in exchange for advice which they believe to be able to collect better themselves on the Internet. Secondly, if a segment of retailing still wants to be able to charge a surplus for advice given in the store, their retail staff has to be increasingly well educated in order to match, or overpower, the information level of the consumer asking for advice. This is because the consumer very often has gained some knowledge from the Internet before visiting the store to ask for advice.

8. Dansk Sammenfatning

Internettet har haft stigende betydning i både det internationale og det danske samfund. Således er der, ifølge Danmarks Statistik (Danmarks Statistik, 2002a), 75 % af den voksne befolkning i Danmark, der har adgang til internettet derhjemme eller på arbejdspladsen. På grund af denne stigende betydning i samfundet har der været en tendens til, at virksomheder mente, at det var nødvendigt at være til stede på internettet. Hvad denne tilstedeværelse skal bestå i, har der derimod været mere tvivl om, blandt andet fordi virksomhederne mangler viden om baggrunden for forbrugernes anvendelse af internettet.

Der er derfor behov for forskning, der beskæftiger sig med forbrugernes anvendelse af internettet til informationssøgning. I den forbindelse er det relevant at undersøge, hvorvidt den eksisterende teori er anvendelig til at pege på forhold, der styrker eller hæmmer forbrugerens anvendelse af Internettet som en kilde til information i forbindelse med en given købsbeslutning, eller om en tilpasning af den eksisterende teori er påkrævet. Særligt er det interessant at undersøge, hvorvidt virtuelle fællesskaber er i stand til at afløse eller supplere primære referencegrupper, som traditionelt er vigtige kilder til forbrugerinformation.

Ud fra ovenstående stilles følgende spørgsmål som udgangspunkt for denne Ph.D. afhandling:

1: Hvad er mulige determinanter for forbrugerens anvendelse af internettet i forbindelse med informationssøgning og

2: I hvilket omfang har søgning efter information i virtuelle fællesskaber på internettet betydning for forbrugeren?

Afhandlingen beskæftiger sig udelukkende med købsadfærd, der relaterer sig til produkter, hvor der foregår overvejelser og informationssøgning, inden købet foretages. Afhandlingen beskæftiger sig ligeledes ude-

lukkende med søge-elementet i købsprocessen, og således ikke med, for eksempel, selve købet.

Afhandlingen er udarbejdet ud fra et neopositivistisk videnskabssyn. Som konsekvens heraf følger den gennemførte forskning i alt væsentligt den hypotetisk-deduktive metode. Afhandlingens opbygning er således ret traditionel: I indledningen præsenteres forskellige teoretiske indgange til forskningen i forbrugeradfærd, og der vælges fagligt ståsted inden for forbrugeradfærdsforskningen: Den kognitive skole. Med udgangspunkt i dette valg studeres herefter i kapitel 2 den forskning, der hidtil har været indenfor de problemfelter, som afhandlingen beskæftiger sig med. Først behandles teori om forbrugerens søgning efter information inden et køb. Herefter behandles i afsnit 2.3 forbrugerens anvendelse af internettet, i afsnit 2.4 internettet set som flere typer informationskilder og endelig overvejes i afsnit 2.5 årsager til valg af informationskilde. Teorigennemgangen munder i afsnit 2.6 ud i opstilling af en model der beskriver, hvad der er bestemmende for, om forbrugeren anvender internettet i forbindelse med sin søgning efter information inden en købsbeslutning samt en række hypoteser om anvendelsen af en bestemt type virtuelt fællesskab – nyhedsgruppen – i forbindelse med informationssøgning.

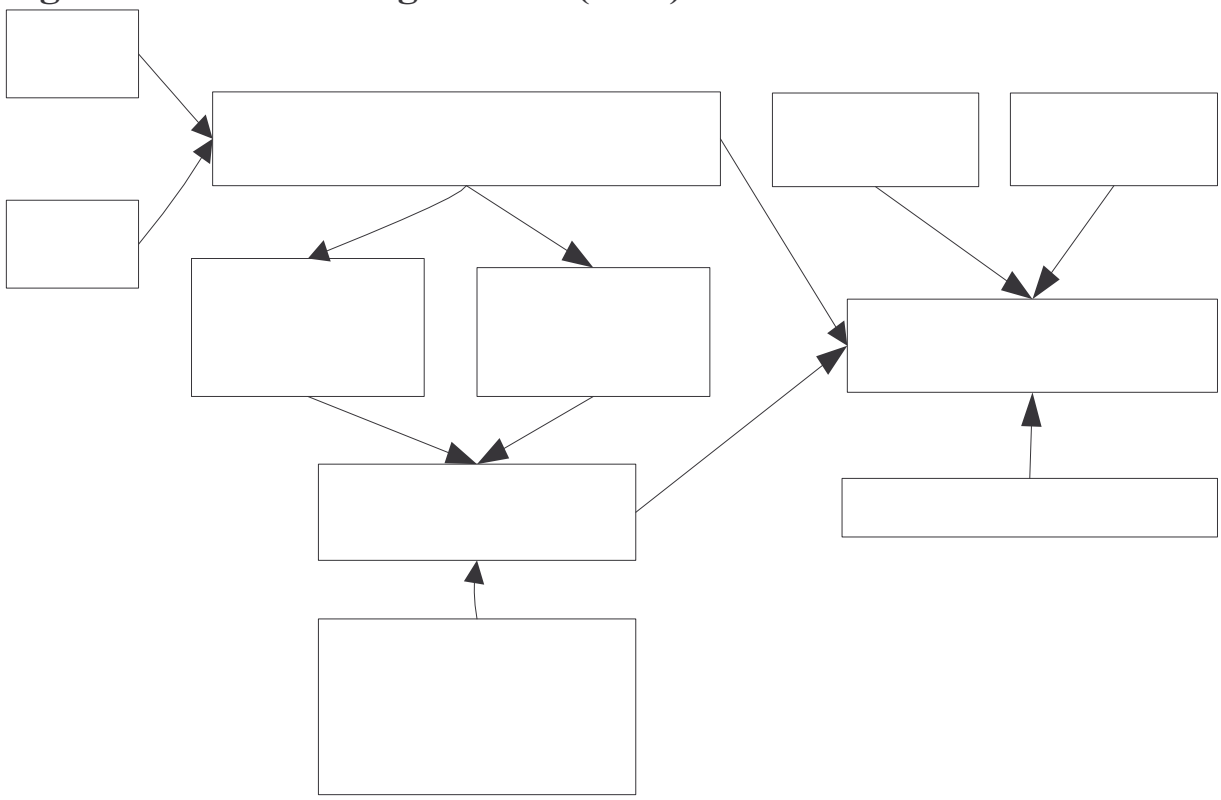
Næste trin her efter er den empiriske test af hypoteserne. Denne planlægges i kapitel 3 og gennemføres efterfølgende i kapitel 4 for så vidt angår informationssøgning på internettet i bred forstand og i kapitel 5 for så vidt angår anvendelse af nyhedsgrupperne. I kapitel 6 redegøres efterfølgende for problemer i de gennemførte teoretiske og empiriske analyser, og der peges på emner for yderligere forskning. Endelig findes i kapitel 7 konklusion og praktiske implikationer af de fundne resultater.

I det følgende gøres kort rede for de opstillede hypoteser, de gennemførte empiriske undersøgelser, og de resultater, der er fremkommet gennem disse undersøgelser. For en mere uddybende gennemgang henvises til selve afhandlingen.

Hypoteser vedrørende anvendelse af internettet som sådan som informationskilde

Med henblik på at besvare det første undersøgelsesspørgsmål opstilles i afsnit 2.6 en model (gengivet i figur 2.9 og i dansk oversættelse i figur 8.1 nedenfor), der forklarer forbrugerens anvendelse af internettet til informationsøgning. Til modellen er knyttet en række antagelser om sammenhænge i modellen. I det følgende redegøres kort for disse antagelser.

Figur 8.1: Internet Søge Model (ISM)



Anvendelse af internettet til informationssøgning kræver en ændring i forbrugerens søgeadfærd i forhold til den adfærd, forbrugeren, gennem sin opvækst, er socialiseret ind i. For at forbrugeren vil foretage en sådan ændring, skal der være fordele ved den nye adfærd i forhold til den tidligere adfærd. Det er klart, at for nogle forbrugere vil de relative fordele ved at anvende internettet til informationssøgning være større, end de vil være for andre forbrugere. Ligeledes vil omkostningerne ved selve adfærdsændringen være mindre for nogle forbrugere, end de vil være for andre.

Alder

Adfærdsændringen forventes lettere kunne ske og efterfølgende blive vedvarende, såfremt forbrugeren i forvejen er vant til og kan lide at bruge internettet. Effekten forventes at ske ad to veje: Direkte og gennem mindskede søgeomkostninger. Den direkte effekt skyldes, at det for forbrugere, der i forvejen anvender internettet som en naturlig del af deres hverdag, også vil være naturligt at anvende internettet til informationssøgning i forbindelse med et køb. Den indirekte effekt fremkommer, fordi anvendelse af Internettet i al almindelighed forventes at give en øget tillid til, at det er muligt at finde pålidelig information på internettet. En sådan tillid, til at den ønskede information er tilgængelig og pålidelig, forventes at påvirke de forventede søgeomkostninger og forventet benefit fra søgningen. I tråd med teori om rational adfærd må det yderligere forventes, at sådanne lavere søgeomkostninger og bedre resultat vil resultere i en øget anvendelse af Internettet i forbindelse med søgning efter information i forbindelse med en konkret problemløsningsproces. Opfattelsen af internettet som en vigtig del af forbrugerens liv forventes i øvrigt at falde med stigende alder og stige med den tid, forbrugeren har anvendt internettet.

En vedvarende adfærdsændring antages også at blive fremmet af en positiv attitude overfor at kunne foretage informationssøgning og indkøb hjemmefra, idet forbrugeren i så tilfælde formentlig i mindre grad finder det problematisk ikke at kunne se varerne fysisk. Ydermere antages det at lette en adfærdsændring, hvis forbrugeren kan lide at prøve nye måder at shoppe på. Endelig antages en travl hverdag at kunne øge de oplevede besparelser i omkostninger ved at anvende internettet, idet anvendelse af internettet til informationssøgning sparer tid, fordi forbrugeren ikke fysisk skal transporteres mellem forskellige informationskilder. Ydermere kan informationssøgningen foretages, når det passer forbrugeren og ikke indenfor en afgrænset tidsperiode.

Ydermere antages en generel interesse i og viden om produktgruppen at fremme anvendelsen af internettet til informationssøgning, idet forbrugere med høj viden om og interesse i produktet har forbedrede muligheder for at anvende en effektiv søgestrategi og dermed også for at finde den ønskede information på internettet uafhængigt af erfaringer med at anvende internettet som sådan.

Sidst, men ikke mindst, antages en forkærlighed for shopping i bred forstand at være hæmmende for anvendelse af internettet til informationssøgning. Dette skyldes, at shopping ofte indebærer social kontakt med medshoppere eller med personale i detailhandelen. Fraværet af denne kontakt antages derfor at nedsætte lysten til at anvende internettet til informationssøgning for de forbrugere, der ser kontakten som noget positivt.

Modellen antages at være den samme for kvinder og mænd.

Som supplement til modellen for anvendelse af internettet set som en samlet kilde opstilles en hypotese vedrørende anvendelsen af de forskellige typer kilder på internettet. I hypotesen antages det, at forbrugeren, i forbindelse med en målrettet informationssøgning, anvender de typer kilder på internettet, han anvender løbende med andre formål for øje.

Hypoteser vedrørende brug af nyhedsgrupper til informationsindsamling

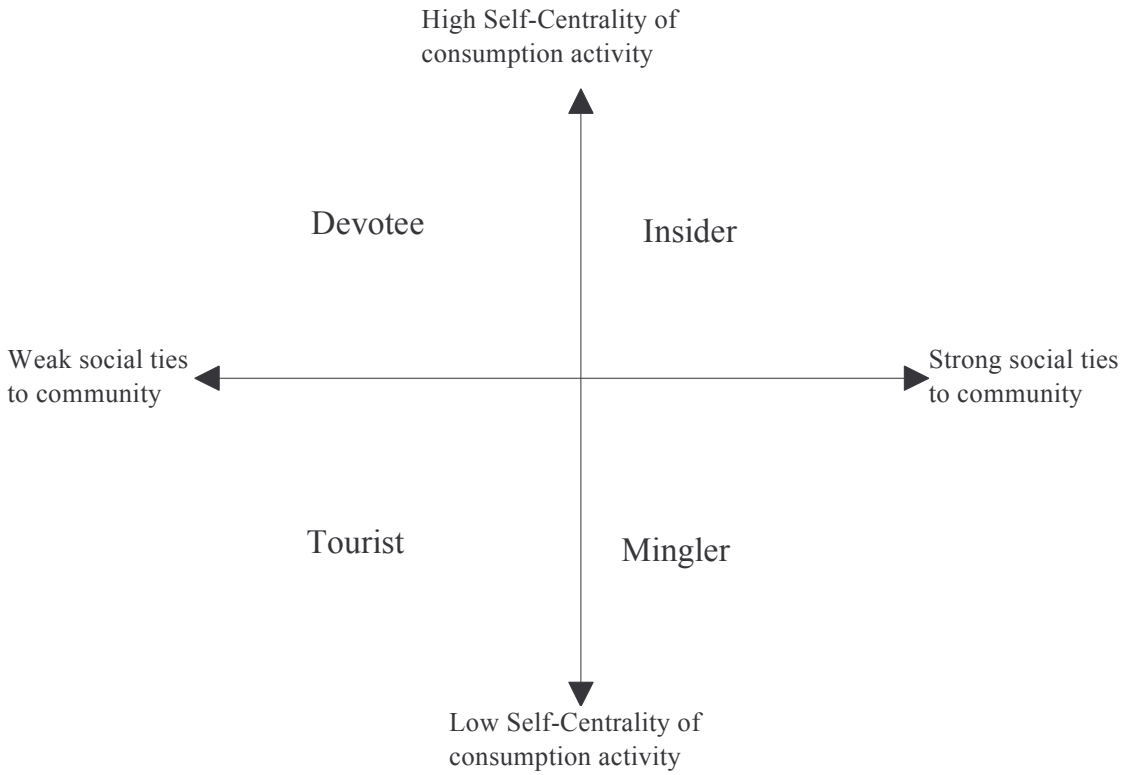
Det andet undersøgelsesspørgsmål i afhandlingen vedrører, hvorvidt virtuelle fællesskaber på internettet anvendes til informationssøgning i købsprocessen, og om informationssøgning i disse fællesskaber har indflydelse på forbrugerne. Da der er et langt svagere teoretisk grundlag for at besvare dette spørgsmål, end der er for at opstille den generelle model, er det nødvendigt at undersøge sammensætningen af og kommunikationen i nyhedsgrupperne, før det forsøges at besvare det egentlige undersøgelsesspørgsmål. I det følgende gøres der rede for denne proces og de dertil hørende hypoteser.

På basis af den indledende teorigennemgang og observationer på internettet vælges det i afsnit 2.6 at opstille hypoteser udelukkende vedrørende anvendelse af nyhedsgrupper og ikke vedrørende andre typer virtuelt fællesskab. Dette skyldes primært, at diskussionsemnerne i nyhedsgrupper, i modsætning til andre typer virtuelle fællesskaber, ofte er centreret om produktvalg. Afledt af dette opstilles hypotesen, at nyhedsgrupper anvendes af en del af forbrugerne i forbindelse med infor-

mationssøgning før et planlagt køb, og at nyhedsgrupper generelt anvendes mere end chatrooms til denne form for kommunikation.

Efterfølgende undersøges sammensætningen af nyhedsgrupperne. Med udgangspunkt i teori om valg af valg af informationskilde (afsnit 2.4.1) og her særligt Kozinet's (1999) begrebsramme (gengivet i figur 2.7 og i figur 8.2 her nedenfor), synes det rimeligt at antage, at der er forskel på, hvordan nyhedsgrupper anvendes til informationssøgning, afhængigt af den type produkt, der er i fokus, og personen, der søger efter information i nyhedsgruppen. Det antages derfor, at der i grupper, der omhandler produkter, der er mere centrale i forhold til personens selvopfattelse, er tættere bånd, end det er tilfældet i grupper, der omhandler produkter, der ikke er knyttet til forbrugerens selvopfattelse. Det antages også, at de forbrugere, der benytter nyhedsgrupperne, har forskellige motiver for at gøre dette. De forskellige motiver er: At søge information, at give information til andre, kontakt til andre, og endelig kan forbrugeren været motiveret af mulighed for underholdning. Yderligere antages det, at der er en sammenhæng mellem tilknytning til gruppen og såvel den absolutte motivation som den relative styrke af de fire motivationsfaktorer. Insiders antages at være mest motiverede for at anvende nyhedsgruppen, mens Tourists forventes at være mindst motiverede for at kontakte gruppen, og antages primært at være motiveret af muligheden for at indhente information i gruppen.

Figur 8.2: Medlemssegmenter i et virtuelt fællesskab



Forbrugerne må forventes at søge information i nyhedsgruppen fordi de forventer at få kvalificerede svar på de spørgsmål, de stiller. Forudsætningen for, at dette sker, er, at der i gruppen også er personer, der kan og vil afgive denne information. Gruppedynamikken i nyhedsgruppen afhænger således – i det mindste for så vidt angår udveksling af forbrugerinformation – af interaktion mellem opinionssøgere og opinionsledere. Hypotese 15 antager, at denne dynamik er til stede i gruppen og at dynamikken forstærker motivationen for at anvende nyhedsgrupper til informationssøgning i købsprocessen.

Såfremt disse indledende hypoteser kan bekræftes, er det muligt at teste de to sidste hypoteser. Disse er direkte relateret til afhandlingens andet hovedspørgsmål. Den første af hypoteserne (hypotese 16) antager, at informationssøgning i nyhedsgrupper primært har erstattet anvendelsen af kilder, der er domineret af markedsfører interesser, og at erstatningen primært er sket for Tourists. Den sidste af de to hypoteser (hypotese 17) antager, at der også er sket nogen erstatning af primære referencegrupper som informationskilde, og at en sådan erstatning mest vil være sket

for personer med nær tilknytning til gruppen – de såkaldte Insiders – og i de grupper, hvor fokus er på produkter, der er knyttet til personens selvopfattelse. Begrundelsen for at antage en større substitutionsgrad i forhold til markedsføringsdominerede kilder end i forhold til primære referencegrupper er, at de virtuelle fællesskaber, især i løsere knyttede grupper, er mindre nære end traditionelle referencegrupper, fordi medlemmerne ikke kender hinanden personligt.

Metode

Næste skridt i forskningsprocessen er at teste de opstillede hypoteser empirisk. Hypoteserne vedrører 2 forskellige niveauer i forbrugerens informationssøgning, nemlig anvendelse af internettet som sådan til informationssøgning og anvendelsen af en bestemt type kilde på internettet. Da undersøgelsesspørgsmålene vedrører to forskellige emner, og kun en forholdsvis lille del af alle internetbrugere forventedes at anvende nyhedsgrupper til informationssøgning, blev der gennemført to separate, men relaterede, undersøgelser. Planlægningen af disse undersøgelser diskuteres i detaljer i kapitel 3 i afhandlingen.

Indledningsvist overvejes i kapitel 3, hvilken type dataanalyse, der forventes gennemført og de krav, dette stiller til det indsamlede datamateriale. Næste trin herefter er udviklingen af måleinstrumenter til test af de opstillede hypoteser. Efter udvikling og konkretisering af målemodellerne i afsnit 3.3 og 3.4 er næste trin valg af kontaktform. Det er naturligt at anvende spørgeskema i denne type analyse. I denne forbindelse skal der foretages et valg af kommunikationsmedium. Dette kan være en form for personlig kontakt, et postudsendt skema eller et elektronisk spørgeskema, publiceret via e-mail eller internet. Undersøgelser og diskussioner i afsnit 3.5 førte frem til, at den kortere forventede svartid, økonomi, målgruppe (internetbrugere) samt højere forventet svarprocent opvejede den ekstra arbejdsindsats, der måtte investeres i at udvikle et spørgeskema til og gennemføre undersøgelsen via internettet. Det endelige spørgeskema kan ses i appendix A.1 og A.2.

Efter beslutningen om at gennemføre undersøgelsen via internettet er næste overvejelse i afsnit 3.6, hvorledes respondenter til de to undersøgelser skal udvælges. I forbindelse med den generelle undersøgelse an-

sås det for vigtigt, at stikprøven så vidt muligt blev repræsentativ for den almindelige internetbruger. Derfor blev der gennemført forsøg med forskellige former for udvælgelse via internettet. Skuffende svarprocenter resulterede i, at der blev gennemført en tilnærmet simpel tilfældig udvælgelse af respondenter ved hjælp at telefonnumre. Disse respondenter blev efterfølgende kontaktet telefonisk med henblik på screening for internetadgang og accept af medvirken. De respondenter, der ville deltage, opgav deres e-mail adresse. Til denne blev sendt et link til et spørgeskema, publiceret på internettet.

For så vidt angår respondenterne i nyhedsgrupperne blev det besluttet at udvælge 4 danske nyhedsgrupper, dækkende 4 forskellige typer produkter. Indledningsvist forsøgtes at sende e-mail direkte til brugerne af en af grupperne ved hjælp af adresser, høstet i nyhedsgruppen. Dette gav dog megen negativ respons fra brugerne, og derfor blev respondenter efterfølgende rekrutteret ved hjælp af selvudvælgelse efter opfordring i en posting i gruppen. Spørgeundersøgelsen blev suppleret med struktureret observation af kommunikationen i grupperne med henblik på analyse af sammensætningen af grupperne samt validering af data fra spørgeundersøgelsen.

Resultater fra undersøgelsen af internettet som sådan til informationssøgning i købsprocessen

I kapitel 4 i afhandlingen behandles de data, der er indsamlet med henblik på test af den generelle model. Dataindsamlingen med henblik på test af den generelle model resulterede i 233 brugbare svar efter kontrol for manglende svar på enkeltspørgsmål. Disse 233 brugbare svar var resultatet af en indledende kontakt til 1673 telefonnumre, hvoraf respondenten i 51,9 % af tilfældene angav at have adgang til internettet og også at være egentlig bruger af internettet. Responsraten i forhold til udsendte mails var på 74,73 % og i forhold til de respondenter, der angav at have internetadgang, var den på 43,6 %.

Stikprøven viste sig ikke at være repræsentativ for den gennemsnitlige internetbruger, idet veluddannede, der lever i familier med børn, er overrepræsenteret i undersøgelsen. Stikprøven er derimod repræsentativ

for så vidt angår alder og geografisk fordeling, mens der er en svag overrepræsentation af ikke-erfarne brugere i stikprøven.

Et af formålene med den generelle undersøgelse var at afdække, hvilke kilder, forbrugere anvender generelt, og om det er de samme kilder, de anvender i forbindelse med søgning efter information i forbindelse med en specifik købssituation. Tabellerne 4.9 og 4.11 indeholder information om dette. På det generelle plan viser tabellerne, at forbrugerne bruger søgemaskiner og websites, sponseret af producenter, løbende, og at det også er disse kilder, der er anvendt i forbindelse med den købsproces, som forbrugerne har redegjort for i undersøgelsen. En yderligere undersøgelse viser en tendens til en sammenhæng mellem intensitet i brugen generelt og brugen med målrettet informationssøgning for øje.

Den efterfølgende test af modellen og de hypoteser, der knytter sig til samme (afsnit 4.3) kunne stort set konfirmere de hypoteser, der opstilledes i afsnit 2.6, og som er refereret tidligere i denne sammenfatning. Der er dog en væsentlig undtagelse: Det var ikke muligt at påvise nogen sammenhæng mellem det at betragte shopping som underholdning og forbrugerens anvendelse af internettet til informationssøgning. Den endelige model kan ses i figur 4.5.

Resultater vedrørende anvendelse af nyhedsgrupper som kilde ved informationsindsamling før et køb

Kapitel 5 beskæftiger sig med test af de opstillede hypoteser vedrørende anvendelse af nyhedsgrupper som informationskilde.

På basis af de data, der er indsamlet med henblik på at teste den generelle model, er det muligt at se, at nyhedsgrupper anvendes langt oftere end chatrooms. 21 % har anvendt nyhedsgrupper i forbindelse med købet, og kun 15 % har anvendt chat. Langt de fleste har dog kun anvendt disse muligheder i meget begrænset omfang, idet kun henholdsvis 4,73 % og 2,15 % har angivet at anvende nyhedsgrupper henholdsvis chat mere end lidt. Disse tal bekræfter dog, at nyhedsgrupper anvendes mere end chat i forbindelse med søgning efter forbrugerinformation.

Valideringen af data fra spørgeundersøgelsen i nyhedsgrupper ved sammenligning med observationer i de samme nyhedsgrupper viser afvigelser fra det observerede i form af et højere aktivitetsniveau i gruppen for de medlemmer, der har svaret, end for den gennemsnitlige, observerbare bruger af gruppen. Dette er formentlig et resultat af dataindsamlingsmetoden, og det vælges at gå videre med analysen på trods af denne afvigelse.

Det kan konstateres, at brugerne af nyhedsgrupperne gennemgående er erfarne internetbrugere, og at den største del af respondenterne har abonneret på gruppen i mere end 1 år, ligesom en stor del af brugerne (81 %) logger på nyhedsgruppen mindst flere gange om ugen. Ser man på sammensætningen indenfor de udvalgte nyhedsgrupper viser det sig, både ud fra de observerede data og spørgeundersøgelsen, at de nyhedsgrupper, der centrerer sig om produkter, der er tættere relateret til forbrugerens selvopfattelse, som hunde og biler som hobby, har langt tættere bånd indenfor gruppen end de to øvrige nyhedsgrupper, der er mere centreret om praktisk hjælp og udveksling af objektiv information. Dette afspejler sig også i medlemssegmenterne, idet der i grupper, centreret om produkter, der er tættere relateret til forbrugerens selvopfattelse, er flere Insiders end i de to andre grupper. Disse har til gengæld, som forventet, en overrepræsentation af Tourists og Minglers.

Hypotesen om motivationsniveau og –faktorer kunne også bekræftes. Insiders er højest motiveret på alle faktorer og Tourists er lavest motiveret på alle faktorer. Såfremt fokus er på den relative betydning af de forskellige motivationsfaktorer, ses det derimod, at Tourists, som forventet, er motiveret af at kunne hente information og mindre motiveret af muligheden for kontakt end de respondenter, der kan henføres til de øvrige segmenter.

Det næste spørgsmål vedrørende forbrugerens anvendelse af nyhedsgrupper var, om kommunikationen i nyhedsgrupperne er understøttet af, at nogle i almindelighed er søgende efter information og hjælp fra andre til at træffe det rigtige valg (opinion seekers), mens andre i almindelighed har lyst til at vejlede andre i deres valg (opinion leaders). Den anvendte model til måling af de to begreber antager, at den samme person

godt kan besidde begge egenskaber. I afsnit 5.5 testes både målemodellen og en model over sammenhængen mellem opinionslederskab og –søgerskab, motivationsfaktor og det at stille og besvare spørgsmål i gruppen. Den endelige model kan ses i figur 5.6, og det empiriske test bekræfter hypoteserne om kommunikation mellem de forskellige typer medlemmer i gruppen.

Efter, at disse indledende hypoteser således er blevet bekræftet, står tilbage at besvare afhandlingens andet hovedspørgsmål: Har nyhedsgrupper betydning i forbrugerens søgen efter information. For at besvare dette spørgsmål mere konkret blev der, som tidligere redegjort for i denne sammenfatning, opstillet to hypoteser: En, der vedrører erstatning af kilder, der er domineret af markedsførerinteresser og en anden, der vedrører substitution af primære referencegrupper. Resultaterne fra undersøgelsen viste, at blandt de markedsfører-dominerede kilder, har nyhedsgruppen i særlig grad erstattet brugen af personale i detailhandelen. Dette var før var en meget anvendt informationskilde. Også brochurer og reklamer er blevet erstattet, og dette er særligt fremtrædende blandt de medlemmer, der ikke føler stærke bånd til hverken til produkttypen eller til nyhedsgruppen, nemlig Tourists (se tabellerne 5.21-23). Til gengæld giver det empiriske materiale ikke støtte til hypotesen om, at information, hentet i nyhedsgruppen, i væsentlig grad kan erstatte primære referencegrupper som familie og venner. Men information fra nyhedsgruppen påvirker medlemmerne og supplerer de primære referencegrupper. Dette gælder i særlig grad for de medlemmer, der føler stærke bånd til gruppen: Insiders og Minglers.

Afslutning

Sammenfattende kan det konkluderes, at de resultater, der er fremkommet i forbindelse med denne Ph.D. afhandling har vist, at Internettet er en vigtig informationskilde for forbrugerne, når de søger efter information i forbindelse med et køb, de står overfor at skulle foretage. Den opfattede vigtighed af internettet stiger med den tid, forbrugeren har haft adgang til internettet og betydningen på det generelle plan vil derfor være stigende over tid, eftersom flere og flere har haft adgang længe. Undersøgelsen har vist, at internettet anvendes som informationskilde fordi internettet opfattes som en billig måde at få information

på målt i tid, penge og anstrengelse. Dette bekræfter, at forbrugerne hovedsageligt vælger informationskilde ud fra en cost/benefit betragtning og dermed en kognitiv overvejelse, hvilket er i tråd med teoridannelsen i den skole indenfor forbrugeradfærdsforskning, som de opstillede hypoteser hviler på. Det var forventet, at fornøjelse ved selve det at shoppe ville være en modificerende faktor i forhold til anvendelse af internettet til informationssøgning, men den empiriske undersøgelse kunne ikke bekræfte dette.

For så vidt angår anvendelse af nyhedsgrupper i forbindelse med informationsindsamling viste undersøgelsen, at disse anvendes af de erfarne internetbrugere, og at den måde nyhedsgruppen anvendes på, afhænger af det bånd, en given person føler til gruppen. Således har internettet i højere grad suppleret primære referencegrupper for forbrugere, der bruger nyhedsgruppen ofte. Den altdominerende substitution er dog sket fra rådgivning i detailhandelen til nyhedsgruppen, og denne substitution er stor for alle brugere af grupperne. De empiriske data understøttede, at kommunikationen i nyhedsgrupperne lever gennem, at nogle forbrugere har behov for svar, og andre forbrugere kan lide at give disse svar, hvorfor den ønskede rådgivning faktisk kan indhentes i grupperne.

De fremkomne resultater betyder i forhold til virksomhedernes markedsføringsfunktion, at det er nødvendigt løbende at overveje, hvorledes det kan gøres attraktivt og frem for alt let for forbrugeren at finde og gøre brug af information om virksomheden og dens produkter. Dette skyldes, at den opfattede omkostning ved brug af en informationskilde spiller en væsentlig rolle for valg af kilde. Samtidigt er det relevant at følge den kommunikation, der foregår i nyhedsgrupper, hvor diskussionerne er relateret til virksomhedens produkter, idet denne kommunikation kan give inspiration til produktudvikling ved at pege på problemer, forbrugerne oplever, og spørgsmål, de har brug for at få besvaret.

Alt i alt kan det konkluderes, at resultaterne i denne Ph.D. afhandling har vist, at internettet er en vigtig kilde til forbrugerinformation. Resultaterne betyder dog på ingen måde, at eksisterende teori må forkastes, idet anvendelsen af internettet som informationskilde skyldes, at dette

opfattes som en effektiv måde at erhverve information på, hvilket er i tråd med de traditionelle teorier om forbrugeradfærd.

References

Aaker, David A., Rajeev Batra & John G. Myers (1992): *Advertising Management, 4th ed*. Prentice-Hall.

Alba, Joseph, John Lynch, Barton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer & Stacy Wood (1997): Interactive Home Shopping: Consumer, Retailer and Manufacturer Incentives to Participate in Electronic Marketplaces, *Journal of Marketing,* vol. 61, pp. 38-53.

Andersen & Wolder (2000): *Speciale om dagligvarer på nettet*. Århus Universitet 2000.

Arbuckle, James L. & Werner Wothke (1999): *Amos 4.0 User's Guide*. SmallWaters Corporation.

Ariely, Dan (2000): Controlling the Information Flow: Effects on Consumer's Decision Making and Preferences, *Journal of Consumer Research,* vol. 27, pp. 233-248.

Armstrong, Arthur & John Hagel III (1996) in a reply to Siegelman, Russel: Electronic Communities, *Harvard Business Review* July/august 1996, p. 164.

Assael, Henry (1984): *Consumer Behavior and Marketing Action,* 2nd ed, Kent Publishing.

Astrup, Andreas (2001): *Send e-mail med FormMail,* http://html.dk/artikler/00012/.

Bagozzi, Richard P. & Youjae Yi (1988): On the Evaluation of Structural Equation Models, *Journal of the Academy of Marketing Science,* vol. 16, pp. 74-94.

Bagozzi, Richard P. & Utpal M. Dholakia (2002): Intentional Social Action in Virtual Communities, *Journal of Interactive Marketing,* vol. 16, no. 2, pp. 2-21.

Bakos, J. Yannis (1997): Reducing Buyer Search Costs: Implications for Electronic Marketplaces, *Management Science,* vol. 12, pp. 1676-1692.

Balasubramanian, Siva K. & Wagner A. Kamakura (1989): Measuring consumer attitudes toward the marketplace with tailored interviews, *Journal of Marketing Research,* vol. 26, pp. 311-326.

Bauer, Hans H., Mark Grether & Mark Leach (2001): Building Customer Relations over the Internet, *Industrial Marketing Management*, Special Issue on Internet Based Business-to-Business Marketing, vol 31, no 2, pp. 1-9.

Baumgartner, Hans & Christian Homburg (1996): Applications of Structural equation modeling in marketing and consumer research: A review, *International Journal of Research in Marketing,* vol. 13, pp. 139-161.

Bearden, William O., Richard G. Netemeyer & Jesse E. Teel (1989): Measurement of Consumer Susceptibility to Interpersonal Influence, *Journal of Consumer Research,* vol. 15, pp. 473-481.

Bearden, William O., Richard Netemeyer & Mary F. Mobley (1993): *Handbook of Marketing Scales. Multi-Item Measures for Marketing and Consumer Behavior Research.* Sage Publications in Cooperation with the Association for Consumer Research.

Belk, Russell (1975) Situational Variables & Consumer Behavior, Enis, Cox &Mokwa (eds.): *Marketing Classics*. 8th ed, pp. 168-178. Prentice-Hall.

Bellenger, Danny N. & Moschis, George P. (1982): A Socialization Model of Retail Patronage, *Advances in Consumer Research,* vol. 9, pp. 373-378.

Bettman, James R. (1973): Perceived Risk and Its Components: A Model and Empirical Test, *Journal of Marketing Research,* vol. 10, pp. 184-190.

Bettman, James R. (1979): *An information Processing Theory of Consumer Choice*. Addison-Wesley.

Bettman, James R., Mary Frances Luce & John W. Payne (1998): Constructive Consumer Choice Processes, *Journal of Consumer Research,* vol. 25, pp. 187-217.

Bickart, Barbara & Robert M. Schindler (2002): Internet Forums as Influential Sources of Consumer Information, *Journal of Interactive Marketing,* vol. 16, no. 2, pp. 31-40.

Blanchard, Anita & Tom Horan (1998): Virtual Communities and Social Capital, *Social Science Computer Review,* vol 16, no. 3, pp. 293-307.

Bloch, Peter H. & Marsha L. Richins (1983): Shopping without Purchase: An investigation of Consumer Browsing Behavior, *Advances in Consumer Research,* vol. 10, pp. 389-393.

Bloch, Peter H., Nancy M. Ridgeway & Daniel Sherrell (1986): Consumer Search: An Extended Framework, *Journal of Consumer Research,* vol 13, pp. 119-127.

Bloch, Peter H., Nancy M. Ridgeway & Daniel Sherrell (1989): Extending the Concept of Shopping: An investigation of Browsing Activity, *Journal of the Academy of Marketing Science,* vol 17, pp. 13-21.

Bloch, Peter H., Nancy M. Ridgeway & James E. Nelson (1991): Leisure and the Shopping Mall, *Advances in Consumer Research,* vol. 18, pp. 445-452.

Blunch, Niels J. (2000): *Videregående Dataanlyse med SPSS og AMOS*. Systime.

Bollen, Kenneth A. & J. Scott Long (1993): *Testing Structural Equation Models.* Sage Publications.

Borgatti, Everett & Freeman (2001): *UCINET 5 for Windows. Software for Social Network Analysis. User's Guide.* Analytic Technologies.

Brim, Orville G., Jr., David C. Glass, David E. Lavin & Norman Goodman (1962): *Personality and Decision Processes*. Stanford.

Brown, Stephen (1997): Six Sixty-six and All That (Or, What the Hell Is Marketing Eschatology?), *European Journal of Marketing,* vol. 31, pp. 639-653.

Bruner II, Gordon C. & Paul J. Hensel (1992): *Marketing Scales Handbook. A Compilation of Multi-item Measures.* American Marketing Association.

Brynjolfsson, Erik & Michael D. Smith (1999)*:* Frictionless Commerce? A Comparison of Internet and Conventional Retailers, at: *http://e-commerce.mit.edu/papers/ERF/ERF6.pdf*.

Brynjolfsson, Erik & Michael D. Smith (2000): The Great Equalizer? Consumer Choice Behavior at Internet Shopbots, at: *http://ebusiness.mit.edu/papers/tge/tge.pdf*.

Burke, Raymond R. (1997): Do You See What I See? The Future of Virtual Shopping, *Journal of the Academy of Marketing Science,* vol. 25, pp. 352-360.

Byrne, Barbara M. (2001): *Structural Equation Modeling with AMOS.* Lawrence Erlbaum Associates.

Campbell, Margaret C. & Amna Kirmani (2000): Consumer's Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent, *Journal of Consumer Research,* vol. 27, pp. 69-83.

Chaudhuri, Arjun (2000): A Macro Analysis of the Relationship of product Involvement and Information Search: The Role of Risk, *Journal of Marketing Theory and Practice,* pp. 1-13.

Chhabra, Surjit & Richard W. Olshavsky (1986): Some Evidence for Additional Types of Choice Strategies, *Advances in Consumer Research,* vol. 13, pp. 12-17.

Churchill, Jr., Gilbert A. (1979): A Paradigm for Developing Better Measures of Marketing Constructs, *Journal of Marketing Research,* vol. 16, Feb, pp. 64-73.

Cobanoglu, Cihan, Bill Warde & Patrick J. Morei (2001): A Comparison of Mail, Fax and Web-based Survey Methods, *International Journal of Market Research*, vol. 43, no.4, pp. 441-452.

Coupier, Mick, Michael W. Traugott & Mark J. Lamias (2001): Web Survey Design and Administration, *Public Opinion Quarterly,* vol. 65, pp. 230-253.

Danmarks Statistik (2002a): *Statistiske Efterretninger. Serviceerhverv.* Befolkningens brug af internet, vol. 18, 25, and 41.

Danmarks Statistik (2002b): *Statistisk Årbog.*

Danmarks Statistik (2002c): www.statistikbanken.dk (Statbank Denmark).

Davis, Duane L. & Ronald S. Rubin (1983): Identifying the Energy Conscious Consumer: The Case of the Opinion Leader, *Journal of the Academy of Marketing Science,* vol. 11, pp. 169-190.

Dholakia, Utpal M. (2001): A Motivational Process Model of Product Involvement and Consumer Risk Perception, *European Journal of Marketing,* vol. 35, no. 11/12, pp. 1340-1360.

Dillman, Don A. (2000): *Mail and Internet Surveys. The Tailored Design Method*. John Wiley and Sons.

Dillon, William R., John B. White, Vithala R. Rao & Doug Filak (1997): Good Science, *Marketing Research,* pp. 22-31.

Dommeyer, Curt J. & Eleanor Moriarty (2000): Comparing Two Forms of an E-mail Survey: Embedded vs. Attached, *International Journal of Market Research,* vol. 42, no 1, pp. 39-50.

Dowling, Grahame R. & Richard Staelin (1994): A Model of Perceived Risk and Intended Risk-Handling Activity, in *Journal of Consumer Research,* vol. 21, pp. 119-134.

Eighmey, John & Lola McCord (1998): Adding Value in the Information Age: Uses and Gratifications of Sites on the World Wide Web, *Journal of Business Research,* vol. 41, pp. 187-194.

Engel, James F., David T. Kollat & Roger D. Blackwell (1970): A model of Consumer Motivation and Behavior, Kollat, Blackwell & Engel (eds.): *Research in Consumer Behavior.* Holt, Rinehart and Winston.

Engel, James F., Roger D. Blackwell & Paul W. Miniard (1993): Consumer Behavior. 7th ed. Dryden Press.

Engel, James F., Roger D. Blackwell & Paul W. Miniard (2001): Consumer Behavior. 9th ed. Dryden Press.

Fan, Xitao & Lin Wang (1998): Effect of Potential Confounding Factors on Fit Indices and Parameter Estimates for True and Misspecified SEM Models, *Educational and Psychological Measurement,* vol 58, no. 5, pp. 701-735.

Farley, John U. & Donald R. Lehman (1977): An Overview of Empirical Applications of Buyer Behavior Systems Models, *Advances in Consumer Research,* vol. 4, pp. 337-341.

Farley, John U. & L. Winston Ring (1974): A Simultaneous-Equation Regression Test of the Howard-Sheth Model, pp. 207-225, in Farley, John U., John A. Howard & L. Winston Ring (eds.): *Consumer Behavior. Theory and Application.* Allyn & Bacon.

Firat, A. Fuat, Nikhilesh Dholakia & Alladi Venkatesh (1995): Marketing in a Postmodern World, *European Journal of Marketing* Vol. 29 pp. 40-56.

Fischer, Eileen, Julia Bristor & Brenda Gainer (1996): Creating or Escaping Community? An Exploratory Study of Internet Consumer's Behaviors, *Advances in Consumer Research,* vol 23, pp. 178-182.

Flynn, Leisa Reinecke & Dawn Percy (2001): Four Subtle Sins in Scale Development: Some Suggestions for Stregthening the Current Paradigm, *International Journal of Market Research,* vol. 43, no. 4, pp. 409-423.

Flynn, Leisa Reinecke, Ronald E. Goldsmith & Jacqueline K. Eastman (1996): Opinion Leaders and Opinion seekers: Two New Measurement

Scales, *Journal of the Academy of Marketing Science,* vol 24, no. 2, pp. 137-147.

Formisano, Roger A., Richard Olshavsky & Shelley Tapp (1982): Choice Strategy in a Difficult Task Environment, *Journal of Consumer Research,* vol. 8, pp. 474-479.

Foxall, Gordon R (1980): Marketing Models of Buyer Behavior, *European Research,* 195-.

Foxall, Gordon R (1986): The Role of Radical Behaviorism in the Explanation of Consumer Choice, *Advances in Consumer Research,* vol. 13, pp. 187-191.

Furse, David H., Girish N. Punj & David W. Stewart (1984): A Typology of Individual Search Strategies Among Purchasers of New Automobiles, *Journal of Consumer Research,* vol. 10, pp. 417-431.

Granitz, Neil A. & James C. Ward (1996): Virtual Community: A Sociocognitive Analysis, *Advances in Consumer Research,* vol. 23, pp. 161-166.

GVU Gatech: 10th WWW User Survey. October 1998, at: http://www.gvu.gatech.edu/user_surveys/survey-1998-10/

Hair, Joseph F., Rolph Anderson, Ronald Tatham & William C. Black (1998): *Multivariate Data Analysis 5th ed.* Prentice Hall.

Hirschman, Elizabeth C. (1993): Ideology in Consumer Research, 1980 and 1990: A Marxist and Feminist Critique, *Journal of Consumer Research,* vol. 19, pp. 537-555.

Hodkinson, Chris, Geoffrey Kiel & Janet R. McColl-Kennedy (2000): Consumer Web Search Behavior: Diagrammatic Illustration of Wayfinding on the Web, *International Journal of Human - Computer Studies,* vol. 52, pp. 805-830.

Hoffman, Donna L. & Thomas Novak (1996): Marketing in Hypermedia Computer-mediated Environments: Conceptual Foundations, *Journal of Marketing,* vol. 60, pp. 50-68.

Hoffman, Donna L. & Thomas P. Novak (1997): A New Marketing Paradigm for Electronic Commerce, *The Information Society,* vol. 13, pp. 43-53.

Hoffman, Donna L. & Thomas P. Novak (1999): eLab research, at: http://ecommerce.vanderbilt.edu/papers.html.

Hoffman, Donna L., Thomas P. Novak & Ann Schlosser (2000): Consumer Control in Online Environments, at: http://ecommerce.vanderbilt.edu/papers.html.

Holbrook, Morris B. & Elizabeth C. Hirschman (1982): The Experimental Aspects of Consumption: Consumer Fantasies, Feelings and Fun, Kassarjian, Harold H. & Thomas S. Robertson (eds.): *Perspectives in Consumer Behavior,* pp. 304-316. Prentice-Hall 1991.

Howard, John A. & Jagdish N. Sheth (1969): *The Theory of Buyer Behavior,* New York: Wiley.

Hulland, John, Yiu Ho Chow & Shunyin Lam (1996): Use of Causal Models in Marketing Science: A Review, *International Journal of Research in Marketing,* vol. 13, pp. 181-197.

Ilieva, Janet, Steve Baron & Nigel M. Healy (2002): Online Surveys in Marketing Research: Pros and Cons, *International Journal of Marketing Research*, vol. 44, no. 3, pp. 361-376.

Jackson, Peter (1999): Consumption and Identity: The Cultural Politics of Shopping, *European Planning Studies,* vol. 7, pp. 25-39.

Jacoby, Jacob & Leon B. Kaplan (1972): The Components of Perceived Risk, in M. Venkatesan (ed.), *Proceedings of the Third Annual Conference of the Association for Consumer Research,* College Park, MD: Association for Consumer Research, pp. 382-393.

Jacoby, Jacob (1972): Opinion Leadership and Innovativeness, in M. Venkatesan (ed.), *Proceedings of the Third Annual Conference of the Association for Consumer Research,* College Park, MD: Association for Consumer Research, pp. 632-649.

Jansen, Erin (2002): *Netlingo*. Netlingo Inc. Available at: www.netlingo.com

Jensen, Jan Møller (1990): *Familiens købsbeslutninger Et 'købscenter' perspektiv. En empirisk undersøgelse af mandens og kvindens roller i købsbeslutningsprocessen ved køb af langvarige forbrugsgoder.* Ph.D. Thesis, Institute for Marketing, University of Southern Denmark.

Jepsen, Anna Lund (2001): *Consumer Search Behaviour on the Internet. Development of a Preliminary Model.* Ph.D. course assignment, University of Southern Denmark.

Jepsen, Anna Lund (2002): *Spørgeundersøgelser afviklet ved hjælp af Internettet.* Ph.D. course assignment, University of Southern Denmark.

John, Deborah Roedder (1999): Consumer Socialization of Children: A Retrospective Look at Twenty-Five Years of Research, *Journal of Consumer Research,* vol. 26, pp. 183-213.

Kargaard, Anders & Thomas Johansen (2001): *Værdien af bannereksponeringer. En effekt uden et klik?* Kandidatafhandling, Handelshøjskolen i København.

Klein, Lisa R. (1998): Evaluating the Potential of Interactive Media through a New Lens: Search versus Experience Goods, *Journal of Business Research,* vol. 41, pp. 195-203.

Kollat, David T., Roger D. Blackwell & James F. Engel (1972): The Current Status of Consumer Behavior Research: Developments During the 1968-1972 Period, *Proceedings of the Third Annual Conference of the Association for Consumer Research,* pp. 576-585.

Korgaonkar, Pradeep & Lori D. Wolin (1999): A Multivariate Analysis of Web Usage, *Journal of Advertising Research,* March/April, pp. 53-68 .

Kotler, Philip (2000): *Marketing Management. The Millenium Edition,* Prentice-Hall.

Kozinets, Roberts V. (2002): The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities, *Journal of Marketing Research,* vol 39, pp. 61-72.

Kozinets, Robert V. (1999): E-tribalized Marketing?: The Strategic Implications of Virtual Communities on Consumption, *European Management Journal,* vol. 17, no.3, pp. 252-264.

Kristensen, Kai, Henning Madsen & Peter S. Mortensen (1982): *Analyse af kvalitative data,* Systime.

Laaksonen, Pirjo (1999): Involvement, in Earl, Peter E. & Simon Kemp (eds.):*The Elgar Companion to Research and Economic Psychology.* Edward Elgar Publishing.

Lascu, Dana Nicoleta and George Zinkhan (1999): Consumer Conformity: Review and Applications for Marketing Theory and Practice, *Journal of Marketing Theory and Practice,* vol. 2, pp. 1-12.

Lavik, Randi (1984): Forbrukeratferd og forbrugerholdninger blant kvinner, in *Grønmo, Sigmund (red.): Forbruker, marked og samfunn,* p. 89-99, Universitetsforlaget, Oslo.

Lehman, Donald R.(1999): Consumer Behavior and Y2K, *Journal of Marketing,* vol. 63, pp.14-18.

Lichtenstein, Donald R. & William O. Bearden (1989): Contextual Influences on Perceptions of Merchant-Supplied Reference Prices, *Journal of Consumer Research,* vol. 16, pp. 55-66.

Lumpkin, James R. & James B. Hunt (1989): Mobility as an Influence on Retail Patronage Behavior on the Elderly: Testing Conventional Wisdom, *Journal of the Academy of Marketing Science,* vol. 17, pp. 1-12.

Lumpkin, James R. & John M. Hawes (1984): Understanding the Outshopper, *Journal of the Academy of Marketing Science,* vol. 12, pp. 200-218.

Lunn, J.A. (1974): Consumer Decision-Process Models, in Sheth, Jagdish H. (eds.): *Models of Buyer Behavior.* Harper & Row pp. 34-69.

Lynch, John G. & Dan Ariely (1998): Electronic Shopping for Wine: How Search Costs for Information on Price, Quality and Store Comparison Affect Consumer Price Sensistivity, Satisfaction with Merchandise and Retention, at: *http://ebusiness.mit.edu/papers/*

Maher, Jill K., Lawrence J. Marks & Pamela E. Grimm (1997): Overload, Pressure and Convenience: Testing a Conceptual Model of factors Influencing Women's Attitude toward, and Use of, Shopping Channels, *Advances in Consumer Research,* vol. 24, pp. 490-498.

Maignan Isabelle & Bryan A.Lukas (1997): The Nature and Social Uses of the Internet: A Qualitative Investigation, *Journal of Consumer Affairs,* vol. 31, pp. 346-371.

Malhotra, Naresh K. (1999): *Marketing Research. An Applied Orientation*3rd*,* Prentice-Hall.

McAlexander, James H., John W. Schouten & Harald F. Koenig (2002): Building Brand Community, in *Journal of Markerting,* vol. 66, pp. 38-54.

McWilliam, Gil (2000): Building Stronger Brands through Online Communities, in *Sloan Management review,* Spring, pp. 43-54.

Miller, Daniel (1998a): *A Theory of Shopping*, Polity Press, Cambridge.

Miller, Daniel, Peter Jackson, Nigel Thrift, Beverley Holbrook & Michael Rowlands (1998b): *Shopping, Place and Identity.* Routledge, London.

Mitchell, Vincent-Wayne (1999): Consumer Perceived Risk: Conceptualisations and Models, *European Journal of Marketing,* vol. 33, no. ½, pp. 163-195.

Morrison, David E. and Julie Firmstone (2000): The Social Function of Trust and Implications for E-commerce, *International Journal of Advertising,* vol. 19, pp. 599-623.

Muniz, Albert M. Jr. & Thomas C. O'Guinn (2001): Brand Communty, *Journal of Consumer Research,* vol. 27, pp. 412-432.

Nancarrow, Clive, John Pallister & Ian Brace (2001): A New Research Medium, New Research Populations and Seven Deadly Sins for Internet Researchers, *Qualitative Market Research: An International Journal,* vol. 4, no. 3, pp. 136-149.

NetValue Worldwide, at www.netvalue.com: Press releases (Now taken over by ACNielsen Netratings).

Nicosia, Franco M. (1986): The Post-Affluence Consumer: Consumer Decision Processes Revisited, *Advances in Consumer Research,* vol. 13, p 509.

Nie, Norman & Lutz Erbring (2000): Internet and Society, at: *http://www.stanford.edu/group/siqss*.

Novak, Thomas P. & Donna L. Hoffman (1997): Modeling the Structure of the Flow Experience Among Web Users, at: *http://ecommerce.vanderbilt.edu/papers/pdf/flow.PDF*.

Novak, Thomas P., Donna L. Hoffman & Yui-Fai Yung (2000): Measuring Customer Experience inOnline Environments: A structural Modeling Approach, *Marketing Science,* vol. 19. iss.1, pp. 22-42.

Olshavsky, Richard W. (1985). Towards a More Comprehensive Theory of Choice, *Advances in Consumer Research,* pp.465-470.

Olshavsky, Richard W. & Donald H. Granbois (1979): Consumer Decision Making – Fact or Fiction?, Kassarjian, Harold H. and Thomas S. Robertson (eds) (1991): *Perspectives in Consumer Behavior.* Prentice-Hall, pp 89-100.

Percy, Larry (1998): Pragmatic Reality, *European Advances in Consumer Research,* vol. 3, pp. 254-258.

Peter, J. Paul (1979): Reliability: A review of Psychometric basics and Recent Marketing Practices, *Journal of Marketing Research* vol. 16, Feb, pp. 6-17.

Peter, J. Paul (1981): Construct Validity: A review of Basic issues and Marketing Practices, *Journal of Marketing Research,* vol. 18, May, pp. 133-145.

Peter, J. Paul & Gilbert A. Churchhill (1986): Relationships among Research Design Choices and Psycometric Properties of Rating Scales: A Meta-Analysis, *Journal of Marketing Research,* vol. 23, pp. 1-10.

Peterson, Robert A. (1994): A Meta-analysis of Cronbach's Coefficient Alpha, *Journal of Consumer Research,* vol. 21, September, pp. 381-391.

Peterson, Robert R, Sridhar Balasubramanian & Bart J. Bronnenberg (1997): Exploring the Implications of the Internet for Consumer Marketing, *Journal of the Academy of Marketing Science,* vol. 25, pp. 329-346.

Pew Internet and Life Report (2000): *Tracking Online Life: How Women Use the Internet to Cultivate Relationships with Family and Friends.* The Pew Internet & American Life Project, at: www.pewinternet.org .

Proquest, at: www.proquest.com via the library at University of Southern Denmark.

Rassuli, Kathleen M. & Gilbert D. Harrell (1990): A New Perspective on Choice, *Advances in Consumer Research,* vol. 17, pp.737-744.

Ratchford, Brian T. (1987): New Insights About the FCB Grid, *Journal of Advertising Research,* vol 27, pp. 24-38.

Rau, Pradeep & Saeed Samiee(1981): Models of Consumer Behavior: The State of the Art, *Journal of the Academy of Marketing Science,* vol. 9, pp. 300-316.

Rehm, Marsha L. (2000): An Aestethic Approach to Virtual Community, *Family and Consumer Sciences Journal,* vol 29, pp. 153-172.

Rogers, Everett M. (1968): *Diffusion of Innovations.* The Free Press.

Rogers, Everett M. (1995): *Diffusion of Innovations.* 4 th ed. The Free Press.

Roselius, T. (1971): Consumer rankings of Risk Reduction Methods, in *Journal of Marketing,* vol 35, pp. 56-61.

Schiffman, Leon G. & Leslie Lazar Kanuk (2000): *Consumer Behavior*. 7th edition. Prentice-Hall.

Schiffman, Leon G. & Leslie Lazar Kanuk (1987): *Consumer Behavior*. 3rd edition. Prentice-Hall.

Schillewaert, Niels, Fred Langerak & Tim Duhamel (1998): Non-Probability Sampling for WWW Surveys: A Comparison of Methods, *Journal of the Marketing Research Society*, vol. 40, no. 4, pp. 307-322.

Sénécal, Sylvain & Jacques Nantel (2001): *Online Interpersonal Influence: A Framework.* Working paper published, at: http://ecommerce.mit.edu/papers/ERF/ERF117.pdf.

Sheth, Jagdish H. (1972): The Future of Buyer Behavior Theory, *Proceedings of the Third Annual Conference of the Association for Consumer Research,* pp. 562-575.

Sheth, Jagdish H. (1974): *Models of Buyer Behavior.* Harper & Row.

Sheth, Jagdish N., Banwarth Mittal & Bruce I. Newman (1999): *Customer Behavior. Consumer Behavior and Beyond.* Dryden.

Simonson, Itamar & Stephen M. Nowlis (2000): The Role of Explanations and Need for Uniqueness in Consumer Decision Making: Unconventional Choices Based on Reasons, *Journal of Consumer Research,* vol. 27, pp. 49-68 .

Socialforskningsinstituttet (1997): *Levevilkår i Danmark*. Statistisk Oversigt.

Solomon, Michael, Gary Bamossy & Søren Askegaard (1999): *Consumer Behavior. A European Perspective.* Prentice-Hall.

Srinivasan, Narasimhan & Brian T. Ratchford (1991): An Empirical Test of a Model of External Search for Automobiles, *Journal of Consumer Research,* vol. 18, pp. 233-242.

Stanton, Jeffrey M. (1998): An Empirical Assesment of Data Collection Using the Internet, *Personnel Psychology,* vol.51, pp. 709-725.

Steenkamp, Jan-Benedict E.M. & Hans C.M. van Trijp (1991): The Use of LISREL in Validating Marketing Constructs, *International Journal of Research in Marketing,* vol. 8, pp. 283-300.

Steenkamp, Jan-Benedict E.M. & Hans Baumgartner (2000): On the Use of Structural Equation Models for Marketing Modeling, *International Journal of Research in Marketing,* vol. 17, pp. 195-202.

Stone, Gregory P. (1954): City Shoppers and Urban Identification: Observationson the Social Psychology of City life, *American Journal of Sociology,* vol. 60, pp. 36-45.

Tse, Alan C.B., Ka Chun Tse, Chow Hoi Yin, Choy Boon Ting, Ko Wai Yi, Kwan Pui Yee & Wing Chi Hong (1995): Comparing Two Methods of Sending Out Questionnaires: E-mail versus Mail, *Journal of the Marketing Research Society*, vol. 37, pp. 441-453.

Tse, Alan C.B.: (1998): Comparing the Response Rate, Response Speed and Response Quality of Two Methods of Sending Questionnaires: E-mail vs. Mail, *Journal of the Marketing Research Society*, vol. 40, pp. 353-361.

Tønnesen, Jens (2001): *Statistik over dk.hierarkiet: Antal brugere opdelt efter grupper,* at: http://www.usenet.dk/statistik/brugere.html.

UCLA Centre for Communication Policy (2000): The UCLA Internet Report: 'Surveying the Digital Future', at: http://www.ccp.ucla.edu/pages/internet-report.asp.

Urban, Glen L., Fareena Sultan & William Qualss (1999): Design and Evaluation of a Trust Based Advisor, at: *e-commerce.mit.edu/forum/.*

Venkatesh, Alladi (1998): Cybermarketscapes and Consumer Freedoms and Identities, *European Journal of Marketing,* vol. 7/8, pp. 664-676.

Venkatesh, Alladi (1999): Postmodern Perspectives for Macromarketing: An Inquiry into the Global Information and Sign Economy, *Journal of Macromarketing,* vol. 19, no. 2, pp. 153-169.

Viswanathan, Madhubalan, Terry L. Childers & Elizabeth S. Moore (2000): The Measurement of Intergenerational Communication and

Influence on Consumption: Development, Validation, and Cross-Cultural Comparision on the IGEN Scale, *Journal of the Academy of Marketing Science,* vol. 28, pp. 406-424.

Ward, Michael R. & Michael J. Lee (1999): Internet Shopping, Consumer Search and Product Branding, at: *http://ux6.cso.uiuc.edu/~ward1/brand5.PDF*.

Wellman, Barry & Scot Wortley (1990): Different Strokes for Different Folks: Community Ties and Social Support, *American Journal of Sociology,* vol. 96, pp. 558-88.

Wellman, Barry (2001): Networks in the Global Village. *Life in Contemporary Communities*. Westview Press.

Wellman, Barry & Milena Gulia (1999): Net-surfers don't ride alone, *Networks in the Global Village*. Westview Press, pp. 331-366.

Wilkie, William L. & Peter R. Dickson (1985): Shopping for Appliances: Consumer's Strategies and Patterns of Information Search, In Kassarjian, Harold H. and Thomas S. Robertson (eds.) (1991): *Perspectives in Consumer Behavior.* Prentice-Hall, pp. 1-26.

Wolfinbarger Mary & Mary Gilly (2001): Shopping Online for Freedom, Control and Fun, *California Management Review*, Winter, vol. 43 (2), pp. 34-53.

List of Figures

Figure 1.1: The competing schools: Differences in the view on consumer behaviour as a science 13
Figure 1.2: Overview of the thesis 23
Figure 2.1: A model of external search for automobiles 31
Figure 2.2: A model of external search – revision 1 36
Figure 2.3: A model of external search - revision 2 40
Figure 2.4: Hoffman and Novak's model 50
Figure 2.5: Traditional sources of information for the consumer 54
Figure 2.6: Information sources on the Internet 56
Figure 2.7: Membership Segments in a Virtual Community 65
Figure 2.8: Clusters in information search 79
Figure 2.9: Internet Search Model (ISM) 95
Figure 3.1: The Internet Search Model with measurement variables 126
Figure 4.1: Initial Measurement Model 205
Figure 4.2 Expected Covariances in the Measurement Model 206
Figure 4.3: Final Measurement Model 215
Figure 4.4: Internet Search Model – Initial Model 218
Figure 4.5: Final model of determinants of use of the Internet for Consumer information search. 225
Figure 4.6: Final Internet Search Model 230
Figure 5.1: Model of motivation 255
Figure 5.2: Measurement model, Opinion leadership 268
Figure 5.3: Answering questions as a result of level of opinion leadership 272
Figure 5.4: Asking questions as a result of possibility to get information 273
Figure 5.5: Model of newsgroup interaction as a result of opinion leadership 275
Figure 5.6: Model of newsgroup interaction after adjustments 276
Figur 8.1: Internet Søge Model (ISM) 309
Figur 8.2: Medlemssegmenter i et virtuelt fællesskab 313

List of Tables

Table 3.1: Variables for the general study 127
Table 3.2: Variables for the newsgroup study 139
Table 3.3: Response time using different survey methods 144
Table 3.4: Response using two different sampling frames 155
Table 4.1: Civil status of families in the sample 171
Table 4.2: Product class thought about while answering the questions 173
Table 4.3: Interest in and knowledge about the product 174
Table 4.4: Importance of the decision 175
Table 4.5: Time used for information search 175
Table 4.6: Time, respondents and population have been on-line 177
Table 4.7: Time respondents use the Internet pr. week 178
Table 4.8: Percentage of Internet use that is work-related 179
Table 4.9: Use of Internet opportunities in general 181
Table 4.10: Locus of control 182
Table 4.11: Use of the various types of sources on the Internet 184
Table 4.12: Product bought and relative use of the Internet for search for information 187
Table 4.13: Attitude towards availability of information 188
Table 4.14: Attitude towards credibility of information 189
Table 4.15: Perception of cost of using the Internet for information search 190
Table 4.16: Innovativeness in shopping 191
Table 4.17: Inclination to in-home shopping 192
Table 4.18: Role Overload 193
Table 4.19: Shopping enjoyment 194
Table 4.20: Importance of contact with sales personnel 195

Table 4.21: Fit indices, Measurement Model after Modification 1 Omission of 'Role Overload' 208
Table 4.22: Fit indices after modification 2 of the Measurement Model 209
Table 4.23: Fit indices after modification 3 of the Measurement Model 210
Table 4.24: Fit indices after modifications 3, 4 and 5 of the Measurement Model 211
Table 4.25: Estimated Covariances in the Measurement Model 212
Table 4.26: Fit indices, Measurement Model 216
Table 4.27: Reliability Measures for Scales used in Estimation of Structural Equations 216
Table 4.28: Fit Indices, Initial Structural Model 219
Table 4.29: Regression weights, paths in the model 221
Table 4.30: Fit indices, Modification 1 – omitting perceived credibility of the information from the model 222
Table 4.31: Fit indices, Modification 2 - omitting 'Shopping Enjoyment' and 'Contact to Sales Staff' from the model 222
Table 4.32: Fit indices, Modification 3 – adding a path in-home shopping to perceived availability of information 223
Table 4.33: Fit indices, Modification 4 – omission of e1b1 from the model. 224
Table 4.34: Estimated regression paths for the Internet Search Model 226
Table 4.35: Squared multiple correlations for the Internet Search Model 229
Table 5.1: Activity in and size of the chosen newsgroups 238
Table 5.2: Number of meetings with the same person 239
Table 5.3: Sample from Hus-og-Have ('House and Garden') 242
Table 5.4: Sample from dk.fritid.bil ('Car') 243
Table 5.5: Result for sample from dk.fritid.dyr.hund (Dogs) 244

Table 5.6: Sample from dk.forbruger (‘Consumer Issues’) 245
Table 5.7: Length of membership in the newsgroup 250
Table 5.8: Frequency of Log-on to the newsgroup 251
Table 5.9: Time spent pr. Logon 252
Table 5.10: Number of members in the segments in each newsgroup 253
Table 5.11: Correlations from the confirmatory factor analysis of model of motivation factors. 257
Table 5.12: Relative weight of motivation factors for the 4 segments 260
Table 5.13: Index of activity of named opinion leaders, all newsgroups 265
Table 5.14: Fit indices for model measuring Opinion Leadership and –Seekership 269
Table 5.15: How often questions are posed to and answered in the newsgroup by the respondents 271
Table 5.16: Final reliability measures, model of newsgroup interaction 276
Table 5.17: Fit indices for Model of Newsgroup Interaction 277
Table 5.18: Estimated regression paths for the Model of Newsgroup Interaction 277
Table 5.19: Chosen based on advice from newsgroup 281
Table 5.20: Will choose based on advice from newsgroup 281
Table 5.21: Use and replacement of consumer information sources 283
Table 5.22: Use and replacement of brochures as an information source 284
Table 5.23: Use and replacement of mail distributed advertisements as an information source 284
Table 5.24: Use and replacement of friends seen often as an information source 285
Table 5.25: Use and replacement of family seen often as an information source 286