Drivers and Impediments of Consumer Online Information Search: Self-controlled versus Agent-based Search in a High Involvement Context

Sarah Spiekermann, Martin Strobel, Dirk Temme
INTRODUCTION

One of the most enduring sequences of research effort in consumer behaviour has been the study of information search behaviour prior to purchase (Beatty and Smith, 1987; Moorthy et al., 1997; Punj and Staelin, 1983; Srinivasan and Ratchford, 1991). The goal has been to investigate drivers and impediments of the search activity as well as search intensity.

Since we can observe an increasing use of the WWW as a source of product purchase information, many of these older studies risk to become partially outdated. This is, because the new electronic medium promises to considerably reduce traditional search cost relevant in offline markets (Alba et al., 1997; Bakos, 1997), offers an exciting amount of new product information sources and efficiently supports the search process through agent technology (Häuble and Trifts, 2000). Within a few years, the Internet has evolved as a major source of product information retrieval and it is expected to play an even bigger role in the buying process once 2nd generation agent applications are fully deployed (Vulcan, 1999; West et al., 1999). As a result, information search behavior is in transition.

When the market dynamics of the Internet economy are studied, scholars have a tendency to assume the widespread deployment and use of highly performing search engines or personal agent technologies. Based on the fascinating idea that there is a highly efficient and reliable technology at work, they have started to integrate a ‘minimal search cost perspective’ in their models and investigated, for example, the consequences for pricing (Brown, 2000) and market dynamics (Alba et al., 1997). Little research emphasis has been attributed, however, to the fact that consumers may not exclusively want to base their purchase decision on agent recommendations. For example, consumers may not trust the technology to the necessary end and may therefore wish to complement agent suggestions with a personally conducted and more controlled search activity. Urban et al. (1999), for example, who tested consumers satisfaction with a ‘trust-based advisor’ found some evidence that subjects with different pre-dispositions in the purchase process (e.g. different levels of product knowledge) react differently to the support technology. It is therefore important to ask to what extend and why consumers recur to agent recommendations for their purchase decisions and in their search activity or, in contrast, prefer personally controlled information search. What are the influential factors behind the use of agent technologies versus self-controlled product inspection?
The study presented hereafter proposes and tests a comprehensive model of Internet based information search. It explicitly respects the existence of two different types of search conducted online: interaction with an electronic decision-support system on one side and personally controlled, detailed inspection of product descriptions on the other. For both search activities driving and impeding factors are being tested in a structural equation model. All factors hypothesised to influence search have either been derived from traditional studies in information search or from recent work on online navigation. This allows us to generate new insights in the relative importance of different Internet search activities and their drivers, but also to test the validity of traditional search constructs for the new online environment.

Our empirical analysis is based on data from 151 online shopping sessions that were collected from subjects looking for compact cameras in an experimental online store. Subjects had to spend their own money if they decided to buy something. The reason why we chose compact cameras to build an Internet search model was that cameras represent a reasonably complex product class that invites information search. In contrast to experience goods where consumers can gauge product quality only by using the product, cameras also display strong search-good characteristics, meaning that consumption benefits can be reliably predicted prior to purchase on the basis of factual product information (Nelson, 1970; Weiber and Adler, 1995). Cameras therefore represent a product class for which the Internet offers relatively strong information advantages. Finally, cameras are in a price range (between $100-$450) frequently confronted by consumers in purchase situations.

Following this introduction, the current article is separated into four parts. Section 2 summarizes the structural equation model we propose and derives all hypotheses made in it. Section 3 then describes the methodology we employed to test our model. Section 4 reports on the results obtained commenting on model fit as well as the acceptance or denial of individual hypotheses tested. In this section we also expend on the relative importance of agent supported search versus detailed product inspection. Section 5 then concludes with a summary of major findings and points to some limitations of the current study as well as opportunities for future research.
PROPOSED MODEL OF ONLINE INFORMATION SEARCH

We propose a model of drivers and impediments of online information search, based on the assumption that consumers seek information on the WWW in order to make better and more satisfying purchase decisions. At the centre of our model is therefore the amount of search activity displayed by subjects. This search activity is hypothesised to be dependent on a number of variables including purchase involvement, product experience, product class knowledge, perceived risk, stage in the buying process, privacy concerns, cost and benefits of search as well as the achievement of a flow status. Figure 1 gives an overview of the hypothesised constructs and their relationships.

![Proposed Structural Equation Model for Online Information Search](image-url)
Endogenous Model Constructs

There are two prominent ways in which product information is sought online: One is to obtain information by consulting detailed product descriptions. Here, the use of product fact sheets, comparison matrices and photographs is a common means employed in online environments. Another way to search for products online is the use of interactive systems (e.g. recommendation agents) that allow for more efficient attribute sorting of products and comparative shopping. Based on customer input data product offers are being suggested by the system (e.g. MySimon.com, PersonalLogic.com).

For modelling purposes we distinguish between these two different types of online information search. Interactivity, on one side, is a multidimensional construct which implies the reciprocity of information exchange, the availability of information on demand, response contingency, customisation of content, and real-time feedback (Alba et al., 1997; Häuble and Trifts, 2000). Detailed product screening, on the other side, displays none of these characteristics. “(Interactivity)...is the extent to which messages in a sequence relate to each other, and especially the extent to which later messages recount the relatedness of earlier messages” (Rafaeli, 1997, p.3). In line with this definition, simple web browsing and reading of product information is therefore not considered as an interactive search activity hereafter. Instead, interaction with an agent system and detailed product screening will be considered in the model as two distinct, but complementing ways to search the web for relevant purchase information.

One construct frequently investigated in the context of information search is perceived risk. Perceived product risk denotes a consumer’s assessment of the consequences of making a purchase mistake, as well as of the probability of such a mistake occurring (Cunningham, 1967). As a result of this initial risk assessment consumers were shown to engage in information search in order to reduce the perceived risk to an acceptable level (Dowling and Staelin, 1994; Sundaram and Taylor, 1998). More precisely, risk was shown to be a multidimensional construct with consumers differentiating between functional, financial, social and psychological risk (Kaplan et al., 1974). Functional risk is the uncertainty that a product may not perform as expected, financial risk that the product will not be worth the financial price and/or would have been available cheaper elsewhere, socio-psychological risk
that a poor product choice will harm the consumer’s ego or may result in embarrassment before one’s friends, family or work group.

Probably, most risk dimensions relevant in the physical purchase process will continue to play a role in online environments. However, it could be that the degree to which individual risks are perceived is different online than offline. For example, being not able to touch and really see the product anymore, the socio-psychological risk might be higher in online markets than for their offline counterpart. In addition, there might be a new dimension of risk gaining relevance online, which is the delivery risk attached to a purchase. Buyers might fear that products won’t arrive on time or be in perfect condition. Because there was no delivery service included in the experimental store, delivery risk has not been respected for the measurement of risk in the current model.

The influence of perceived purchase risk on information search has been investigated in a myriad of studies (Cox, 1967; Dowling and Staelin, 1994; Kaplan et al., 1974; Srinivasan and Ratchford, 1991). Also for in-home shopping environments its relevance has been confirmed (Sundaram and Taylor, 1998). In his meta-analysis of the risk construct Gemünden (1985) concludes, however, that perceived risk seems to be particularly valid for high-involvement goods and less so for commodities, because lower levels of product risk do not trigger information search as a risk reduction strategy. As a result of these findings, perceived risk has been included in our model of online information search. It was expected that higher levels of perceived risk would lead participants to use both means of search in a relatively intensive manner. As former models of information search have suggested a mediating role of risk between exogenous variables such as product knowledge and information search (Srinivasan and Ratchford, 1991), perceived risk was considered as an endogenous variable in our model and it was hypothesized that:

H1: The more product risk a consumer perceives prior to the purchase of a camera, the more he or she will interact with an electronic advisor agent.

H2: The more product risk a consumer perceives prior to the purchase of a camera, the more will he or she consult detailed product information.
Exogenous Model Constructs

Referring to earlier information search studies, the concepts of cost and benefit of search, product knowledge, product experience and purchase involvement were included in the model of online search.

A recognized construct in structural equation models of information search (Punj and Staelin, 1983; Srinivasan and Ratchford, 1991) (and theoretical reflections thereon) (Moorthy et al., 1997) are the costs and benefits of search. Costs of search in these studies represent the accumulation of physical and cognitive effort as well as monetary expenditures necessary to find the right product. Benefits of search have been described as satisfaction with the product chosen or cost savings realized through the search activity (Punj and Staelin, 1983). Benefits have also been recognized in relation to the degree of uncertainty present in the choice of environment, risk aversion and the importance a buyer gives to the product category sought (Moorthy et al., 1997).

In an online context, perceived cost and benefits of search will probably continue to trigger search effort. Yet, especially the cost side may be of different nature online than offline. As was mentioned above, academics have pointed to a reduction of search costs in online environments (Alba et al., 1997). In fact, many traditional search cost variables (such as the physical effort to travel to stores, the implied transportation cost or the cost of cognitive effort to handle the complexity of product comparison) may be comparatively less important in online environments than offline. At the same time, two traditional information search cost factors, namely information processing time and ease of access to information, were shown to continue to play a role for online environments, their design and consumer product choice (Hoque and Lohse, 1999; Lynch and Ariely, 2000). Both of these cost factors are linked to the time investment a user is willing to make as part of the online search process. Thus, even though the time required for an online search is already minimal in comparison with the offline world, it still appears to play a role in the way consumers search for information. As a result, time cost has been included in our model of online information search. While the named studies investigated the information search cost construct only for user driven information search, referring mostly to product listings, the current model hypothesizes that time cost may be equally important in an interaction process with an agent. After all, consumers may weigh the number of specifications they make and potentially skip interactive
search categories (in our case any of the 7-agent-question cycles) in order to minimize time investment. Two hypotheses have been derived:

H3: The more time cost a consumer perceives while searching for product information, the less will he or she interact with an electronic sales agent.

H4: The more time cost a consumer perceives while searching for product information, the less they will consult detailed product information.

Costs of search have traditionally been outweighed by their benefits. For online environments this argument will probably continue to hold true. As in offline environments the benefits of search reside in the identification of an appropriate product. If consumers feel that interacting with an agent helps them to identify the right product they will probably be ready to invest into a relatively extensive dialogue (at least in a high-involvement context). If agent interaction is, however, thus beneficial, they will probably invest less effort into manual searching. To stress the relevance of perceived benefits from agent interaction for online information search, it was hypothesised that:

H5: The more benefits a consumer perceives from interacting with an agent, the more they will interact with it.

H6: The more benefits a consumer perceives from interacting with an agent, the less will he or she consult detailed product information.

Another construct that has continuously been shown to influence offline information search is product knowledge (Srinivasan and Ratchford, 1991, Beatty and Smith, 1987, Punj and Staelin, 1983). Yet, what consumers actually know about a product category (objective knowledge) and what they think they know (subjective knowledge) is often differing and may have diverging effects on search (Brucks, 1985). As a result, the empirical findings on how knowledge influences search have been contradictory. For the purpose of the current study there has been a focus on the knowledge consumers claim to have on a product category, because in the end it is this subjective feeling that will drive search effort. Subjective product knowledge was expected to limit searches by allowing responses to become routine or by allowing relevant information to be easier separated from the irrelevant, especially when
interacting with an agent system. On the other hand, it was thought that higher levels of subjective product class knowledge would lead subjects to increase manual search, since it allows one to delve deeper into information material. In addition, it was argued that those consumers who have more knowledge on a product also perceive less purchase risk (Srinivasan and Ratchford, 1991; Sundaram and Taylor, 1998). It was therefore hypothesized that:

H7: The more knowledge a person states to have about a product category, the less will he or she interact with an electronic advisor agent.

H8: The more knowledge a person states to have about a product category, the more will he or she consult detailed product information.

H9: The more knowledge a person states to have about a product category, the less risk will he or she perceive when confronted with a buying situation in the respective category in an online context.

A concept that has gained considerable recognition in the study of information search is the level of involvement a consumer has with the purchase situation (Beatty and Smith, 1987; Punj and Steward, 1983). Purchase involvement can be described as “a person’s perceived relevance of the object based on inherent needs, values and interests” (Zaichkowsky, 1985, p.341). Involvement is seen as a motivational factor in consumer choice behavior and is attributed mainly to three causes (Deimel, 1989): personal predisposition (i.e. subjective needs or goals), situational factors (e.g. time pressure) or stimulus-dependent factors (e.g. influence of product or communication). While situational involvement has been integrated in the model as a separate construct, stimulus-dependent involvement has been seized indirectly through the construct of product knowledge and perceived risk. Involvement was expected to play on both, agent interaction and manual search. A number of authors have argued that purchase involvement is also closely related to the consequences element of perceived risk (Beatty and Smith, 1987). It was therefore hypothesised that:

H10: The more involvement a consumer has with a purchase situation, the more will he or she interact with an electronic sales agent.
H11: The more involvement a consumer has with a purchase situation, the more will he or she consult detailed product information.

H12: The more involvement a consumer has with a purchase situation, the more risk will he or she perceive when confronted with a buying situation in an online context.

A number of studies have addressed the subject of consumer interactivity, and information exchange with first generation computer mediated environments. For example, based on the theory of exchange developed in marketing literature, Swaminathan et al. (1999) tested the impact of vendor characteristics, transaction security, privacy concerns and customer characteristics on the likelihood of electronic exchange. Other studies observed the importance of secure financial transactions for consumers’ perceived risk in online transactions (Parachiv and Zaharia, 2000). By far the greatest research attention has, however, been attributed to the impact of privacy concerns on information exchange (Culnan and Milberg, 1999; Swaminathan et al., 1999) and to the existence of flow in online navigation (Hoffman und Novak 1996, 2000). These two constructs, privacy and flow, have therefore been integrated in our online search model.

Privacy concerns of online users are a hotly debated issue. As mentioned above, studies confirm that consumers have great concerns about breaches of privacy. Ackermann et al. (1999), for example, found three distinct groups of online users with different levels of privacy concern: marginally concerned users, a pragmatic majority and privacy fundamentalists. Yet, despite these concerns many Internet users do not possess even rudimentary levels of online surveillance knowledge, and they do not use the available tools to protect themselves (Pew Internet & American Life Project, 2000). As a result, the relationship between privacy concern and subsequent behavior is unclear. Would users restrict online exchange in order to protect themselves? Swaminathan et al. (1999) suggested in an empirical study among 428 users that this might be so. However, as is the case with most privacy surveys, they only based their model findings on questionnaire data and (lag observations of consistent action. How might people react to a friendly anthropomorphic agent that gives good product advice in exchange for private information?

Privacy can be sacrificed by both interacting with an agent, or by simply navigating online sites. All activities are usually logged by several servers that host the content displayed on
users’ screens. However, as was outlined above, when consumers interact with advisor agents on website (which ask for key-words or retrieve personal data through dialogue-based systems) they reveal a considerable amount of direct personal information. Consumers were therefore expected to be particularly cautious when using the interactive applications leading to the hypothesis:

H13: The more privacy concern a consumer expresses over the revelation of personal data, the less will he or she interact with an electronic sales agent.

Another phenomenon apparently occurring when navigating in online environments is ‘flow’. The flow status is a psychological state that has been investigated in the context of intrinsic motivation since the 1960’s (Csikszentmihaly and Csikszentmihaly, 1995). Hoffman and Novak observed its relevance for online environments (1996, 2000) and defined it here (2000, p.23) as a “state occurring during network navigation which is: (1) characterized by a seamless sequence of responses facilitated by machine interaction, (2) intrinsically enjoyable, (3) accompanied by a loss of self-consciousness, and (4) self-reinforcing.” Thus, when consumers search for information online, it is possible that they lose perception of time and keep on navigating longer and in more directions than they initially planned to. Seen the creation of flow in online environments, the aim was to control this phenomenon with the following hypotheses:

H14: The more flow a consumer perceives, the more will he or she interact with an electronic sales agent.

H15: The more flow a consumer perceives, the more will he or she consult detailed product information.

Finally, it is intuitive to suggest that online consumers who used physical retail channels to get an overview of the product category and are thus more advanced in the buying process than their peers, would engage in less information search online than those who entered the online search process unprepared. The reason for this is that in interacting with an agent, informed customers might already know what selection criteria are the most important for them and are able not only to reduce the number of search criteria to a reasonably small size, but can also make up their mind more quickly regarding the specifications they prefer. As
they know what they want, they may also be able to view product alternatives quicker and understand detailed product information more easily. Even though the stage in the buying process and product knowledge are related concepts, they have been distinguished for modelling purposes. Consumers could have felt knowledgeable about a product category without having gone to a store in advance of the online shopping trip. At the same time, subjects may have gone to a store before shopping online, but still felt little knowledgeable about the product category. Given this, it was hypothesized that:

H16: The further a consumer is advanced in the buying process, the less will he or she interact with an electronic sales agent.

H17: The further a consumer is advanced in the buying process, the less will he or she consult detailed product information.

H18: The further a consumer is advanced in the buying process, the less risk will he or she perceive when confronted with a buying situation in an online context.

**METHOD**

In winter 2000 an experiment was carried out with 151 participants to observe consumer information search behavior during an online shopping trip for a compact camera.

**Participants, Incentive Scheme and Briefing**

The experiment was advertised at Humboldt University Berlin, Germany. Its goal was described as a test of user interaction with a highly innovative and performing product search engine developed for online shop systems at the Institute of Information Systems at Humboldt University. The system we told students would be tested out on the basis of a ‘real-world’ shopping trip for cameras. The online environment we said would be hosted by the industrial sponsor of the project who did not wish to be named for the time being. All navigational data would be transferred to this company. If people were interested in a camera they were asked to sign up to participate in one of the shopping sessions organized in a computer laboratory at a pre-arranged time. If they chose to buy something in the store they had to spend their own
money. The main incentive to participate in the experiment was a 60% discount offered on all prices of cameras displayed in the store\(^1\) even though this still implied a minimum expenditure of 80 DM (around $60 to $80 in purchase power) in case of buying. In addition to the discount, participants were promised a personal feedback on the behavior we would measure for our purposes during the shopping session. 95% of the resulting participants were students from different university faculties, while the remaining 5% held different jobs. 55.8% decided to buy a camera during the experiment.

**Online Material and Apparatus**

The shopping trip took place in an online store called “MCJC Store” explicitly programmed for the experiment (for screenshot see Appendix). The main reason for choosing a self-developed experimental store instead of using log-file material from some conventional online retailer was that we wanted to observe online search behavior with a view to \(^2\)nd-generation interface-agent systems. For this purpose we needed a highly interactive environment offering users the possibility not only to specify hard product attributes, but to enroll in an online sales conversation. An animated 3-D shopping agent image licensed from Artificial Life was therefore used to assist the user in product search. 56 purchase related questions were developed that would treat different ‘harder’ and ‘softer’ aspects connected to the purchase situation (Annacker et al., 2000). On the basis of answers given by shoppers the agent would calculate a reliable Top-10 ranking from more than 50 different compact camera models in the store. All participants had high-speed access to the store so that no significant time delays were present in page loading.

The navigation opportunities participants encountered in the store were organized in three phases: When participants entered the experimental store they had a space for *orientation* (phase 1) where they had the possibility to view all products on offer one by one from a list. From there, users proceeded to the search engine where the anthropomorphic 3-D shopping agent Luci introduced herself and her purpose to the user and started a *communication* or *interaction* phase. The agent interaction phase relevant in our model (phase 2) was organized in 7 cycles of 7-10 purchase related questions that a user could run through with the agent.

\(^1\) Since project finances did not allow us to offer this discount to all buyers, however, the incentive structure was slightly refined such that a lottery after the shopping session decided on one out of 10 participants who would have the right to take the product for the 60% off.
The 7 question cycles were displayed to the user on a category survey page leaving him the choice to run through the agent questions in any order he preferred and to the depth he deemed necessary. Through a ‘dialogue control box’ (situated on the upper left part of the screen) we ensured during each cycle that users were aware of the questions still to come and control for the questions already answered or skipped (thus, looked at but explicitly left blank). Based on any number of multiple-choice answers provided by the participating shopper, Luci could be asked to calculate the Top-10 ranking of products. From this ranking list, customers could then view a more detailed description of each product and enlarge its photograph (phase 3). The detailed product description contained a brief marketing text on the respective model displayed, a small photograph and a fact sheet summarizing major product attributes. However, no brand names were displayed in the store on any of the products. The reason for this manipulation was that brand names were shown to serve as information chunks for consumers (Jacoby and Hoyer, 1981; Weinberg, 1981). “Information chunks are information that particularly relevant for the judgment of products and that are able to substitute or bundle a number of other information” (Kroeber-Riel and Weinberg, 1999, p.280). By avoiding brand names, it was ensured that all participants navigated under the same conditions and that superior levels of brand knowledge of some participants would not lead to uncontrollable ‘short-cuts’ by some subjects in identifying the right product.

With the three shopping phases orientation, agent interaction and detailed product inspection navigation resembled a typical offline store visit. The shopping process could be exited at any time and a purchasing decision could be made after the request for a product information page. In order to avoid information chunks and have people investigate products ‘neutrally’, no brand information was displayed, neither in verbal product descriptions nor on photographs. In average participants spent 24.8 minutes in the online store looking around for products with a time spread of 6.9 to 51.4 minutes.

---

The remaining participants received a small financial compensation. If someone had not bought, but won the lottery, he or she would go out empty.

2 Users were not forced to provide any answers though. Prior to purchase subjects were told that if they did not wish to communicate with the agent at all the ranking of products would be in random order.
**Procedure**

Before and after the shopping trip, all participants filled out a battery of questions in which model variables were integrated. Before the shopping trip information was requested on demographics, online experience, product knowledge and experience, perceived risk, involvement, privacy concerns and the stage in the buying process. After the shopping trip participants were questioned about flow variables, agent perception and the probability of having chosen a wrong product. As we asked for so many different factors before the experiment (in the first questionnaire), we do not expect to have primed subjects on any particular one of the variables tested.

Before entering the store, all participants were asked to sign a privacy statement and a consent of payment in case of purchase. The privacy statement made clear that log-files would not only be used for research purposes, but also handed on to a third party. The goal of the privacy statement was to create a navigational context similar to the WWW where data is usually collected not only by the host server, but also by third party servers (e.g. advertising companies). The consent of payment was necessary as we did not offer automatic credit card debit and also had no postal distribution service integrated in the online service. Both signatures supported the aim of raising participants’ consciousness for the consequences of purchase and surfing online. We felt this to be necessary, as the laboratory environment and the University context might have otherwise led to an ‘unnaturally unconcerned’ type of interaction (experimenter effect).

Finally, as we wanted people to take their time shopping and not rush through the store we asked them to remain in the laboratory for at least 30 minutes. In order not to adversely affect their personal interaction feedback, however, we also told them to remain no longer than necessary in the store and leave it once they felt shopping to be over. Employing this time-manipulated set-up we artificially avoided some of the influence of time cost that is usually present when people surf and buy online (Hoque and Lohse, 1999). We did so consciously, because if we had given people freedom in time we would have had many users with different personal time agendas leading to uncontrollable earlier break-ups. We wanted to avoid this, for in the current study it was more important for us to control model variables than to observe the absolute time investment users make to decide on a purchase (other studies that are based on conventional log-file analysis can do so much more effectively).
MEASURES

Measurement of Endogenous Model Constructs

Measurement of the information search behavior

In the literature on offline information search, search activity has typically been operationalized by the time employed, the number of stores visited, the number of product alternatives inspected, the number of friends consulted etc. (Beatty and Smith, 1987; Punj and Staelin, 1983). For the purpose of the current study, measuring information search levels had to be adjusted to the electronic medium. While the relative amount of time spent searching was kept as one factor representing the search effort, the number of page requests was added as a second measure. Time was recorded for interaction with the electronic agent (phase 2) and for the two product inspection periods (phases 1 & 3). The time for interaction with the agent has been represented through the total time spent on answering agent questions and going back to the 7 category survey-page. The number of page requests in the context of agent interactivity stands for the intensity of exchange a user sought with the electronic agent. As was described above, the agent asked 56 purchase related questions, each of them representing a separate page. Users could return to this interactive functionality at any time during the shopping process and modify answers initially given. This activity of modifying specifications added to the number of pages requests in the interaction cycle as well as the time spent on the functionality. Finally, calls for the Top-10 ranking originating from the agent dialogue or the 7 category survey-page have been added to the number of page requests representing the breadth of agent interaction.

The number of individual product alternatives viewed added to the manual search construct. Each camera model on offer in the online shop was described on a separate html-page that could either be viewed in phase 1 or in phase 3. In addition to this detailed description, users had the possibility (in phase 3) to enlarge the photograph of each object in a separate page. The number of photo enlargements have been added as additional page requests to the construct of manual search. Finally, product descriptions were always requested from a page that listed the models available; either the Top-10 product ranking or the initial product orientation list (in phase 1). Together, product model lists, factual descriptions and photo enlargements made up the number of page requests for the dependent manual search.
construct. For all these pages time has been recorded and taken as a second measure. Both measures, time and page requests, are extremely precise measures of search when compared to the effort recall measures traditionally used in offline studies on information search.

Both time and page requests were recorded until a participant ended the search process which could be done either by pressing the ‘buy-button’ or the ‘exit-button’. Time and page requests were also the only model constructs that were automatically recorded by the system. All the other measures were derived from participants’ answers to pre- and post-shopping questionnaires.

It could be argued that the choice of time as a metric for the search undertaken is questionable since subjects have been asked to stay for a specified minimum of time at the lab. The time-cost factor that is usually present in shopping activities was therefore slightly manipulated. In fact, briefing the participants in this way may have led to a reduction in the variance of the time variable. However, the variance finally observed can be attributed more effectively to the constructs tested and is less subject to personal motivations in time management that would otherwise have gone uncontrolled. In addition, most of the subjects spent more time in the laboratory than they had to. It can therefore be argued that time is still a good measure; particularly as it was only important to observe the relative differences in behavior present in treatments with the same time conditions.

Measurement of Perceived Product Risk

Previous work was referred to in order to measure product category risk. As was outlined above, perceived risk has been characterized as a multidimensional construct with people differentiating between several negative consequences of a purchase including functional, financial, sociological and psychological risk (Kaplan et al., 1974). For the current model, risk dimensions have been combined into one overall index (that has been proposed and tested by academics in earlier studies (Peter and Tarpey, 1975)). As a result, risk has been captured in the following way:

$$OPR_j = \sum_{i=1}^{n}(PL_{ij} \cdot IL_{ij})$$
with \( \text{ORP}_j = \text{overall perceived risk for brand } j \)
\( \text{PL}_{ij} = \text{probability of loss } i \text{ from the purchase of brand } j \)
\( \text{IL}_{ij} = \text{importance of loss } i \text{ from purchase of brand } j \)
\( n = \text{risk facets (here } n = 4) \)

\( \text{OPR} \) contains two components: “…a chance aspect where the focus is on probability (of losing) and a ‘danger’ aspect where the emphasis is on severity of negative consequences of purchase” (Kogan and Wallach, 1964 cited in Peter and Tarpey, 1975, p.30).

In the pre-shopping questionnaire, risk perception was measured by employing a 15-point scale for both dimensions, probability and importance of loss. In order to calibrate the way in which different people respond to scales, each individual had to rate not only camera purchases, but also potential dangers and probabilities of loss associated with ‘extreme products’ in terms of risk, namely toothpaste and used automobiles.

**Measurement of Exogenous Model Constructs**

In order to measure time cost, earlier studies were considered which have introduced the idea of measuring time cost as opportunity cost. For example, Srinivasan and Ratchford (1991) measured time cost by asking people for their general time constraints and implied that this perception would be a measure for the opportunity cost perceived while searching for product information. In the present study, time cost was therefore grasped similarly by asking participants after shopping whether they had had the feeling during search that they would have rather done something else.

The problem in specifying the benefit construct is that, strictly speaking, benefits are not an antecedent, but a result of search. More precisely, perceived benefits of search are the anticipated result of each additional search step performed (Moorthy, 1997; Weitzman, 1979). Studies that measure the benefits of search should therefore try to capture either expected or ongoing benefits of search. This, however, has turned out to be a challenge. Either studies referred to the post satisfaction with the product bought (Srinivasan and Ratchford, 1991) or employed very general measures testing for consumers’ backward belief in the merits of the search activity. Doing so, self justification may have impacted responses. On the other hand, measuring expected benefits of search prior to the actual search taking place carries the risk to
prime subjects’ behavior. The measurement problem was attempted to be circumvented by taking the perceived quality of agent recommendations as an indicator for perceived search benefits. Doing so, neither self-justification effects were present in our measure nor have subjects been primed. Instead, it has been possible to capture participants’ ongoing impression of the quality of exchange.

For the measurement of product knowledge and involvement, measures have been used in the current study that have been proposed and tested in earlier empirical works (Moore and Lehmann, 1980; Srinivasan and Ratchford, 1991).

For the measurement of the two variables privacy and flow identified to be relevant for online environments parts of recent studies on these subjects have been employed. To measure privacy concerns some of the scales developed by Ackerman et al. (1999) were used. Participants were asked ten questions reflecting to what degree they would be ready to reveal certain types of information about themselves on a web site, including identification information (e.g. address or name) and profiling information (e.g. hobbies or income). The arithmetic mean of answers given to these 10 questions provided an index for participants’ privacy concerns.

Flow is a construct that is relatively complex to measure. In psychological experiments conducted by Csikszentmihalyi and Csikszentmihalyi (1995), the so-called Experience Sampling Method (ESM) has been employed which involves permanent and unexpected measurement of the current state of consciousness during an activity. Thus, upon a notification signal of a transmitter that subjects have to carry with them, they are required to respond to a short questionnaire (so called random activity information sheet) testing their current state of being. As a constant measurement of flow was not practicable in the shopping experiment, an additive index has been developed that is based on a number of questions capturing the flow experience as defined by Csikszentmihalyi and Csikszentmihalyi (1995) and Hoffman and Novak (2000). The questions used to measure flow were derived from the random activity information sheets used in ESM experiments and attempted to capture what Hoffman and Novak (2000, p.24) characterized as the cognitive state of flow on the Web which would be “determined by (1) high levels of skill and control, (2) high levels of challenge and arousal and (3) focused attention and (4) is enhanced by interactivity and telepresence”.

19
Finally, the fact that some participants had gone to a physical retail outlet was taken into account in advance of the experiment. There, some had already chosen products of interest for themselves that they now wished to buy for a 60% discount in our online store. Even though the online store made it difficult for them to rapidly identify their consideration set, because there were no brand names displayed, these subjects might still have behaved differently to those who were not informed. Subjects were therefore asked in advance of the buying session whether they had informed themselves of the product they wanted to purchase before coming to the lab and also to what degree they had already decided on products (consideration set). The two answers given were then combined to one index entitled *Stage in the Buying Process*.

A major limitation of construct measurement is that constructs usually did not have more than 1 or 2 indicators. More precisely, the model captures 4 constructs (privacy concern, flow, perceived risk, stage in the buying process) with the help of an index, 4 other constructs (purchase involvement, product knowledge and the online search variables) with the help of 2 indicators and finally, costs and benefits of search with only one indicator. The reason why model constructs had to be concentrated in this way is that for structural equation modelling the recommended ratio of sample size to number of free parameters is about 5:1 (Bentler and Chou, 1987). As was mentioned above, the study was restricted in terms of sample size, which implied that the number of free model parameters had to be minimized. Using reliable indices as construct representatives was a reasonable strategy to do so.

**RESULTS**

**Data**

Before model estimation, the data was screened for outliers which led to an exclusion of 6 from 151 observations. In addition, 29 subjects had missing data which we originally wanted to impute. However, imputing missing values by using a Maximum-Likelihood approach (Little and Rubin, 1987) implies multivariate normal data. Using PRELIS 2.30 (Jöreskog and Sörbom, 1996) we tested the assumption that the variables are normally distributed. The multivariate tests (Bollen, 1989) after listwise deletion of 29 cases with missing data show that the remaining data is significantly skewed ($z = 5.42$, $p = .000$) whereas multivariate kurtosis represents a borderline case ($z = 2.45$, $p = .014$). An omnibus test on multivariate skewness and kurtosis ($\chi^2 = 35.37$, $p = .000$) further indicates that the data is not normally distributed.
distributed, although deviation from normality seems to be rather modest and in the first place concerns indicators for information search behavior. As a result, model estimation had to be based on 116 cases.

**Model estimation and fit**

We used a structural equation modelling approach to simultaneously test model constructs and their relations. The model was estimated by the software program Mplus (Muthén and Muthén, 1998) which uses Maximum-Likelihood (ML) estimation as a standard modelling approach. Yet, since our data violate the normality assumption that underlies ML-estimation we decided to use the more robust MLM estimation option available in Mplus. This choice has an effect on the estimated standard errors for parameter estimates as well as the Chi-square test statistic, which corresponds to the re-scaled test statistic developed by Satorra and Bentler (1988, 1994). Although a recent simulation study has shown that this fit measure is not without shortcomings in small samples (Bentler and Yuan, 1999) the use of the robust estimation procedure seems to be warranted given the modest non-normality in our data and its superiority to the standard ML-estimation.

In an initial model estimation thus conducted with MLM four of our latent variable indicators had negative measurement error variances. These so-called “heywood cases” are a problem often encountered in structural equation modelling under the conditions of a small sample size and only two indicators per latent variable (Anderson and Gerbing, 1984; Boomsma, 1982). As both of these two conditions were unchangeable in our case we solved the problem of impropriety by employing a strategy pursued by earlier studies on information search where negative error variances have been set to zero (Punj and Staelin, 1983). Re-estimating the model with an error variance for the time variable of detailed product inspection (‘z2_dpd’) fixed to zero resolved the negative error variance problem also for the other variables.

Fit measures for this model were highly satisfactory. The RMSEA of .038 is considerably below the cut-off value of .05 (Browne and Cudeck 1993; Hu and Bentler 1999) and both CFI = .974 and TLI = .952 are above the threshold value of .95 (Hu and Bentler 1998). The explained variances of the endogenous variables of information search are of only moderate size (R² for “Interaction with agent” is .21 and for “Self-controlled search” .19) but are
considerably higher than those reported in former studies on information search behavior (e.g., Punj and Stealin, 1983).

The rather small sample size prevented a highly sophisticated operationalization of our theoretical constructs by multiple indicators. Nevertheless, based on parameter estimates for our model we tried to assess the reliability and validity of our two-indicator measurement models (see Table 3). For this purpose we used indicator reliability (Bagozzi, 1982), factor reliability (squared correlation between a construct and an unweighted composite of its indicators; see Bagozzi and Baumgartner, 1994) and the average variance extracted (Fornell and Larcker, 1981). Both, factor reliability and average variance extracted can also be regarded as measures for convergent validity. Since all these values are above their corresponding threshold values (Bagozzi and Yi, 1988) and as factor loadings were all significant, our construct measurements can be regarded as reliable and valid (see table 1).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Indicator</th>
<th>Indicator reliability</th>
<th>Factor reliability</th>
<th>Average variance extracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involvement</td>
<td>1</td>
<td>.908</td>
<td>.841</td>
<td>.747</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.431</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product class knowledge</td>
<td>1</td>
<td>.978</td>
<td>.811</td>
<td>.688</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.438</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction with agent</td>
<td>1</td>
<td>.848</td>
<td>.761</td>
<td>.615</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product inspection*</td>
<td>1</td>
<td>1.000*</td>
<td>.864</td>
<td>.761</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.626</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required level</td>
<td>≥.4</td>
<td>≥.6</td>
<td>≥.5</td>
<td></td>
</tr>
</tbody>
</table>

* NOTE. — error variance fixed to zero

Table 1: Reliability and Validity of Measurement Models

Model Relationships Found

Fit measures of the model indicate that the overall relationships hypothesized to exist for online information search sufficiently reflect reality. Interesting for the better comprehension
of online information search is, however, to what extend the hypotheses made hold true and at what level of significance they could be supported. Table 2 gives an overview of the findings.

<table>
<thead>
<tr>
<th>Explanatory/Exogenous variables</th>
<th>Perceived risk</th>
<th>Interaction with agent</th>
<th>Self-controlled search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived risk</td>
<td>-</td>
<td>-.022</td>
<td>.139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-.25)</td>
<td>(1.69)</td>
</tr>
<tr>
<td>Time cost of search</td>
<td>-</td>
<td>-.161</td>
<td>-.299</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-1.55)</td>
<td>(-3.63)</td>
</tr>
<tr>
<td>Benefits of search</td>
<td>-</td>
<td>-.190</td>
<td>-.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-1.65)</td>
<td>(-.20)</td>
</tr>
<tr>
<td>Product knowledge</td>
<td>-.232</td>
<td>-.375</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>(-2.06)</td>
<td>(-2.95)</td>
<td>(.05)</td>
</tr>
<tr>
<td>Involvement</td>
<td>.016</td>
<td>.315</td>
<td>.367</td>
</tr>
<tr>
<td></td>
<td>(.16)</td>
<td>(2.54)</td>
<td>(3.66)</td>
</tr>
<tr>
<td>Privacy</td>
<td>-</td>
<td>-.259</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.77)</td>
<td></td>
</tr>
<tr>
<td>Flow</td>
<td>-</td>
<td>.152</td>
<td>.164</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.57)</td>
<td>(1.75)</td>
</tr>
<tr>
<td>Stage in the buying process</td>
<td>-.081</td>
<td>-.077</td>
<td>-.059</td>
</tr>
<tr>
<td></td>
<td>(-.93)</td>
<td>(-.83)</td>
<td>(-.76)</td>
</tr>
</tbody>
</table>

**NOTE.**— Standardized solution. *t*-values are given in parentheses.

Table 2: Estimation Results for a Model of Online Information Results

In hypotheses 1 and 2 it was postulated that the more purchase risk a consumer perceives the more will he or she search for information. In fact, hypothesis 1 that users use an electronic agent more intensively when they perceive higher levels of risk was not confirmed by the data. In contrast, it was observed that participants tended to rely less heavily on the interactive functionality the more risk they perceived, even though this relation is not significant. At the same time, they consulted significantly more detailed product information the more risk they
perceived, confirming hypothesis 2. This finding suggests that consumers may engage more in manually controlled forms of search the more product risk they perceive. At the same time, they do not necessarily like to rely on an interactive functionality like agent Luci. In the section 4.5. below this phenomenon is commented on in more detail.

All exogenous constructs that were hypothesized to influence the perception of risk, namely product knowledge (H9), purchase involvement (H12) and the stage in the buying process (H18) proved to be in the right direction. However, none of them were statistically significant, except for product knowledge.

As far as the time cost of search is concerned, hypothesis 4 was supported. The data revealed that the more participants had wished to do something else while shopping online, the less they manually sought for information. The same was true for agent interaction (hypothesis 3), however not to a significant level. The results might indicate that agent functionality is relatively less impacted by consumers’ time constraints than are user-driven search forms. This, however, would have to be proven by more research.

In contrast to hypothesis 5, the more benefits a user derived from their interaction, the less he or she was willing to invest in the interaction process. In fact, since benefits of search were measured in the form of perceived accuracy of agent recommendation, it is intuitive to argue that the better the initial recommendation made by the agent, the less participants had an incentive to return to the interactive functionality to enhance or modify search parameters. However, even if this explanation is straightforward, the finding is still interesting because it raises awareness that one of the most basic assumptions made in information economics, which is that the more benefits one retrieves from information search, the more one searches for information, might be significantly impacted by agent technology (at least if benefits are measured in terms of identifying the right model). This impact resides in the possibility that the perceived utility of search renders decreasing marginal returns of search much quicker than this was the case for offline markets. The result is an inverse relationship between perceived search benefits and the activity of search. More research is certainly needed to investigate this finding and test its impact on the cost-benefit construct in information search theory. Hypothesis 6 that the more benefits a consumer perceives from interacting with an agent, the less will he or she consult detailed product information was supported by the data, however not at a significant level.
The traditional concept of product knowledge proved to be a reliable indicator for the prediction of interaction with the agent. Hypothesis 7 that the more knowledge a person states to have about a product category, the less will he or she interact with an electronic sales agent was shown to be significant at the highest level. In contrast to this finding, there was almost no effect of product knowledge on self-controlled search. Thus, people who think that they know a lot about a product relied less on an advisor agent, spending less time and effort on interaction with it. At the same time, almost no relationship seems to exist between product knowledge and the tendency to invest time and effort in manual search.

Another traditional search factor which proved highly significant for both parameters of search, agent interaction and detailed product inspection, was product involvement (H10 and H11). The more involvement a participant had with the purchase situation, the more he or she used both information sources available from the online store.

In summary, most of the traditional information search factors identified for offline markets were supported by the online model, with more than half of them at a significant level. Only two relationships did not hold true, namely the impact of perceived risk, and search benefits on the interaction process with the agent.

Hypothesis 13 that privacy concerns would be negatively related to consumer willingness to interact with the agent system was corroborated by model results. In fact, the data does not only support hypothesis 13, but also suggests that privacy concerns may have the strongest impact on agent interaction amongst all variables tested. This finding means that marketers who employ highly interactive technologies on their web sites should, in their own interests, pay attention to the privacy conditions they offer to their customers. However, it should also be noted here that in average more than 85% of the agent’s questions were answered by the participants. This is surprising, because answering agent questions is much more informative about a user than his navigating a site. Users’ privacy concerns seem to have expressed themselves more in a restriction of navigation (measurable in time and page requests) than in a reduction on information disclosed.

The flow construct introduced by Hoffman and Novak (1996, 2000) for Web navigation proved significant to the model. The data confirmed that participants who perceived more
flow searched significantly more manually (hypothesis 15). This positive effect was, however, not significant in as far as the shopping agent was concerned (hypothesis 14).

Finally, the data supported at a non-significant level that the more participants were advanced in the buying process, the less would they interact with the advisor agent (hypothesis 16) or manually search for information (hypothesis 17). As there were no brand names displayed in the store, the strength of this finding must, however, be regarded with caution. In case of brand display the negative effect on information search could have been stronger, with participants going directly for their consideration set.

CONCLUSION

The structural equation model we proposed for drivers and impediments of online information search displayed a very good level of fit and supported the majority of hypotheses made. As a result, we were able to show that determinants of information search identified in former offline studies, including product knowledge, purchase involvement and time cost seem to hold true for the online world. Furthermore, prove could be made of the influence of new variables such as privacy concerns and the achievement of a flow status.

In addition to the confirmation of these relationships a number of interesting findings have been made that, in our view, deserve further research confirmation. These include the observation that consumers who perceive higher levels of risk prior to a purchase seem to rely less strongly on agent advice than their peers and prefer to consult the more controllable information detail available on products. Another aspect is consumers’ curious handling of privacy concerns that on one side seem to be significantly addressed by decreased levels of interaction, but on the other hand also seem to be ignored when it comes to actual information disclosure.

The particular benefit of the study in the way we conducted it is that we were able to observe the ‘pure’ and instantaneous impact of different behavioural constructs on information search. Thus, we were able, for example, to exclude the impact of brands on behaviour. Also we observed actual search behaviour taking place and did not have to rely on self-reported activities (as former studies did). By using the electronic shopbot we were also able to win
insights into peoples’ dealings with this emerging type of technology and its relative importance in the information search process in comparison to detailed product descriptions. Here it was interesting to see that agents really represent only one way of searching for information and that, for example in situations of higher purchase risk, they may not be the preferred tool for users to decide on their products. This finding particularly questions the ‘zero-search-cost-assumptions’ of online information search discussed in the introduction.

Limitations of the current study are the relatively small sample investigated. Also, we were only able to observe information search in one store environment. It may be interesting to observe search in a more ‘natural’ way by respecting the use of different sites for different purposes (and thus also respecting different levels of handling the electronic medium). In addition, it appears sensible to compare search behaviour for a wider spectrum of products. For example, it would be interesting to see to what extend the model holds true for less expensive or less complex products or to what extend it can be transferred to other product categories such as experience and credence goods.
APPENDIX: Overview of Navigation in the Store

Communication and Interaction

Agent Interaction (Phase 2)

Manual Search (Phase 3)

Agent Question Cycles

Address Provision

Orientation (Phase 1)

EXIT

buy
REFERENCES


Pew Internet & American Life Project, Trust and Privacy Online: Why Americans Want to Rewrite the Rules (August 2000), [www.pewinternet.org](http://www.pewinternet.org)


Rafaeli, Sheizaf and Fay Sudweeks (1997), “Networked Interactivity”, *Journal of Computer Mediated Communications*, 2 (4)


Bezalel Gavish, American Telecommunication Systems Management Association, Dallas, Texas, 387-402.


