DIFFERENTIAL EFFECTS OF PRODUCT CATEGORY ON SHOPPERS' SELECTION OF WEB-BASED STORES: A PROBABILISTIC MODELING APPROACH

Oded Lowengart
Department of Business Administration
School of Management
Ben-Gurion University of the Negev
odedl@bgumail.bgu.ac.il

Noam Tractinsky
Department of Information Systems Engineering
Ben-Gurion University of the Negev
noamt@bgumail.bgu.ac.il

ABSTRACT

This study analyzes the probability of buying products from a particular online store given a set of alternative vendors. We use a multinomial logit choice model to analyze experimental data of consumer choice in two product categories - books and computers, each of which represents goods that differ along the risk dimension. Books represent search-quality, cheaper products, while computers represent experience-quality, more expensive products. Our modeling approach enabled us to capture the competitive environment of Internet shopping and at the same time obtain diagnostic information about the salient factors in the choice process. The results indicate the existence of differences in terms of the dimensions considered by consumers when buying high- vs. low-risk goods. Most notably, we found that when purchasing computers online, aspects of uncertainty and risk were more salient than when purchasing books online. We discuss the implications of these findings and areas for future research.

1. Introduction

One of the prominent questions in the area of Internet shopping from, both theoretical and practical viewpoints, is what factors affect consumer decisions to purchase from a specific Internet-based store. Attempts to answer this question should take into account three key dimensions of the purchasing decision.

One such dimension relates to the characteristics of the consumer. By this we refer to the set of general beliefs that consumers bring to the shopping environment, e.g., their general attitude towards the Internet or towards shopping (see, for example, McFadden, 1986, for a conceptual framework). A second dimension includes store characteristics. Consumer perceptions of stores are influenced by those store characteristics, which are used, in turn, to evaluate the store from which they want make their purchase. A third dimension refers to the type of product under consideration, since consumers may engage in different purchasing decision processes in different product categories (Kotler, 2000). For example, merchandise can be characterized using the distinction between search-goods versus experience-quality goods (Nelson, 1974). This distinction relates to how product quality can be determined. For some products, quality can be determined before the product is actually purchased (e.g., thickness of cloth), while for other products, quality cannot be determined prior to purchase and use (e.g., sharpness of an image on a television screen). Thus, search goods are products whose physical characteristics are known prior to purchase (e.g., books), while experience goods have some uncertainty with respect to their quality or higher likelihood of physical malfunctioning (e.g., a computer).

Evaluating purchasing decisions and interrelationships between the three dimensions involves different methodological approaches. One prominent distinction of research paradigms in consumer behavior distinguishes preference models from choice models. Preference models are commonly used to examine consumer preference formation toward a particular product and are usually based on the expectancy value model or a variation thereof (Fishbein, 1967). This modeling approach enables identification of the overall judgment of *one* alternative at a time and determination of the relative weight of the attributes that affect the preference toward that alternative. Choice models, in contrast, are simultaneous multiattribute designs that try to predict consumer behavior when faced with a set of alternatives. Unlike most preference models that focus on one alternative at a time, multinomial choice models, in general, can capture the extent of the effect perception have on choice among concurrent alternatives (Gensch and Recker, 1979). The choice modeling approach enables researchers to capture the *competitive* setting of a given scenario. That is, estimating the choice probability of

each store out of a set of competing stores. Furthermore, these models provide diagnostic information about the salient characteristics involved in the choice process.

The vast majority of empirical studies of consumer behavior in the e-commerce environment (or in a similar simulated environment) have focused on testing preference models (e.g., Jarvenpaa et al., 1999, 2000). Thus, it is important to gather evidence that relies on a different method to either support current theories or to suggest different ones. Furthermore, the current literature on consumer Web purchasing decisions has mainly concentrated on identifying the factors that affect the willingness of consumers to engage in Internet shopping (e.g., Jarvenpaa and Todd, 1997; Swaminathan et al., 1999). Research focusing on *which* store to choose from, however, is very scarce (see Lohse and Spieler, 1999, for an exception) and therefore does not provide adequate description of the competitive environment of Internet store selection. Gaining more understanding of this competitive environment will provide more insight into the choice process of Internet shopping and will help store managers focus on significant design dimensions that will eventually improve their competitive position.

The current research, therefore, is aimed at addressing the following issues. 1) At the methodological level, we study consumer purchase decisions among Web-based stores in a competitive environment by using a probabilistic choice modeling approach. 2) In terms of the type of the decision made by the consumer, we concentrate on consumer decisions regarding *which store to buy from* rather than *whether to buy or not to buy from an Internet store,* and the relevant characteristics that have an effect on this decision. Furthermore, we capture the heterogeneity between product categories by comparing decisions to purchase low-risk (search) goods with decisions to purchase higher-risk (experience-quality) goods.

2. Previous Research

Several empirical studies have examined the general tendency of consumers to shop on the Web and the factors that affect their decisions. In one of the earliest studies on this subject, Jarvenpaa and Todd (1997) used an experimental survey of 220 Internet shoppers to validate a model of attitudes and shopping intention towards Internet shopping in general. The model included several perceptual indicators that belonged to four major categories: the value of the product sought by the consumer, the shopping experience, the quality of service offered by the Web site, and risk perceptions of Internet retail shopping.

In a survey of 428 Web shoppers who were recruited by Web advertisements (i.e., not randomly sampled), Swaminathan et al. (1999) studied Internet users' attitudes towards Web shopping. They operationalized three characteristics of the shopping environment: vendor characteristics, perceived security of shopping transactions on the Internet, and consumer privacy concerns. In addition, they took into account the preference consumers have to deal with people and the degree to which convenience affected their preferred shopping mode. They found that with both frequency of shopping and amount spent online as dependent measures, consumer characteristics dominated all other variables in terms of variance explained. In addition, vendor characteristics also impacted consumer behavior. Security and privacy issues, however, played only a minor role in affecting the reported shopping behavior of study participants.

In another study that surveyed general attitudes toward Web shopping, Li, Kuo, and Russell (1999) examined the effects of three perceived channel utilities: communication, distribution, and accessibility and four types of consumer shopping orientations: recreational, experiential, convenience, and economic. The survey results of 999 self-selected participants found that self-reported online buying behavior was affected by a mix of consumer shopping orientation and perceived channel utilities.

Vellido, Lisboa and Meehan (2000) studied responses of 2,180 individuals to 44 items from GVU's Ninth WWW User Survey (Kehoe et al., 1998). They extracted nine factors associated with user opinions of online shopping. Among those factors, risk perception of users was demonstrated to be the main discriminator between those who reported that they purchased products online and those who reported that they did not. Other discriminating factors were control and convenience over the shopping process, affordability of merchandise, customer service, and ease of use of the shopping site.

While the aforementioned studies have dealt with factors that affect shopping in general on the Web, other studies have focused on what makes shoppers buy from one specific store as opposed to another. Lohse and Spieler (1999) found, based on real sales data from 1996, that usability features of a Web store (such as the existence of a product list to reduce navigational barriers) contributed significantly to variations between stores in terms of consumer traffic. Consequently, they suggest that information search costs are a major determinant of both store traffic and sales. However, their study did not distinguish between types of stores. Thus, the lack of distinction between store types or the different types of products sold by the different stores might have tainted the results of this study.

Jarvenpaa, Tractinsky and Vitale (2000) tested a model of consumer attitudes towards specific Web-based stores, in which perceptions of the store's reputation and size affect consumer trust of the retailer. The level of trust, in turn, was positively related to the attitude toward the store and inversely related to the perception of the risks involved in buying from that store. Finally, the attitude and the risk perception affected the consumer's

intention to buy from the store. The model was supported by data gathered in three countries: Australia, Israel and Finland (Jarvenpaa and Tractinsky, 1999).

Perhaps because of the relatively early stage of the research on Internet-based retail, most studies of consumer behavior on the Web still do not distinguish between different types of retail merchandise sold through this medium. For example, Lohse and Spiller (1999) studied Internet "superstores" that offer a large selection of products. Vellido et al. (2000) sampled responses to a survey that did not distinguish among type of products or stores. Consequently, their results may constitute an "averaged" effect of various factors that affect consumer behavior, but which might mask effects that are contingent on the type of goods sought by the consumer. Indeed, some evidence already exists to support the notion that merchandise type has a moderating effect. For example, Jarvenpaa et al. (2000) tested a model of antecedents of consumer intentions to buy from Web-based stores. They found that perceptions of the size of an Internet-based retailer more significantly affected consumer trust in the store when the merchandise was expensive and of the experience-quality type (e.g., travel packages) than when the products under consideration were relatively inexpensive search-quality goods (e.g., books).

In summary, risk was found to be a major determinant of Internet shopping. Risk is comprised and affected by security, privacy issues, trust, reputation of the retailer, size of the store, and overall risk perception. These elements of risk were found to have a discriminating effect in on-line purchasing decisions. When considering salient factors that might affect consumers considerations for online purchasing, product value, shopping experience, service quality, user control, price, and ease of use and information search were found to be relevant.

3. Conceptual Framework
In light of this review, there is a need for comparative research that will explore whether online shoppers react differently to different types of goods. In the following section we present a short review of a product typology that is based on the level of risk associated with the product.
3.1. High-Risk vs. Low-Risk Products
One of the most salient obstacles to the success of Internet retail has been consumer perception that Internet shopping involves excessive risk (Hoffman, Novak and Chattarjee, 1995, Jarvenpaa and Todd, 1997). It is reasonable to assume, therefore, that under varying degrees of perceived risk, shoppers will behave differently. It seems, however, that research providing empirical evidence for this claim is still wanting. We believe that Nelson's categorization of products meshes nicely with such research. The main focus of the research stream that compared search goods and experience-quality goods was on the information asymmetries that distinguish between the two types of products (Nelson, 1974). Consumers do not have complete information about experience-quality goods. The lack of information increases uncertainty and hence the risk associated with buying such products. In the case of search-quality products, however, consumers are already aware of the product's qualities and attributes, thus very little uncertainty is involved and the risk is relatively small, *ceteris paribus*. Information asymmetries that result from such differences (i.e., experience and search) also exist in the service sector (Nayyar, 1990). Similarly, whereas conventional shopping entails less uncertainty, as relatively more qualities of the store can be detected by the consumer, Internet shopping involves greater uncertainty and higher risk. Consequently, it is apparent that different types of products (e.g., books vs. computers), sold in different types of outlets (e.g., brick-and-mortar vs. Internet) might be perceived and evaluated differently by consumers with regard to their implied risks. As a result, different product and service/outlet types will have differing effects on the consumer's purchasing decision.

Since this study deals with comparing the selection processes of web-based stores for different product categories (i.e., high-risk and low-risk products), we selected two different products that are sold extensively on the Internet: books and computers. Books represent a product category (search quality) for which consumers have no doubts about the quality of the product. They might like or dislike the content, but will experience almost no uncertainty about the quality of the physical aspects of the book (e.g., readability, weight, type of cover, etc). Thus, buying books exemplify the context of purchasing low-risk products for studying the choice of a web-based store. Computers, on the other hand, represent a category of products whose true quality can only be evaluated after being used by the consumers. Not only does the complexity of the product create uncertainty about its quality, but it also integrates numerous components, all of which have to perform flawlessly in order for the entire product to be of high quality.
It might be argued that given the wide and extensive use of computers in many daily activities, some consumers already see computers as a commodity. Yet, we argue that this is not the case. Various trade magazines regularly conduct customer satisfaction surveys regarding computers' reliability, which are aimed mainly at providing information to potential users. For example, in a recent PC magazine poll, there were considerable differences in reliability among several leading computer brands (Howard, 2001). Contrast this with books: to the best of our knowledge, surveys regarding the reliability and user satisfaction from their book performance do not exist. Furthermore, a market that is characterized by product differentiation, branding of products by firms, building

brand equity and commending price premium over competitors, as the case in the computer market, defies the notion of a commodity market. We take this as indication that marketers, computer professionals, as well as users think that computers are not yet a commodity.

As such, computers are riskier than books in terms of the uncertainty associated with their performance. Uncertainty, however, is not the sole source of risk in product evaluations. Consumers also assign certain risk levels to the potential damage they might incur if their expectations of product performance are not met. This type of risk increases with the price level of the product. Since computers are more expensive than books, if anything malfunctions, the price of an erroneous purchase of computers is higher than that of a book. Thus, shopping for a computer represents the context of high-risk products in studying the choice of web-based stores.

3.2. Study Approach

To summarize, our approach to this study is based on the premise that consumer online purchasing decisions involve several related aspects that are depicted in Figure 1.
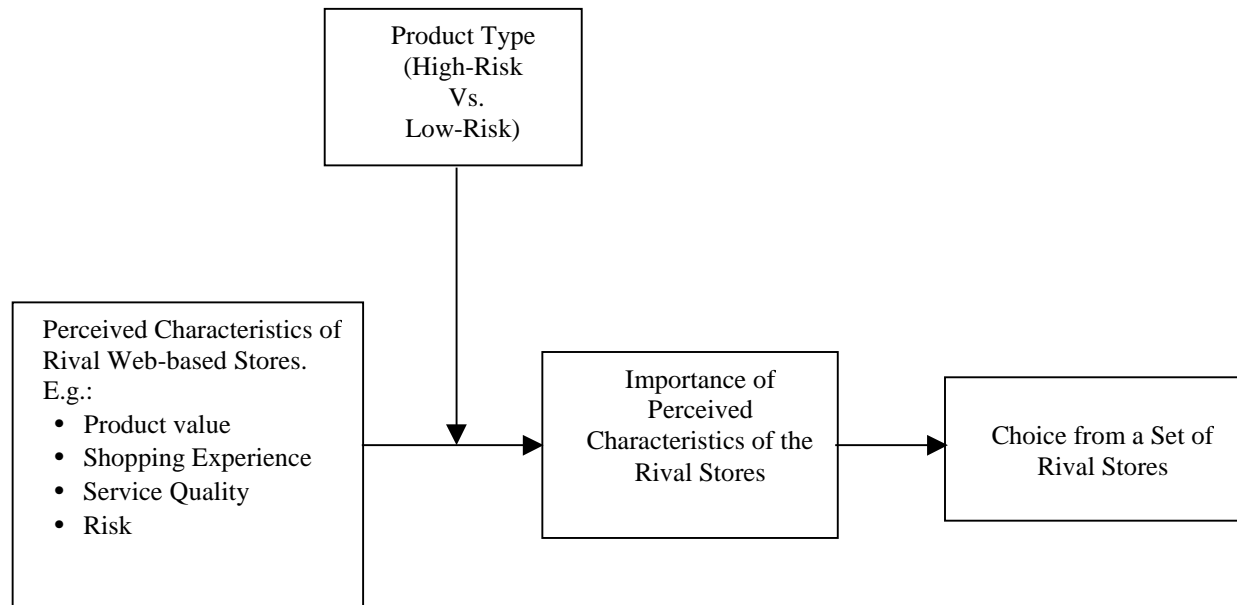

Figure 1: The Research Model: The model suggests that product risk affects the importance of various perceived characteristics of the store in determining choice among stores in a competitive environment.

The first aspect is the type of product that the consumer intends to buy. Here we distinguish between two product types that elicit different evaluations from consumers. Such product evaluations are based on the different risk levels associated with the products, which stem from the uncertainty associated with the functionality of the product and the product price. Because less uncertainty about functionality is inherent to search-quality products, they are associated with lower risk than experience-quality products. It then follows that more expensive products are associated with greater risk than cheaper products. In this research study we covary the dimensions of product uncertainty and monetary expenditure to better distinguish between the products. Consequently, we regard expensive, experience-quality goods such as computers as high-risk purchases and search-quality, inexpensive goods such as books as low-risk purchases. Of course, by this we are trading off the ability to distinguish between the effects of price and product type in order to strengthen the distinction between high- and low-risk products.

The second aspect of the shopping process relates to the characteristics of the stores that consumers consider when shopping online. Here, different characteristics of the Internet store carry different importance as a function of the product sought. For example, buying a low-risk product on the Internet would probably require less attention to the retailer reputation (which signals long-term operations with fewer failures). Purchasing high-risk products, on the other hand, would probably be more influenced by such characteristics.

The third aspect involves the comparison of a set of rival stores. In an ordinary shopping process, store characteristics are generally evaluated with respect to the characteristics of other stores. To remain loyal to this process and to increase the study's external validity, our research simulates the competitive environment of buying a particular product on the Internet and its impact on consumer evaluations of the different shopping sites.

The integration of these aspects implies a process in which consumers first decide what product to buy and then engage in a simultaneous comparison of alternative stores that offer the product. Thus, our major research goal is to find whether shopping for different product types will be reflected in consumer considerations of

which Internet site to buy from. Consequently, the study's main research question is: What are the similarities and differences in a store choice process of Internet shopping between different contexts of product risk level?

4. Method

Our research approach included several stages. Based on our review of the existing literature, primarily on Jarvenpaa and Todd (1997), we first devised a 22-item questionnaire regarding factors that affect consumer intentions to buy from a specific Internet store. Next, we designed an experiment in which participants simulated two shopping tasks from actual Internet stores. The participants then answered questionnaires about the stores they visited. The results were analyzed in two phases. In the first phase we conducted factor analysis (one for each product category) to identify the relevant dimensions according to which participants evaluated the different stores. In the second phase we estimated the probability of each site being selected and the effect of the various dimensions on this choice process by employing a multinomial logit probabilistic choice model.

4.1. Measures

Participants' beliefs and perceptions about various aspects of Internet stores served as the variables of interest in the current study. Many studies have enumerated such aspects, with much overlapping among them. In developing measures for this study, we chose to rely mainly on the aspects that were identified in Jarvenpaa and Todd's seminal work (1997), in which the behavior of Web shoppers was examined through open-ended questionnaires. Four major dimensions were identified in that study: (1) Product value, which includes the variety of products and stores available, the quality and the price of the goods on the Web. (2) Shopping experience, which includes the effort required to shop on the Web, compatibility with the consumer's style of shopping, and the playfulness associated with the act of shopping. (3) Service quality, which includes the responsiveness of stores on the Web, the tangibility of the store's site (i.e., the quality of the information about the products), the empathy with which merchants accommodate consumer needs, and the reliability of the store in keeping its promises. (4) Risk perceptions of the consumers, which are composed of the risk of the product not performing as expected, personal risk primarily related to credit card concerns, loss of privacy, and monetary losses through poor purchase decisions. The items extracted by Jarvenpaa and Todd (1997), however, refer to Internet shopping in general and not to shopping in a particular context. Hence, we reworded the items to reflect relevance to a particular purchase. Some of Jarvenpaa and Todd's items, though, would not be relevant for a specific site (e.g., "shopping on the Web will be perceived as imprudent or socially unacceptable"). Thus, these items were excluded from the current study. In addition to Jarvenpaa and Todd's factors, we introduced seven items, which were raised by other researchers (e.g., Hoffman, Novak and Chattarjee, JCMC 3(1), Jarvenpaa et al., 2000, Lohse and Spieler, 1999).

The items were pretested in a pilot study, which took a very similar form to the experiment described in this section. Seventy-eight student volunteers participated in the pilot study that served to fine-tune the experimental procedure and the perceptual measures. The purpose of the pretest was to check the level of task difficulty and ensure that the respondent's interest and attention are obtained, the meaning of the questions is understood, and variation in responses exists. Based on the results of the pilot study and feedback from the participants, the instructions and several items were reworded to improve their clarity. The revised set of items is presented in Appendix 1.

4.2. Sample

114 first-year engineering students volunteered to participate in the study for class credit. 72 participants were male (63%) and 42 female (37%). The participants' average age was 23. 75% of them rated their knowledge in computers to be at least at the average level or more. All participants have used the Internet before (26% at least once a day and 58% several times a week), but only 11% had previously shopped on the Internet. 94.5% of respondents reported to have current access to the Internet, of which 13.5% have access from home, 35% from the university and 51.5% of respondents have more than one way to access the Internet (i.e., both home and university).

4.3. Tasks

Participants were asked to access the Web sites of three online Israeli bookstores and three online Israeli computer stores. These stores were selected based on their popularity among Israeli surfers (as measured by a leading Israeli portal, http://www.walla.co.il) and because they offered a relatively wide assortment of items of the relevant categories relative to other Internet stores. The computer stores carried generic computers and customers could customize their orders. System prices ranged from about $1,500 to $3,000. The experimental software randomized the order of tasks (book shopping and computer shopping) and the order in which participants accessed the stores within each task. The participants were asked to browse each site, examine its policies and merchandise, and, finally, look for a specific item (a book or a computer system) to purchase. Participants were asked to proceed with the purchasing process up to the final step of submitting an order. Upon completion of the task, the participants returned to the experiment's home page and answered an online questionnaire regarding the store that they had just visited. After completing the questionnaires for all three stores in each task, participants indicated which one of the three stores they would choose to buy from.

4.4. Procedure

The experiment was administered entirely by computerized software written in HTML and Java script that was installed on a LAN server. The software was run on 200 MHz Pentium computers in the departmental computer laboratory. Participants performed the experiment in groups of up to 10 people per session.

5. Results

5.1. Factor Analysis

In order to detect the underlying dimensions of the participants' evaluations of the stores, their perception ratings were subjected to a principal component-based factor analysis. Factors with eigenvalues greater than 1 were extracted and rotated using the VARIMAX method to form an orthogonal structure. Four factors, accounting for 63% of the total variance, were extracted from the data on the bookstores. For the computer equipment stores, 6 factors were extracted, accounting for 66 % of the variance. Cronbach Alpha reliabilities were calculated for the factors. With the exception of two factors all reliabilities ranged between .71 and .90. The low reliabilities of the two factors may be attributable to the small number of items in each factor (three). Because of this and because of their theoretical relevance we decided to retain these factors for subsequent analyses. Table 1 summarizes the results of the factor analysis for the two shopping tasks. The rotated factor structures, the item loadings, and the factor reliabilities are presented in Appendix 2.

The four factors that emerged from the data on bookstores can be interpreted as representing: (a) Shopping process (e.g., easy to find information in the site, shopping from the site is fun, visually appealing and provides detailed information on the products, and good customer service). (b) Product quality (e.g., the site is a source of high-quality products, products do not meet expectations, reliable information, and interesting products). (c) Risks involved in shopping from the store (e.g., good data security, privacy, reliable operators, and potential monetary loss). (d) Size and reputation of the store (interestingly, those are trust precursors from Jarvenpaa et al., 1999; 2000).

The six factors of the computer stores resemble those obtained from the bookstores, with several interesting differences. The first factor remains the same. It reflects, once again, the shopping process as measured by items such as site design and enjoyable and easy shopping. The second factor represents part of the second factor relating to books. In a sense, the data from the computer stores indicate that shoppers break the quality of the product into two: the quality of the information about the product (Factor 2, e.g., up-to-date detailed and reliable information, information about costs, the range of the products and their costs) and the quality of the product itself (Factor 5, e.g., high-quality and interesting products). Similarly, the third factor from the bookstore data has been separated into two distinct factors in the computer store data. One factor (Factor 6) deals specifically with monetary risks, while the other factor (Factor 3) deals with all the other risks mentioned in the questionnaire. Finally, Factor 4 remains the same as it was in the data set for bookstores.

Table 1: Comparison of the final factor structure for the data from the book purchase task and the computer purchase task.

Factor	Book Purchase	Computer Purchase
1	Shopping process	Shopping process
2	Quality: the product and the information about the product	Quality: information about the product
3	Risks: of all types	Risks: general
4	Store size and reputation	Store size and reputation
5	--	Quality: the product itself
6	--	Risks: monetary

These results are in keeping with our expectations regarding the differences between the book-shopping task and the computer-shopping task. That is, in decisions regarding the purchase of computers, issues of financial risk and information about the product (i.e., quality uncertainty) were more pronounced than in decisions regarding the purchase of books. These differences are further reflected by the percentage of variance explained by the various factors. The two risk factors in the high-risk (i.e., computer-shopping) task accounted for 10.9% of the variance compared to 7.6% of the variance explained by the single risk factor in the low-risk (book-shopping) task. Similarly, factors that represent the quality dimension of the products accounted for 13.0% of the variance in the computer-shopping task compared with 8.5% in the book-shopping task.

5.2. Choice Model

As noted earlier, the main objective of this research study is to estimate the probability that potential consumers will choose a specific site, from a set of alternative sites, for their online purchases. Furthermore, we strive to gain diagnostic information about this decision process by determining the salient variables (i.e., site characteristics or dimensions) involved. To obtain these objectives, we employed a probabilistic choice model to our data.

Our analysis includes the utilization of a multinomial logit (MNL) type choice model that incorporates the different characteristics of the different Internet shopping sites. Following the identification of the underlying dimensions involved in choosing an Internet store, we applied the MNL with the factor scores from the previous stage (see for example Gensch and Ghose, 1992 for an application of such method).

5.2.1. Multinomial Logit (MNL) Model

The MNL model is a simultaneous compensatory[1] attribute choice model that incorporates the concepts of thresholds, diminishing returns to scale, and saturation levels (McFadden 1974). It enables us to predict which Internet shopping site has the highest probability to be selected by consumers and to ascertain which store attributes are salient in this choice process. The MNL model is based upon the assumption that the overall consumer preference for a choice alternative (i.e., an Internet shopping site) is a function of the perceived relative utility that the alternative holds for the consumer. Therefore, the probability that Internet shopping store j will be chosen by customer i is given by:

$$P_{ij} = \frac{\exp^{(U_{ij})}}{\sum_{j=1}^{j=m} \exp^{(U_{ij})}} \qquad (1)$$

where U_{ij} is the utility of Internet shopping site j for customer i, and m is the number of alternative Internet shopping stores. The utility function can be separated into a deterministic component V_{ij} (measured in terms of perceived value associated with the characteristics of the Internet shopping sites), and unobserved random component ε_{ij} (independent and identically distributed with a double exponential distribution - Gumbel type II extreme value) such that we get:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \qquad (2)$$

5.2.2. Utility Specifications

Based on the results of the factor analysis, the deterministic component of the utility function for the <u>book</u> category has the following form:

$$V_{ij} = \alpha_1 SHOPPING\_PROCESS_{ij} + \alpha_2 PRODUCT\_QUALITY_{ij}$$
$$+ \alpha_3 RISK_{iJ} + \alpha_4 REPUTATION_{ij} \qquad (3)$$

where

$SHOPPING\_PROCESS_{ij}$ - ease of the purchasing process in book site alternative j

$PRODUCT\_QUALITY_{ij}$ - quality of the products in book site alternative j

$RISK_{iJ}$ - risk involved in purchasing from book site alternative j

$REPUTATION_{ij}$ - reputation and size of book site alternative j

$\alpha_1, \alpha_2 \alpha_3, \alpha_4$ - parameters to estimate.

Similarly, the deterministic component of the utility function for the <u>computer</u> category has the following form:

$$V_{ij} = \alpha_1 SHOPPING\_PROCESS_{ij} + \alpha_2 INFORMATION\_QUALITY_{ij}$$
$$+ \alpha_3 RISK_{ij} + \alpha_4 REPUTATION_{ij} + \alpha_5 PRODUCT\_QUALITY_{ij}$$
$$+ \alpha_6 FINANCIAL\_RISK_{ij} \qquad (4)$$

where

$SHOPPING\_PROCESS_{ij}$ - ease of the purchasing process in computer site alternative j

$INFORMATION\_QUALITY_{ij}$ - information given on products sold in computer site alternative j

$RISK_{iJ}$ - risk involved in purchasing from computer site alternative j

$REPUTATION_{ij}$ - reputation and size of computer site alternative j

$PRODUCT\_QUALITY_{ij}$ - quality of the products in computer site alternative j

$FINANCAIL - RISK_{ij}$ - risk of the transaction in computer site alternative j

---

[1] Compensatory choice process implies that the existence of a perceived high level of a desirable attribute in an alternative can compensate for a perceived low level of another desirable attribute in that alternative.

$\alpha_1, \alpha_2 \alpha_3, \alpha_4, \alpha_5, \alpha_6$ - parameters to estimate.

5.2.3. Data Analysis
Books

We begin our analysis with the book category, where the results of the MNL analysis are presented in Table 2. It can be seen that the only significant factor in this choice model is the shopping process. However, contrary to common wisdom, the results indicate that a less positive shopping process was associated with a greater tendency to purchase from a bookstore site. We discuss this result in the Discussion section. The other factors, particularly risk perception, did not contribute significantly to the model.

Table 2: MNL Coefficients - Book Category

Factor	Coefficient	P-value
Shopping process	-0.2897	0.0000
Product quality	-0.1377	0.1679
Risk	-0.0261	0.8024
Reputation	0.1086	0.1998
Log likelihood function	-112.9073	

In order to evaluate the accuracy of the model's prediction, we computed four measurements of goodness-of-fit. The first is a hit rate (a count of a correct model classification of each observation to the actual chosen site) which is measured at the individual level. The other three are the aggregate level measurements of mean absolute deviation (MAD), root mean square error (RMSE), and a chi-square test between the aggregate actual choices of the three Internet shopping sites and the predicted choice probabilities. For more explanation on goodness-of-fit measurements, see Hair et al., (1998). The results of these tests are presented in Table 3.

Table 3: Goodness-of-Fit Measurements - Book Data

Variable	Computed value
Hit rate	53%
x^2	10.670
MAD	17.971
RMSE	11.543

The various goodness-of-fit measurements reveal that the model predictions were moderate. The hit rate variable indicates that the model increased the accuracy of predicting consumer behavior to 53% compared with an expected 33% prediction accuracy of random choice (i.e., the "naïve" approach). Though this is an improvement over the naive approach, the x^2 entails that there is a significant difference between the predicted and actual sets of probabilities (p < .01), indicating low predictive accuracy. The relatively high values of the MAD and RMSE measurements also suggest that the predictions are only moderately accurate.

It is also possible that certain site-specific variables that were not explicitly measured by our instrument may add to the predictive power of the model. These site-specific variables capture the idiosyncratic effects of the sites (see, for example, Guagdani and Littlle, 1983 for a discussion of such variables in the context of brand choice). Unlike the previous four variables that represent consumer evaluations of the different Internet shopping sites based on common attributes, here we refer to unique characteristics of the different Internet shopping sites that are not shared with the others. In order to avoid singularity, only J-1 variables are included in the unrestricted (full) model, and the deterministic component of the utility function becomes:

$$V_{ij} = \alpha_1 SHOPPING\_PROCESS_{ij} + \alpha_2 PRODUCT\_QUALITY_{ij}$$
$$+ \alpha_3 RISK_{ij} + \alpha_4 REPUTATION_{ij} + \alpha_5 SPECIFIC_A + \alpha_6 SPECIFIC_B \qquad (5)$$

The results of this analysis, termed full analysis, are presented in Table 4.

Table 4: MNL Coefficients - Book Category - Full Model

Factor	Coefficient	P-value
Shopping process	-0.3052	0.0020
Product quality	-0.0727	0.4797
Risk	0.0034	0.9749
Reputation	0.0049	0.9585
Site specific A	-1.1724	0.0014
Site specific B	0.2485	0.4223
Site specific C	0	

Log likelihood function	-101.8319

In order to verify whether the full model better fits the data than the restricted model (i.e., without the store-specific variables), we performed a log-likelihood test on the log-likelihood function values. This test is $-2 \log \lambda$, where $\lambda = (LL_{restricted} - LL_{unrestricted}$, $LL_{restricted}$ is the restricted model's log-likelihood function, and $LL_{unrestricted}$ - is the full model's (unrestricted) log-likelihood function (for more details see Gensch, 1985). This resulted in a statistic value of 22.15, which is x^2 distributed and significant at least at the 0.001 level. This indicates that the two models are different from one another and that the addition of the site-specific variables has significantly improved the model performance. It can also be seen that this difference is most noticeable in site A (i.e., the low market share site).

It can be argued that this site has some unique perceptions that are not shared by the other sites. It should be noted, however, the site-specific variable captures both the unique effect of a site (as explained above) as well as the effect of possibly omitted other common variables (e.g., time spent in the site). Namely, this variable contains two components that cannot be separated. Thus, we cannot claim that all the difference between the restricted and unrestricted is attributed to the site-specific variables.

We also computed the goodness-of-fit measurements as in the previous stage. All measurements, presented in Table 5, indicate improved prediction relative to the restricted model with a higher hit rate and lower values of x^2, MAD, and RMSE.

Table 5: Goodness-of-Fit Measurements - Book data with site-specific variables

Variable	Computed value
Hit rate	56%
x^2	8.477
MAD	15.652
RMSE	9.604

Thus, the result of the log-likelihood test and the goodness-of-fit measurements indicate that the addition of the site-specific variables significantly improved the data fitting of the model.

Computers

We now present the results of the computer category. In Table 6 we present the MNL coefficients and significance level.

Table 6: MNL Coefficients - Computer Category - Restricted Model

Factor	Coefficient	P-value
Shopping process	0.2023	0.0055
Information quality	0.1958	0.0393
Risk	0.0914	0.3888
Reputation	0.0081	0.9247
Product quality	-0.0310	0.7396
Financial risk	-0.1936	0.0476
Log likelihood function	-118.5655	

The MNL results indicate that several aspects influenced the choice among the computer shops. The shopping process, the quality of information about products, and the financial risk are salient dimensions in this choice process. In other words, it appears that if a vendor wants to increase the probability that consumers will purchase experience-quality, high-risk products from its site, it must improve the shopping process, maintain accurate and complete information about its products, and employ risk mitigation mechanisms to convince consumers that they are not "throwing their money away".

A log-likelihood test of the full model that includes site-specific variables against the restricted model (as in the previous section) resulted in a statistic value of 4.23, which is not significant. This result indicates that the addition of the site-specific variables did not have a significant effect on the data fitting of the model.

As in the book category, we also calculated the four measurements of goodness-of-fit and present them in Table 7. Overall, it can be seen that the model performed slightly better on the aggregate level than on the individual level (i.e., predicting market share as opposed to predicting individual consumer choice). The x^2 - of 95% for the restricted model and of 92.5% for the unrestricted model indicate a very good fit of both models at

the aggregate level. While the addition of the site-specific variables somewhat improved the aggregate measurements, it lowered the individual measurement slightly by 1%. We take those marginal changes in predictive accuracy as an indication that the factors identified for the computer-shopping task were adequate, and that additional site-specific factors could not improve this accuracy any further.

Table 7: Goodness-of-fit indices for the restricted model and the full model

Variable	Restricted Model	Unrestricted Model
Hit rate	55%	54%
x^2	0.872	0.312
MAD	4.637	2.898
RMSE	3.279	7.787

Overall, it can be seen that our model yield more accurate results in the computer-shopping task than in the book-shopping task. We elaborate more on this difference in the following section.

6. Discussion

This paper presents a study of factors that affect consumer purchase decisions on the Internet. The study stands out from most previous studies in this area through its emphasis on three main dimensions. First, its dependent variable is consumer choice of one out of a set of Internet stores rather than consumer ratings of the stores (see for example, Einhorn and Hogarth, 1981, for discussion about the differences between choice and judgment). Second, it is about choosing from a specific store rather than about buying on the Web in general. Finally, our study compares the determinants of consumer choice for low-risk versus high-risk products, a distinction that has generally been ignored in previous studies. We believe that this approach allowed us to gain new insights into the factors that affect online shopping.

We have assembled a set of factors purported by researchers in the field of B2C to have an effect on consumer choice in online shopping. Using factor analysis we were able to map these factors into new sets of broader dimensions. The results indicate that the type of products sought affects the dimensions that are relevant for the Internet shopping task. Specifically, it appears that more dimensions of the Internet store are considered when consumers shop for computers relative to when they shop for books. A multinomial logit model that predicts participant choices based on their subjective ratings of the Web sites indicates that shopping for computers involves considerations of the shopping process, the information provided by the shop regarding its products, and perceptions of financial risk involved in the transaction.

In contrast, shopping for books appears to incorporate different considerations. Not only were the dimensions of risk and product information absent from participant considerations, but increased probability of purchase was associated with a less favorable shopping process. This result is counterintuitive as most current practices of site design emphasize the positive effect of the shopping experience as a major point of attracting and retaining shoppers. While there is no unqualified explanation for these findings, it may be the result of several context variables, which were unaccounted for by our study design. First, the experimental task was purchasing a book. This task is very specific in nature, with almost all possible variables associated with it known ahead of entering the shopping site (e.g., title, quality, cover). Thus the only attribute that distinguished one site from the other was the book's price. Consequently, respondents were probably most interested in this information since other information would provide no additional benefit. Any additional information, therefore, provided by the store would require respondents to spend more time examining the item which, under the current scenario, might lead to a negative effect on the probability of choosing a store. The other aspect worthy of consideration is the site characteristics that constructed the shopping process factor. When consumers have positive perceptions of items that relate to information provided in the site, product description, or the graphics of the site (all had high loadings on this factor), it also implies that the time spent in the site increases. Similar to the observation made earlier, under the experimental task (i.e., book purchasing) such a context might lead to an overall negative relationship between the shopping process and the choice probability of a specific site. Recent research suggest that time-related factors of Internet shopping might affects consumers (Rose and Straub, 2001; Rose, Lees and Meuter, 2001). The time spent in the site, therefore, should be taken into account in the design process. Likewise, our understanding of shopping behavior on the Web will gain from studying the time variable in this context.

Finally, we cannot rule out the possibility that idiosyncratic characteristics of the stores, which were not captured in the predefined set of store characteristics, did affect the choices of the participants in the book-buying task, but not in the computer-buying task.

These findings stress the importance of addressing the apparent heterogeneous nature of shopping tasks. This study demonstrated that book-shopping tasks were qualitatively different than computer-shopping tasks in terms of the shopping dimensions taken into account by the participants. These differences allude to the

possibility of a general difference between shopping for high-risk products and shopping for low-risk products. Since our experiment was designed to distinguish products along the risk dimension, we cannot tell whether these differences stem mainly from the price difference between the products or from the difference between search- and experience-quality products. Most likely, both aspects have contributed to the final result. Because of the characteristics of experience-quality goods, shoppers are more likely to pay close attention to the shopping process, the information about the product, and the financial risks involved. On the other hand, our results indicate that the shopping process might be a bit different for search-quality products at least in two aspects. First, in the book-shopping task store-specific characteristics seem to be more influential than common store characteristics. Second, the overall predictive power of the model was less accurate in the book-shopping task. There may be two reasons for why the model did not have high accuracy in predicting consumer behavior: (1) There might be additional common variables that were not included in our questionnaire and such exclusion eroded the model performance. 2) The strong effect of the unique variables (i.e., store specific) raises the question of whether the shopping process for such products is compensatory in nature. Both reasons might be plausible since the field is still young as is the existing knowledge on shopping for low-risk goods over the Internet and about the shopping process itself (i.e., whether it is compensatory or non-compensatory in nature) is rather limited. The latter view can lead to the idea that the modeling process might be either a non-compensatory or a combination of a non-compensatory (e.g., lexicographic) that reduces the choice set to an acceptable sites (i.e., reputation, etc.) followed by a compensatory process (see for example Gensch, 1987). Future research should clarify this issue.

6.1. Study Limitations and Future Research

Obviously, one should not ignore the limitations of this study. The use of students as a population for our experiment increases the internal validity of our results, but at the same time poses some external validity limitations (i.e., homogeneity of sample). This limitation is mitigated to some extent by the fact that our participants represent the population group (21-25 year olds) that is the most active in Internet shopping (e.g., Kehoe et al., 1999).

Another threat to the external validity of the study is the confinement of the study to respondents from one country (i.e., Israel). It would be helpful, thus, to replicate these results in different countries. Yet, it should be noted that an earlier study found similar patterns of behavior among online consumers from three different countries (Jarvenpaa and Tractinsky, 1999).

The choice of books and computers for this study's experimental tasks makes it impossible to separately examine the effects of price and product type (i.e., search vs. experience) on the shopping decision. Hence, future studies should isolate the effects of the two dimensions. The design of such studies might include search-quality products that are pricey as well as experience-quality products that are inexpensive to tease out the effects of each of these two dimensions.

In light of these limitations, further research in this area should include respondents from various countries shopping in sites with different languages as well as different cultural characteristics. Furthermore, investigating heterogeneity in consumer characteristics (e.g., risk aversion, heavy vs. light users) can shed more light on the choice among Internet shopping sites that will enable better segmentation of the market. Another extension can focus on tackling possible hierarchical purchasing decisions (i.e., compensatory vs. non-compensatory processes). This can possibly be addressed by adopting hybrid-modeling approaches that can accommodate both processes.

Our research was conducted on the B2C market. It is reasonable to assume that business markets would react differently to some of the factors identified in this study. For example, the shopping process might be of lesser importance for the business market relative to the consumer market. Investigating B2B market behavior, therefore, would improve the understanding managers have of how to attract potential customers to their shopping sites.

Other aspects of the competitive nature of Internet shopping for such products can be further explored by employing some perceptual mapping techniques that will enable site managers to better evaluate their competitive position. While our modeling approach enables managers to evaluate the relevant site characteristics that will contribute to increasing their market share, a mapping technique would reveal other facets of such competition in the form of distance (i.e., how similar the sites) and closeness to consumer ideal points.

REFERENCES

Einhorn, H.J. and Hogarth, R.M., "Behavioral Decision Theory: Processes of Judgment and Choice," *Annual Review of Psychology*, Vol. 32: 53-88, 1981.

Fishbein, M., "Attitude and Prediction of Behavior." In M. Fishbein, ed., *Readings in Attitude Theory and Measurement*. New York: McGraw-Hill, 1967.

Gensch, D. H., "Empirically *Testing* a Disaggregate Choice Model for Segments," *Journal of Marketing Research*, Vol. 25: 462-467, 1985.

Gensch, D. H. and S. Ghose, "Elimination by Dimensions," *Journal of Marketing Research*, Vol. 29, No. 4: 417-429, 1992.

Gensch, D. H. and W. W. Recker, "The Multinomial Multiattribute Logit Choice Model," *Journal of Marketing Research*, Vol. 16: 124-132, 1979.

Guadagni, P. and J.D.C. Little, "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, Vol. 2: 203-238, 1983.

Hair, J. F., R.E. Anderson, Ronald L. *Tatham*, and William C. Black, *Multivariate Analysis*, 5th ed., Prentice-Hall, 1998.

Hoffman, D.L., T.P. Novak, and P. *Chatterjee*, "Commercial Scenarios for the Web: Opportunities and Challenges," *Journal of Computer-Mediated Communication*, Vol. 1, No. 3, 1995. Available online: http://www.ascusc.org/jcmc/vol1/issue3/hoffman.html.

Howard, B. "The *Best* and Worst: Desktops," *PC Magazine*, *August 1,* 2001. Available online: http://www.pcmag.com/article/0,2997,s%253D1590%2526a%253D7559,00.asp,.

Jarvenpaa, S.L. and P.A. Todd, "*Consumer* Reactions to Electronic Shopping on the World Wide Web," *Journal of Electronic Commerce*, Vol. 1, No. 2: 59-88, 1997.

Jarvenpaa, S.L. and N. Tractinsky, "Consumer Trust in an Internet Store: A Cross-Cultural Validation," *Journal of Computer-Mediated Communication*, Vol. 5, No. 2, 1999. Available online: http://www.ascusc.org/jcmc/vol5/issue2/.

Jarvenpaa, S.L., N. Tractinsky, M. *Vitale*, "Consumer Trust in an Internet Store," *Information Technology and Management,* Vol. 1, No. 1-2: 45-71, 2000.

Kehoe, C., J. Pitkow, K. Sutton, G. Aggarwal, and J.D. Rogers, "Results of GVU's Tenth World Wide Web User Survey," 1999. Available online: (http://www.gvu.gatech.edu/user_surveys/survey-1998-10/).

Kehoe, C., J. Pitkow, and J.D. Rogers, "GVU's 9th WWW *User* Survey," 1998. Available online: http://www.gvu.gatech.edu/user_surveys/survey-1998-04/.

Kotler, P., *Marketing Management*, 10th Ed., Prentice Hall: NJ, 2000.

Li, H. Kuo, C. and Russell, M.G. "The Impact of Perceived Channel Utilities, Shopping Orientations, and Demographics on the Consumer's *Online* Buying Behavior," *Journal of Computer-Mediated Communication,* Vol. 5, No. 2, 1999. Available online: http://www.ascusc.org/jcmc/vol5/issue2/hairong.html.

Lohse, G.L. and P. Spiller, "*Internet* retail store design: How the user interface influences traffic and sales," *Journal of Computer-Mediated Communication*, Vol. 5, No. 2, 1999. Available online at: http://www.ascusc.org/jcmc/vol5/issue2/lohse.htm.

McFadden, D., Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (Ed.), *Frontiers in Econometrics*. New *York*: Academic Press, 1974.

McFadden, D., "The choice theory *approach* to market research," *Marketing Science*, Vol., 5, No. 4: 275-287, 1986.

Nayyar, P. R., "Information *Asymmetries*: A Source of Competitive Advantage", *Strategic Management Journal*, Vol. 1, No. 7: 513-529, 1990.

Nelson, P., "Advertising as *Information*," *The Journal of Political Economy*, Vol. 82, No. 4: 729-754, 1974.

Rose, G.M. and Straub, D., "The *Effect* of Download Time on Consumer Attitude Toward the E-Service Retailer," e-Service Journal, Vol. 1, No. 1: 55-76, 2001.

Rose, G.M., Lees, J. and Meuter, M. "A Refined View of Download Time Impacts on E-Consumer Attitudes and Patronage *Intentions* Towards E-Retailers," *The International Journal on Media Management* (forthcoming, 2001).

Swaminathan, V., E. Lepkowska-*White*, and B.P. Rao, "Browsers or Buyers in Cyberspace? An Investigation of Factors Influencing Electronic Exchange," *Journal of Computer-Mediated Communication*, Vol. 5, No. 2, 1999. Available online at: http://www.ascusc.org/jcmc/vol5/issue2/swaminathan.htm.

Vellido, A., , P.J.G. Lisboa, and K. Meehan, "Quantitative characterization and prediction of on-line purchasing behavior: A latent variable approach," *International Journal of Electronic Commerce*, Vol. 4, No. 4: 83-104, 2000.

APPENDIX 1

Items used in the study's questionnaire

1. The site belongs to a large store
2. The site belongs to a reputable store
3. Finding information in the site is easy
4. The Web site makes shopping easy
5. The site's design is beautiful
6. This site uses good mechanisms of data security

7.   This site provides good service
8.   The site's operators will not put my privacy in jeopardy
9.   The site provides complete information about the costs associated with a purchase
10.   I am interested in the products that are sold on this site
11.   The site has a wide range of products
12.   The site provides attractively priced merchandise
13.   Shopping on this site is fun
14.   The site's operators are reliable
15.   Shopping on this site might lead to monetary losses for me
16.   The site provides detailed description of the products
17.   The information on this site is reliable
18.   This site is a source of high-quality products
19.   Products on this site are displayed in a visually appealing way
20.   Products on this site do not meet my expectations
21.   The site provides up-to-date information to potential customers

APPENDIX 2

Rotated Component Matrix - Bookstores

	Factors and percent of explained variance			
	1	2	3	4
	41.69%	8.50%	7.56%	4.87%
The site belongs to a large store	.09	.18	.02	.91
The site belongs to a reputable store	.07	.11	.15	.91
Finding information in the site is easy	.75	.29	.08	.05
The Web site makes shopping easy	.69	.13	.28	.00
The site's design is beautiful	.72	.09	.01	.26
This site uses good mechanisms of data security	.34	.00	.68	.22
This site provides good service	.67	.35	.35	.10
The site's operators will not put my privacy in jeopardy	.19	.15	.76	.15
The site provides complete information about the costs associated with a purchase	.52	.26	.38	.02
I am interested in the products that are sold on this site	.35	.55	-.10	.14
The site has a wide range of products	.57	.59	.07	.06
The site provides attractively priced merchandise	.40	.56	.06	.13
Shopping on this site is fun	.79	.35	.12	.06
The site's operators are reliable	.25	.37	.62	.05
Shopping on this site might lead to monetary losses for me [R]	-.06	.09	.64	-.10
The site provides detailed description of the products	.72	.40	.18	-.06
Information on this site is reliable	.24	.62	.42	.08
This site is a source of high-quality products	.18	.68	.24	.16
Products on this site are displayed in a visually appealing way	.77	.23	.10	.07
Products on this site do not meet my expectations [R]	.18	.60	.22	.04
The site provides up-to-date information to potential customers	.60	.54	.20	-.01
Cronbach Alpha	.90	.71	.57	.76*

* Pearson correlation coefficient

Rotated Component Matrix - Computer Stores

	Factors and percent of explained variance					
	1	2	3	4	5	6
	36.82%	7.92%	6.14%	5.56%	5.04%	4.71%
The site belongs to a large store	.15	.14	.25	.85	.06	.02
The site belongs to a reputable store	.16	.17	.27	.81	.07	.06
Finding information in the site is easy	.58	.42	.07	.29	.20	-.05
The Web site makes shopping easy	.72	.13	.16	-.01	.10	.00
The site's design is beautiful	.82	.14	.11	.06	.11	.00
This site uses good mechanisms of data security	.16	.09	.69	.22	.06	-.04
This site provides good service	.57	.36	.40	.11	.08	.16
The site's operators will not put my privacy in jeopardy	.19	.00	.75	.18	.04	-.03
The site provides complete information about the costs associated with a purchase	.38	.61	.10	.11	.00	.02
I am interested in the products that are sold on this site	.09	.20	.14	-.03	.76	-.13
The site has a wide range of products	.30	.58	.05	.30	.33	-.04
The site provides attractively priced merchandise	.13	.68	.18	-.08	.15	-.04
Shopping on this site is fun	.74	.29	.20	.14	.10	.02
The site's operators are reliable	.14	.41	.67	.09	.12	.16
Shopping on this site might lead to monetary losses for me [R]	-.04	.07	.07	.02	-.07	.91
The site provides detailed description of the products	.38	.55	.01	.38	.06	-.05
Information on this site is reliable	.14	.61	.35	.15	.23	.21
This site is a source of high-quality products	.13	.31	.38	.20	.57	.02
Products on this site are displayed in a visually appealing way	.72	.28	.08	.24	-.01	-.04
Products on this site do not meet my expectations [R]	.20	-.11	-.31	.13	.56	.45
The site provides up-to-date information to potential customers	.41	.65	.00	.24	-.01	.06
Cronbach Alpha	.88	.83	.71	.76*	.46	NA

* Pearson correlation coefficient