19

Experience and Trust in Online Shopping

Robert J. Lunn and Michael W. Suman

Abstract

A key aspect of everyday life is consumer behavior. The Internet is increasingly being used as a tool for purchasing goods and services. In this chapter we focus on the factors influencing online shopping. We examine the role of a wide variety of variables organized around 14 conceptual themes. Using data from the 2001 UCLA Internet study, we analyze the relative importance of these different components in predicting Internet shopping frequency and dollar amounts spent. The results suggest the importance of experience as a major factor in predicting both amount and frequency of Internet shopping. Experience is conceptualized as a complex multi-dimensional construct consisting of indirect and direct components. These different components interact with the purchase process in different ways. We also discuss the advantages of a sequential modeling approach when attempting to understand predictors of amount and frequency of Internet purchases. Comparisons with other research findings are discussed. We conclude with a discussion of the importance of experience and trust as predictors of net shopping and suggest the implications of our findings for future research.

Introduction

The Internet is having a serious impact on the world's economy. Former US President Clinton asserted that one-third of all new economic growth that occurred during the eight years of his presidency was attributable to high-tech, with the Internet being a key component (Emeagwali, 2000). The value of e-commerce transactions is still small relative to the size of the American economy, and after a period of remarkable growth, the rate of expansion slowed considerably. However, even many traditionalists admit that the Internet's role in our consumer culture will continue to grow.

Beyond the dollar amount of e-commerce transactions, even more significant are the ways in which businesses and consumer activities are being changed. New business models are being created in an effort to lower business costs, improve customer service, and increase productivity and profitability (US Department of Commerce, 1999).

Both Internet-based companies and some traditional producers of goods and services are transforming their ways of conducting business. And as for the customers, they can now shop 24 hours a day, 365 days a year. They are buying computer hardware, software, books, CDs, airline tickets, flowers, automobiles, and a host of other products. They are making hotel reservations, trading securities, checking their 401 K accounts, and conducting a wide variety of other business transactions.

Consumer behavior is one key aspect of everyday life affected by the Internet and that is what we investigate in this chapter. In particular, we examine the factors influencing online purchasing behavior. Do consumers shop on the Internet because they perceive the prices there to be lower than in brick-and-mortar establishments? Do they turn to the net to meet their shopping needs because of convenience, as a way to save time in their busy lives? As Internet experience grows, do consumers buy more products and services using the Internet? Does experience with other means of remote shopping, such as phone and mail purchasing, predispose the consumer to buy online? Are Internet users deterred from online purchasing by their concerns over the security and privacy of their personal and credit card information? Do they opt out because of inferior customer service or because they would miss the social aspect of the shopping experience that they get at the local mall? Are people put off because it is difficult to assess the quality of products or the accuracy of online product descriptions? To what extent do demographic variables predict whether one will be an Internet shopper or not? Do the young shop more than the old, women more than men, the well-off more than those who are less so? Are those with faster Internet connections more likely to buy online than those with slower telephone modem connections?

Previous research

To date there has not been much empirical research on predictors of online buying behavior. There has been conceptual work on factors

influencing online shopping (Alba, Lynch, Weitz, Janiszewski, Lutz, Sawyer, and Wood, 1997; Palmer, 1997). And there have been efforts to classify different types of online shopper (BMRB International, 1999; Harris Interactive, 2000). Lohse, Bellman, and Johnson (1999, 2000) have identified predictors of online buying behavior using panel data from the Wharton Virtual Test Market survey. Based on their first survey panel (WVTM1), Lohse et al. reported that people who spent more money in e-commerce had a more "wired lifestyle," were on the net more hours per week, and received more email than other Internet users. In their second study conducted one year later (Lohse et al., 2000, using WVTM2), they associated net purchasing with less concern about online privacy, more years of online experience, more email messages received per day, more purchasing from catalogs, being male, and more frequent use of the net to search for product information, travel information, financial information, current events information, and news. They identified certain variables that, though they did not predict whether one shopped online, did influence how much one spent there. These factors positively associated with online spending included income, likelihood of downloading software, number of hours worked per week, and hours per week online.

Swaminathan, Lepkowska-White, and Rao (2000) have investigated factors influencing electronic exchange using secondary data based on an email survey. They found that perceived vendor reliability, convenience of placing orders and contacting vendors, price competitiveness, and access to information had a positive influence on the number of online purchases, but did not influence the amount of money spent on the Internet. Their results showed that among females social interaction served as a shopping motivation, and that loss of social interaction deterred female consumers from frequent online shopping. Consumers who valued convenience tended to use the Internet to purchase more frequently, and they seemed to spend more money in their electronic transactions. These researchers conducted a stepwise regression to assess the relative contribution of their different independent variables. The results indicated, with both frequency of shopping and amount spent online as dependent variables, that customer characteristics, such as convenience as a shopping motive, dominated all other variables in terms of variance explained. One problem with this approach is that stepwise regression results are notorious with respect to replication problems.

Data

After a review of the literature, we selected a wide range of variables that have been implicated in predicting purchase frequency and purchase amounts on the Internet.

The source of our data is the 2001 UCLA Internet Study. In this study interviews were conducted with 2,006 households throughout the 50 states and the District of Columbia.

Sample

This was the second year of an American longitudinal dynamic panel study. Out of 2,096 people in the original panel, 1,274 answered the second year's survey. For both the original sample drawn last year and the replacement sample selected this year a national Random Digit Dial (RDD) telephone sample using an Equal Probability Selection Method (EPSEM) was used. This sampling methodology gives every telephone number in the 50 states and the District of Columbia an equal chance of being selected. In the initial call an interviewer spoke to a person in the household 18 years of age or older to obtain a roster of all household members. At this point, a computer system ("CFMC Servent" CATI) randomly selected one individual from among those 12 years of age and over in the household to be the interviewee from that household. If the randomly selected individual was between 12 and 17 years of age, the interviewer asked a parent or guardian for permission to interview the child. In the initial contact, once the selection of a household member was made, only that individual was eligible to complete the interview. Eight call attempts were made to complete an interview. If a household refused twice, it was not contacted again. In re-contacting panel members from the original sample up to 16 call attempts were made to reach them. The same household member who participated the previous year was interviewed again. The only condition in which a new household member was accepted was if the one interviewed the previous year was no longer a member of the household. Those participating in the survey for the second year were paid a monetary incentive. Interviews were conducted in English and Spanish. Interviewing took place between May and July 2001.

After the data were collected they were compared to census data to ensure that the sample was representative in terms of geographic distribution, race, age, sex, family composition, education, and household

income. The sample was very close on all demographic categories except for minor discrepancies on education, gender, and race. To correct for this the data were weighted by these three factors. Sample size was preserved during the weighting process. Since some of the questions were only asked of adults, the analytical sample consisted of only adult Internet users (18 years of age or older). The analytical sample size was 1,173.[1]

Analysis

Overview of the data analysis section

The data analysis consisted of three phases: data conditioning, data transformation, and data modeling. Data conditioning was performed to remove extreme values and handle missing data. Data transformation in this circumstance refers to the principal components transformation of the set of potential regression predictor, that is, independent, variables to statistical independence. The data modeling phase involved two multivariate data analytic techniques: regression and discriminant analysis. Regression analysis provides information on what variables are reliable with respect to predicting the dependent variables of interest, that is, average purchase amount and average purchase frequency, and how important each reliable variable is in terms of predictive ability. It also provides information on the quality, that is, how well they predict, of the predictive equations. Discriminant analysis is a statistical classification procedure. Our usage of discriminant analysis in this study was confined to how well the reliable variables found in our regression analysis could predict respondents who purchase versus those who don't purchase on the Internet. The following sections describe each of the data analysis phases in more detail.

Data conditioning

The variables selected for inclusion in the analysis were examined for outliers. The presence of outliers was small. The outlier values

1 For more information on the survey and to see the questionnaire used in the study go to UCLA Center for Communication Policy's website at http://ccp.ucla.edu

were winsorized (outlier values brought in to two standard deviations from the mean). The proportion of missing data was also small. Missing data were replaced with random imputation around each variable's modal value. Missing data were not imputed for the dependent variables (dollar amount of Internet purchases and frequency of Internet purchase). If data were missing for the dependent variables, the case was dropped. The proportion of cases with missing data on the dependent variables was less than 3 percent. No pattern signifying that cases with missing data were different than cases with data was apparent.

Controlling for redundancy and conceptual confounding

The primary purpose of the research effort was the development of predictive models to help identify characteristics associated with Internet purchase behavior. Specifically, we wanted to understand what factors predict purchase frequency, and what factors predict purchase amount. One complication faced by all model builders concerns the issue of multicollinearity. Multicollinearity refers to the presence of correlations among a set of predictor variables. The presence of multicollinearity causes interpretative difficulties with respect to the size, stability, and computational accuracy of a regression model's predicative weights. There are several approaches to handling multicollinearity. We chose to apply a principal components transformation to the set of predictor variables. The following provides a brief introduction to the concept of principal components analysis.

A principal components analysis determines how a set of variables are interrelated, that is, correlated, and then demonstrates how the correlated variables could be combined into a smaller number of statistically independent "factors" that represent most of the information in the original set of variables.

A brief example will clarify the underlying concept. In our situation, people who rate themselves as highly experienced Internet users spend more time connected to the Internet, send and receive more email, and have been using the Internet more years than people who rate themselves as inexperienced Internet users. Since all these concepts are inter-correlated, we can substantially simplify our analytical task, and remove multicollinearity, by combining items that are cor-

related into one unified information theme. So, in our example, we can combine the four different "experience-oriented" variables into one underlying informational theme, or factor, called "Internet experience." Thus, a principal components analysis combines similar informational items by removing redundancy.

A second benefit associated with a principal components analysis is the removal of certain types of conceptual confounding. A good example of this concerns the pattern of associations between predictor variables like income, online experience, and age with a dependent variable such as average purchase amount. Income, online experience, and age are each individually correlated with average purchase amount. An interpretative confound can occur because income, online experience, and age also exhibit a pattern of correlations with each other. However, the nature of those correlations is not sufficient to consider them as measuring a unified informational theme. In this situation, the principal components analysis recognizes there are three different underlying informational themes, that is, income, online experience, and age, and transforms them to separate, statistically independent factors. Statistical independence means the resultant factors are not inter-correlated.

The end result is a set of predictor variables, that is, factors that have no correlation with each other.

Data modeling

Regression analyses were performed in two stages. The first stage was used to identify significant predictor variables. The second stage removed the non-significant predictor variables from the resultant predictive equations. This allows for a better estimate of the explanatory power of the regression equations. The predictor variables were orthogonalized factors. Because the predictor variables were orthogonal, that is, zero correlated, there was no multicollinearity present. This means the variance figures are strictly additive for higher-level conceptual interpretations.

A discriminant analysis was used to determine the ability of the significant predictor variables in the regression analysis to predict membership in groups representing the zero/one state for our dependent variables. The zero/one state for purchase amount and purchase fre-

quency indicated those who purchased on the Internet and those who did not.[2]

Results

In this section we provide information on the results of the principal components analysis and statistical details concerning how well our different informational themes predicted online purchase frequency (model 1) and online purchase amount (model 2).

Indication of a variable's predictive strength

In a regression analysis some of the predictor variables are better predictors than others. A regression analysis provides "weights" that allow us to assess the *relative* predictive strength for each variable. One of these weights is called a "beta" weight. You would normally see a table of these weights when describing regression results.

However, one of the bonuses involved with using a principal components analysis is that the predictor variables have been transformed to independence. That substantially simplifies our reporting task. When you are dealing with independent predictor variables the sum of the squared beta weights is exactly equal to R^2.

Our results are presented in the regression table, table 19.1, which shows the percent of predicted variation in each model attributable to each of the significant predictor variables. For example, when predicting purchase amount (model 2), the average amount you spend per year making mail and phone purchases, excluding Internet purchases, accounts for 56.83 percent of all the accounted for variation in the dependent variable, that is, average amount of online purchases. With these independent predictor variables we have obtained the *absolute* predictive strength for each significant predictor variable.

2 The discriminant analysis classification results were obtained using a 50 percent random hold out sample. This allowed us to obtain an unbiased estimate of the classification accuracy for each set of variables. We ran the discriminant analyses to obtain classification accuracy figures to compare with the results of Lohse et al. (1999, 2000). Lohse et al.'s models demonstrated that it was possible to achieve a high degree of classification accuracy when attempting to distinguish those who purchase from those who do not purchase on the Internet.

Informational themes

The following 14 informational themes were identified by our principal components analysis: perception of the Internet, Internet experience, privacy/security concerns, negative consequences of shopping on the Internet, perceived availability of goods and services on the Internet, connection speed, age, increased likelihood of buying brand names, perceived price advantage of shopping on the Internet, structural Internet shopping difficulties, income, shop on Internet / buy elsewhere, previous purchase by mail and phone, gender.

The following describes each of the identified informational themes. We also include a brief overview of how similar informational themes have been found to predict average purchase amount and average purchase frequency in past research.

1 Perception of the Internet

This component includes one's evaluation of the Internet overall, perceived ease of finding information on Internet, perceived amount of relevant information available on Internet, how much of the information on Internet one believes is reliable and accurate, and the extent to which one believes that using the Internet saves time. These variables involve the extent to which one has positive perceptions of and experiences with the Internet. These types of positive feelings might be tied to higher levels of online buying.

Some of these concepts have been previously explored. For example, there is evidence suggesting the importance of perceived reliability of an exchange partner on purchase behavior (Morgan and Hunt, 1994; Luedi, 1997). Swaminathan et al. (2000) also connect online purchasing frequency with vendor reliability. And Moorman, Deshpande, and Zaltman (1993) relate the concept of vendor reliability with trust. We expect trust-oriented perceptions to positively correlate with Internet purchasing behaviors.

2 Internet experience

This component includes self-rated ability to use the Internet, total hours of Internet connect time per week, use of email, and number of months of Internet experience. There is evidence that these experience variables are associated with Internet purchase amounts and fre-

quency (Lohse et al., 1999, 2000). Experience with an activity is often tied to increased levels of that and related activities. We would expect these variables to positively correlate with Internet purchasing variables.

3 Privacy/security concerns

This component includes concern about the privacy of personal information when purchasing on the Internet, concern about the security of one's credit card when purchasing on the Internet, concern that new technology will lead to loss of personal privacy, and the perception that people who go online put privacy at risk. Some research suggests that consumers are not very concerned about privacy (Milne and Gordon, 1993). Other studies suggest that they are (Bloom, Milne, and Adler, 1994; Rohm and Milne, 1999). There is also evidence that perceived lack of security is a major deterrent to online purchasing (Zellweger, 1997).

The concept of trust is obviously directly related to these concerns. Many facets of the issues of security, privacy, and trust have been extensively reviewed (Camp, 2000; Cranor, 1999; Feldman, 2000; Fox, 2000; Friedman and Kahn, 2000; Gefen, 2000; Grossman, 2000; Lohse, Bellman, and Johnson, 2000; Minahan, 1997; Olson and Olson, 2000; Sabo, 1997; Salnoske, 1998; Schoder and Yin, 2000; Sklar, 2001; Urban and Sultan, 2000; Wachter, 1999). According to Swaminathan et al. (2000) online consumers are only marginally less concerned about the security of electronic exchanges than non-consumers. But they argue that online consumers are more concerned about some aspects of information privacy. For example, consumers who purchased more on the Internet seem to be more favorable about the creation of laws protecting privacy on the Internet. And they state that another dimension of privacy, that is, consumers' beliefs that marketers want information about them for marketing purposes, has a marginally negative effect on the amount of money spent on the Internet. Despite some discrepancies in previous research, we would expect these variables to be somewhat negatively correlated with the net purchasing variables.

4 Negative consequences of shopping on the Internet

This component includes being uncomfortable with the lack of face-to-face contact when ordering on the Internet, stated concern that one would miss the company of fellow shoppers when buying online, perceived difficulty in assessing product quality and accuracy of product

descriptions, perceived difficulty in returning or exchanging goods ordered on Internet, and the inverse of an intention to make many more purchases online. Some of these factors deal with the concept of convenience. There is evidence that perceived convenience is an important factor encouraging remote (from the home) shopping (Gehrt, Yale, and Lawson, 1996; Gillet, 1976; Reynolds, 1974). In addition, Swaminathan et al. (2000) provide evidence that convenience is powerfully associated with purchase amount and frequency. Research also suggests that the Internet is less attractive to consumers who value the social aspect of shopping (Alba et al., 1997). This finding was confirmed by Swaminathan et al. (2000).

One factor, the negative aspects of Internet shopping, includes a mix of two concepts: convenience and the loss of social interaction. The perceived inability of consumers to assess product quality and accuracy of product descriptions, and difficulties in returning or exchanging goods and services ordered on the Internet are inverse aspects of shopping convenience. The discomfort with lack of face-to-face contact and missing the company of fellow shoppers involve the loss of social interactions in the shopping and transaction process. As these are by their nature structural problems with shopping online, we would expect these variables to negatively correlate with the net purchasing variables.

5 Perceived availability of goods and services on the Internet

Those who perceive a ready availability of goods and services on the Internet are probably more likely to buy them. Therefore, we would expect this variable to positively correlate with the Internet purchasing variables.

6 Connection speed

The faster one's connection the easier it is to use the Internet, and by extension, the easier it is to buy online. This relates to the previously discussed concept of convenience. We would expect high-speed connection to positively correlate with the net purchasing variables.

7 Age

Some studies suggest that demographic variables influence Internet usage patterns (Hoffman, Kalsbeek, and Novak, 1999; UCLA, 2000).

Others relate demographic variables more specifically to shopping motivations and behaviors (Bellenger and Kargaonkar, 1980). Since younger elements of the population are more likely to be on the Internet, we would expect this segment to purchase more frequently online. However, older people have more money than younger people, and people with more money might be expected to spend more online and purchase more frequently. It is difficult to provide a uniform prediction of the net effect of age.

8 Increased likelihood of buying brand name products while shopping on the Internet as opposed to in brick-and-mortar stores

This speaks to a certain level of fear and lack of trust with online vendors. Since the consumer does not know if he or she can trust the online vendor, he or she turns to something that has proved to be historically reliable in their experience, and substitutes brand reliability for vendor trust. We would expect this variable to be correlated with net purchasing variables, although we would expect its role to be a negative one. If you fear something, it is a negative factor. It could be said that people buy brand names on the Internet because they are less expensive on the Internet. We would argue that one of the hallmarks of strong branding is pricing consistency. Many brand name manufacturers do not allow sale pricing of their merchandise because sale pricing detracts from the premium status of the brand.

9 Perceived price advantage to shopping on the Internet

There is evidence that the Internet supplies information that makes price comparison possible and that this information is likely to increase purchases (Zellweger, 1997). Swaminathan et al. (2000) have demonstrated a relationship between price competitiveness and online purchase frequency. We would expect this variable to positively correlate with the net purchasing variables.

10 Structural Internet shopping difficulties

This factor includes the stated position that if sales tax were charged one would probably buy less on the Internet and the belief that shipping charges for online products are too high. The first statement is a negative form of price influence for Internet transactions. The second statement is another negative aspect of Internet price influence, ship-

ping charges. Factors 9 and 10 appear to measure similar concepts of price sensitivity. However, factor 9 is more of an overall impression of price on the Internet versus price elsewhere while factor 10 is more structurally oriented. Factor 10 involves underlying processes which might account for why overall costs on the Internet could be higher than elsewhere. We would expect these variables to negatively correlate with the net purchasing variables.

11 Income

Higher income levels are typically correlated with higher levels of buying and we would expect the same association here.

12 Shop on Internet/buy elsewhere

This factor measures whether a person seeks information about products and services on the Internet and then makes the purchase transaction in a local store. This is a strategy that relegates Internet sites to a mostly information provision status. Since beginning Internet users mostly use the Internet in an information-seeking fashion (UCLA, 2000; UCLA, 2001) we would expect a higher frequency of novice Internet users to adopt this strategy. Since Internet experience is likely to be positively associated with Internet purchase behavior, we would expect this variable to be negatively correlated with the online purchasing variables.

13 Previous purchase by mail and phone

These older forms of from-home shopping may serve as precursors to Internet shopping. This may occur through the experience of establishing a level of comfort and trust with a remote vendor. There is evidence associating online buying with catalog purchasing (Lohse et al., 2000). We would expect buying from mail and phone to be positively correlated with buying online.

14 Gender

Some studies have examined the effect of gender on the process and outcomes of the use of the Internet (Wachter, 1999). Swaminathan et al. (2000) give evidence that males are less motivated than females by social interaction as a shopping motive and are more convenience-

oriented, both of which are associated with increased Internet shopping. Lohse et al. (2000) have also linked being male with increased online buying. We would expect being male to be positively correlated with buying online.

Description of the predictive models

In this section we provide a technical overview of the regression and discriminant analysis procedures. Model 1 describes the results when we are attempting to predict online purchase frequency. Model 2 describes the results when attempting to predict average online purchase amount. The explanatory power for both models is similar.

In the first regression model (model 1) we are trying to predict results for the following question: "During a typical year, how many times do you purchase products or services over the Internet? Do not include payments for your Internet connection." Ten of our 14 informational themes were significant predictors of online purchase frequency. Overall the model predicted 26.3 percent of the total variation in average yearly number of times purchased on the Internet ($F(10,1162) = 34.4$, $p < 0.000$). The shrunken R^2 was 25.50 percent.

In the second regression model we are trying to predict average purchase amount. The actual question was "During a typical year, how much do you spend ordering products or services over the Internet? Please do not include any money that you paid to access the Internet." Seven of our 14 informational themes were significant predictors of average purchase amount. Overall this model predicted 23.9 percent of the total variation in the average yearly amount the Internet user purchased on the Internet ($F(7,1165) = 36.5$, $p < 0.000$). The shrunken R^2 was 23.3 percent.

Table 19.1 provides an overview of the predictive results across both models. The second column of table 19.1 provides the percent importance in terms of predicting average online purchase frequency by the different variables. Variables that do not appear in column one were not reliable predictors. Predictive importance figures in parentheses indicate a negative association. For example, high scores on the factor "negative consequences of shopping on the Internet" reduces average online purchase frequency (see column two, row three). The third column ranks the importance of each significant predictor in terms of predicting average online purchase frequency (the table is sorted by model 1 results). The fourth and fifth columns show the results for

Table 19.1 Overview of predictive models

Factor name	*Model 1 (purchase frequency) % Importance*	*Model 1 rank importance*	*Model 2 ($ amount) % Importance*	*Model 2 rank importance*
Purchase mail/phone	41.59	1	56.83	1
Negative consequences of shopping Internet	(24.78)	2	(11.40)	3
Income	7.69	3	7.21	4
Internet access experience	6.32	4	2.02	7
Privacy/security concerns	(5.77)	5	(3.49)	5
High-speed Internet/home	5.54	6	15.71	2
More likely buy brand products on Internet	4.23	7	NS	—
Availability of good and services	1.45	8	NS	—
Gender (male)	1.32	9	3.34	6
Internet prices lower	1.31	10	NS	—

model 2. The table was designed to facilitate comparisons of differences across the two models.

Classification accuracy

The purpose of the discriminant analyses was to determine how well our reliable predictors variables from each model, models 1 and 2, could classify people who purchase from people who don't purchase on the Internet. This test was performed so classification accuracy could be compared with results from other researchers (Lohse et al. 1999, 2000). The two discriminant analyses were run using a random hold out sample (50 percent). This is done so that resultant classifica-

tion accuracy is based on a set of randomly selected data (50 percent of the sample) that were *not* used to construct the classification equations. This removes a known source of bias from the classification results. The classification accuracy for the variables used in the regression model for Internet purchase frequency was 77 percent. The classification accuracy for the variables used in the regression model for Internet purchase amount was 73.7 percent.

These tell us that the set of significant predictor variables for each model can accurately determine respondents who purchase on the Internet from respondents who don't purchase on the Internet with an accuracy of 77.0 and 73.7 percent respectively.

Interpretation of Results

The results of the two models predicting different measures of Internet purchasing are relatively similar. This is not surprising since the correlation between dollar amount of purchases and frequency of purchases on the Internet is 0.59 ($p < 0.000$). Thus, some degree of similarity in the results of the two models is to be expected. Both models provide similar levels of predictive power. Model 1 accounts for 25.5 percent of the total variation in purchase frequency, and model 2 accounts for 23.3 percent of the total variation in purchase amount.

Experience and trust

The most important predictor of Internet annual purchase amounts and frequency of purchases is the amount of money an individual spends by mail or phone.

Previous research has demonstrated that another type of experience, that is, experience with the Internet itself, has a strong association with Internet purchase frequency and Internet purchase amount (Lohse et al., 1999, 2000; UCLA, 2000, 2001).

These findings imply that the degree of experience with non-Internet, remote shopping modalities, that is, a form of indirect experience, might substitute for direct Internet experience, that is, years of Internet experience and/or average connect time. Previous research has indicated a relatively long latency period between an individual's first usage of the Internet and his/her first Internet purchase (UCLA, 2000, 2001). If purchase experience by catalog or phone substitutes for

Internet experience, we would expect a shorter time period (that is, a shorter latency period) between an individual's first Internet access and his/her first Internet purchase for those who have phone or mail purchasing experience. This is exactly the pattern of results observed. Adult Internet users who shop by catalog or phone, on average, take 18.1 months before they make their first Internet purchase. Adult Internet users who do not shop by catalog or phone take 25.0 months. This is a reliable difference ($t = 3.10$, $p < 0.001$).

One measure of Internet experience is the individual's average connect time per week. If non-Internet remote shopping substitutes for Internet experience, then we should find lower average connect times for individuals who purchase on the Internet with alternative remote shopping experience versus individuals who purchase on the Internet with no alternative remote shopping experience. Once again, this is the pattern of results observed. Adult Internet users who do not shop by catalog or phone but do shop on the Internet have an average weekly connect time of 13.7 hours. Adult Internet users who do shop by catalog or phone and shop on the Internet have an average weekly connect time of 10.9 hours. This is a reliable difference ($t = 2.03$, $p < 0.04$).

The concepts of trust and experience are often found to be strongly associated (Gefen, 2000). In a way, trust in the retail environment is built through repeated successful transactions. In the absence of actual experience with a retailer, vendor reputation, brand reputation, or word of mouth serve as surrogates for actual experience. We now know that experience with a similar shopping modality also substitutes for experience. Without experience, the first purchase that someone makes on the Internet represents a leap of faith. The larger the dollar amount of that first purchase, the greater the risk. The greater the risk, the greater the leap of faith. Reduction of perceived risk levels has been identified as a key antecedent to relationship commitment in general (Sheth and Parvatiyar, 1995) and consumer behavior in particular (Bauer, 1960). Although we don't specifically measure a "trust" factor, the concept of trust is very likely strongly associated with how often and how much one buys in any purchasing experience.

The presence of a high-speed Internet connection in the home is another factor that is strongly influenced by experience. Only 6.7 percent of users with less than one year of connection experience have high-speed access (i.e., cable modem, DSL, ISDN, T1/T3). High-speed access is present for 13.2 percent of individuals with two to less than

four years of experience, and 22.5 percent of individuals with five or more years of experience. The UCLA studies have also demonstrated that the ways in which the Internet is used change as a function of Internet experience. Those newly online primarily use the Internet in chat rooms, for general browsing, and for games and entertainment. More experienced Internet users use the Internet more for work, checking their finances, trading stocks, and shopping (UCLA, 2000, 2001). As people gain experience, they learn how to use the Internet to address and satisfy more practically oriented needs and concerns. And correspondingly, as the Internet becomes more useful, an individual is more likely to commit to a high-speed connection.

We therefore argue that at least three of the factors in our models appear to represent different aspects of a multi-dimensional construct of trust/experience. Since the different factors are independent, their predictive contributions are additive. Using this interpretive approach, trust/experience related factors account for 74.56 percent of the predicted variation in dollar amount of purchases. This compares with only 53.45 percent predictive power in model 1 where we are predicting purchase frequency. Thus, trust and experience factors account for greater predictive power when we are modeling purchase dollar amount as opposed to purchase frequency.

Major barriers to Internet shopping

The predictive power of negative consequences of shopping on the Internet are substantially stronger when we are predicting purchase frequency (24.78 percent), as opposed to when predicting purchase amount (11.40 percent). Negative consequences of shopping on the Internet include variables such as being uncomfortable with the lack of face-to-face contact when ordering on the Internet, perceived difficulty in assessing product quality and accuracy of product descriptions, and perceived difficulty in returning or exchanging goods ordered on the Internet.

The higher predictive power for frequency of purchases suggests an underlying cumulative effect. The more times you purchase, the more likely something is going to go wrong. Or the more times you purchase, the higher the cumulative shipping charges. And the more times you purchase, the higher the likelihood of having to return merchandise, and the greater the hassle involved.

Income: the more you make, the more you spend

Both models suggest that individuals with higher incomes tend to purchase more frequently and spend more money on the Internet. This finding adds face validity to the results since most retail studies demonstrate that higher incomes are associated with more purchases.

Income and age are confounded to some degree. The principal components analysis statistically transformed income and age into distinct factors. Thus, while these two variables may be naturally interrelated, after the principal components analysis, they are transformed to statistical independence. The advantages of this approach are clear in the resultant modeling efforts. The portion of variation due to age alone was not a reliable predictor of online buying behavior, while the proportion of variation due to income was.

These results suggest that any univariate association, that is, correlation, of age with online buying behavior is either spurious or represents an indirect or mediated effect. In other words, just because you are young does not mean you buy more goods and services online. By splitting income from age, we can see that it is most likely that income, and not age, is directly associated with buying behavior.

Privacy and security

Privacy and security concerns appear in both models, but their effects are relatively small. Privacy and security concerns are measured in this study on a five-point scale ranging from "extremely concerned" to "not at all concerned." Let us first look at Internet users with mail or phone shopping experience. For members of this group who have also purchased on the Internet, the mean score on concern for the security of their credit card information on the Internet is 2.74. The mean score on this measure for Internet users with no Internet purchase is 3.40. Among individuals who have made a mail or phone purchase, the Internet purchasers have reliably lower levels of concern than individuals who have not purchased on the Internet ($t = 10.5$, $p < 0.000$).

Among Internet users with Internet shopping experience but no mail or phone shopping experience, the average rating on concern about the security of their credit card information on the Internet is 2.56. For those who have made no remote purchases (mail, phone, or

Internet) the mean score is 3.31. Once again, Internet users who have made an Internet purchase express reliably lower levels of concern than individuals who have not made an Internet purchase ($t = 7.24$, $p < 0.000$).

This same pattern of results extends to concerns about the privacy of personal information. The average ratings of privacy concern among mail and phone shoppers purchasing on the Internet and not purchasing on the Internet are 2.69 and 3.23 respectively ($t = 8.75$, $p < 0.000$). The average ratings of privacy concern among non-mail and phone shoppers purchasing on the Internet and not purchasing on the Internet are 2.59 and 3.12 respectively ($t = 5.10$, $p < 0.000$).

We controlled these mean comparisons by mail and phone shopping experience to demonstrate an important finding. While previous shopping experience using mail or phone shopping modalities appears to substitute for Internet experience when considering impact on purchase frequency and amount, this effect does not extend to security and privacy concerns. These concerns appear to be mostly related to the Internet shopping modality itself.

The role of brand, product availability, price

The Internet purchase frequency model (model 1) includes three factors that do not appear in the Internet dollar amount purchase model (model 2). In modeling frequency of Internet purchases, more frequent purchases are accompanied by a higher likelihood to purchase brand-name products on the Internet, a greater perceived availability of goods and services on the Internet, and a greater perceived price advantage to shopping on the Internet.

The likelihood of those with mail and/or phone experience who are Internet shoppers to buy brand name products is higher than the likelihood of remote shoppers who only purchase through mail and/or phone. The mean score for those who purchase on the Internet is 2.93, and for those who do not the mean is 2.58 ($t = 3.67$, $p < 0.000$). This same pattern is not replicated among people without mail and phone purchase experience. In this case, the mean score for individuals who purchase on the Internet on the same statement is 2.69, while for those who do not purchase on the Internet the mean is 2.51 ($t = 1.15$, NS (not significant)). Apparently experiences drawn from mail and phone purchases predisposes one to buy brand name products. Thus, remote shopping experiences have "taught" the customer that buying brand

names somehow reduces the inherent risk associated with not touching and feeling the goods.

The finding that perceived price advantage is not an important predictor of Internet purchasing behaviors might seem puzzling given the attention paid to price by many e-retailers. Examining perceived lower prices by our familiar mail/phone/Internet groupings revealed no reliable differences by any combination of these variables. Perceived price advantage does not predict purchase amounts, and is only a minor predictor for Internet purchase frequency. This finding could reflect a cost versus convenience trade-off, or Internet sites could be claiming lower prices on an item-by-item basis, but when shipping charges are added to the bottom line, prices may be seen as equivalent. Alternatively, selection and customer service often substitute for low price. In any event, price is not a major predictor of Internet purchasing behavior, and Internet sites that concentrate on trying to convince consumers that they have lower prices than local retail stores are likely to encounter a good deal of skepticism. This is particularly true when dealing with traditional (brick-and-mortar, mail and phone) retail operations that also have an online presence. If vendors selling through traditional retail operations were to offer lower prices for their online shopping outlets, they would undercut their own traditional retail outlets.

The availability of goods and services is a minor predictor of purchase frequency, accounting for only 1.45 percent of the predicted variation. There is a reliable mean difference ($t = 7.06$, $p < 0.000$) between people who purchase on the Internet (4.01) and those who don't purchase on the Internet (3.66) on perceived product availability. Apparently, this difference is irrelevant when attempting to predict the dollar amount of purchases, and is only of minor relevance when attempting to predict frequency of purchase.

The role of gender

Gender is a reliable minor predictor of purchase frequency (1.32 percent) and purchase amount (3.34 percent). Interestingly, males, on average, spend nearly three times as much on the Internet as females ($899 per year versus $311, $t = 4.25$, $p < 0.000$). Males also purchase more frequently than females (7.14 times per year versus 4.39, $t = 3.46$, $p < 0.001$). In the aggregate, the average purchase for males is $126 per year versus $71 for females.

It is instructive that our dependent variables exhibit such large mean differences by gender, yet gender itself is a minor predictor in our regression equations. One likely explanation is that gender is confounded with other predictor variables, specifically experience.

Among people who shop on the Internet, the average experience in months for males is 51.9, the average for females is 44.6 ($t = 3.21$, $p < 0.001$). In addition, males spend nearly twice as much money as females using the mail/phone order remote shopping modality, excluding Internet purchases ($758 per year versus $402, $t = 2.22$, $p < 0.02$). Thus, males have more direct (time online) and more indirect (spend more on male/phone orders) experience than females. When variation due to experience is factored out of gender, the resultant gender specific variation is a poor predictor of online shopping behavior. Another reason for this poor performance is that gender is a two-value variable, that is, male or female. A two-value variable is never going to do a good job at predicting a continuous range variable like purchase amount. Although we are not attempting a causal analysis in this chapter, it does seem logical that the more (favorable) experience you have with remote shopping, the more likely you will be to increase your remote shopping. Gender could never have that type of effect since gender does not vary beyond two values. A gender effect would have to work in an indirect fashion, influencing other variables that exhibit a range of values.

A note on location

"Location" is certainly a critical factor for real-world shopping facilities. However, it would not have a one-for-one correspondence with a similarly named concept in cyberspace. For example, distance and ease of access would probably define location for real-world shopping facilities. However, distance is irrelevant in cyberspace. We suspect a higher-level concept like convenience would encompass location in its definition, but the sub-components that define convenience in the two shopping domains would very likely be different. It would be interesting to compare models for brick-and-mortar buying behavior with different modalities for remote shopping behaviors, but that would very likely involve designing a study just for that purpose. The UCLA study contains a wealth of information about remote shopping behavior, but it has no information on brick-and-mortar shopping behavior. Thus it offers no information on location as a variable.

Classification results

We used two levels of modeling in our analysis. The regression models were used in an attempt to predict the "amount" of the dependent variables. The discriminant analysis models were used to determine accurate classification rates for a yes/no predictive model. (We could have also used a logistic regression approach in the latter case, but discriminant analysis is adequate for the task.)

The classification results were relatively strong. Each of the classification models correctly predicted over three-quarters of the modeled behaviors.

Summary and Conclusions

The single most important factor predicting Internet shopping behavior is experience. Experience is actually a complex multi-dimensional construct that consists of direct and indirect components. Direct experience refers to aspects of Internet utilization, for example, the amount of average Internet connect time per week or the presence of a high-speed Internet connection. In this study, indirect experience, such as the amount of money a person spends purchasing by mail or phone, was found to substitute for direct Internet experience such as average Internet connect time. This has important theoretical and methodological implications. For example, when predicting Internet purchase amounts or frequency using a direct measure of Internet utilization such as Internet average connect time, you will get different results depending on the amount of mail or phone shopping experience present in the sample.

Our findings with respect to experience generally support those of Lohse et al. (1999, 2000). They also associated Internet purchasing and Internet purchase amounts with more years of online experience, and more email messages received per day (that is, direct measures of Internet experience). They also found a relationship between purchasing from catalogs (indirect experience) and online purchasing and amounts spent online. Where the studies differ is in the strength of the effects. This study used a multivariate procedure to remove redundancy present in the different predictor variables, and then predicted both the Internet purchase dollar amounts and purchase frequencies. Lohse et al. (1999, 2000) predicted Internet buying or not buying using

a logistic regression approach with correlated predictor variables. In their study, remote shopping experience was only the third most important predictor. In practice, it is difficult to compare effect sizes between studies predicting different aspects of the same variables. Our study modeled amounts of the dependent variables, whereas Lohse et al. (1999, 2000) modeled purchase/no purchase.

Results from our multivariate discriminant analysis agree with Lohse et al.'s (1999, 2000) finding that it is possible to achieve a high degree of classification accuracy when attempting to distinguish those who purchase from those who do not purchase on the Internet. We successfully classified 77 percent of those who purchased from those who did not purchase on the Internet using a 50 percent random hold out sample with the reliable predictor variables from the purchase frequency regression model. This is the same predictive power found in Lohse et al. (1999, 2000).

Although we did not focus on this in the results section, the predictive power of the different variables was different in the regression and discriminant analysis models. This difference clearly reflected the difference in the analytical tasks. This finding and the importance of experience-oriented factors suggest that our predictive understanding of factors that influence Internet purchases could be enhanced by a sequential or hierarchical modeling approach. That is, first understand the factors that are most predictive of making that first "leap of faith," that is, the first Internet purchase. Then model the factors that predict low to moderate levels of Internet purchases, and moderate to high levels of Internet purchases. It seems plausible that different aspects of experience and trust may come into play as the number of Internet purchases increases. For example, the positive influence of mail and phone purchases may decrease in terms of their predictive power for people who already make a lot of Internet purchases. In other words, the effects of surrogate experiences may saturate in terms of their predictive power. In this situation, the importance of having a high-speed Internet connection, a measure of direct Internet purchase experience, may be a stronger predictor for people who frequently purchase on the Internet as opposed to those who only purchase at moderate levels.

A second important conclusion from this research is that previous experience with mail and phone shopping acts to offset potential negative consequences of remote shopping on the Internet. This seems to occur because the basic structural risks are the same. For example, if you are shopping remotely, you do not have face-to-face contact, and

your perception of goods and services is dependent on "descriptive" materials provided by the vendor. In addition, remote shopping involves paying shipping charges, and more hassle if products need to be returned. Apparently, people who shop remotely trade off these negative aspects of remote shopping in exchange for higher levels of convenience. However, no amount of mail or phone order experience decreases concerns about privacy of personal information on the Internet or concerns about security of credit card information on the Internet. It appears that only direct Internet shopping experience acts to reduce these types of risks. Conceptually this implies that there are two types of risks. One is a basic structural consequence of remote shopping in general; the second is specific to Internet shopping.

This raises an important question. What is there about Internet purchasing in general that makes security and privacy concerns unique to that purchase modality? The security and privacy fear might be linked to the prevalence of hackers on the Internet. Some of the largest Internet companies have experienced major security intrusions where unauthorized access to confidential information has been gained. Privacy concerns may, in part, be related to "cookies" and related technologies that track how a person uses the Internet. These types of concerns appear to be relatively unimportant in our current models.

The results of Swaminathan et al. (2000) stress the importance of perceived convenience in predicting both online purchase frequency and amount. Once again, it is difficult to compare the results across studies because the studies have different research goals and the measures are not comparable. In the UCLA study the concept of convenience was measured mostly by our "perception of the Internet" factor. This factor contained elements such as "using the Internet saves time" and "ease of finding information on the Internet." We found that this factor was not a reliable predictor of either Internet purchase frequency or purchase amount. However, Swaminathan et al. (2000) did not include any measure for mail or phone shopping. People often use mail and phone shopping for purposes of convenience, and it's possible that a good deal of convenience information was captured by our mail and phone shopping factor. This possibility will be examined in more depth in future research.

In our analysis, social interaction as a motive for shopping was (inversely) part of the "negative aspects of shopping" component. We found some predictive value here and Swaminathan et al. (2000) did as well. Those who value the social aspect of shopping appear to buy less online. Swaminathan et al. (2000) also found that perceived

vendor reliability had a positive influence on the number of online purchases, but did not influence the amount of money spent on the Internet. Reliability was part of our factor on perceptions of the Internet, which was not a significant predictor. However, we also noted that reliability is related to trust which was a significant predictive factor in our analysis, especially as we related it to experience. Comparisons are especially challenging here as they do not include in their analysis the experience variables, both in terms of the Internet and other remote shopping modes, that we find to be the most important predictors.

Swaminathan et al. (2000) found that price competitiveness had a positive influence on the number of online purchases, but did not influence the amount of money spent on the Internet. We also found that perceived price advantage does not predict purchase amounts. But according to our evidence this variable is only a minor predictor for Internet purchase frequency.

In the realm of privacy and security Swaminathan et al. (2000) found that online consumers are less concerned about the security of electronic exchanges than non-consumers, but only marginally. This corresponds with our findings. On this matter Lohse et al. (1999, 2000), who found that purchasing and purchase amounts were negatively associated with privacy concerns, also agree.

In the realm of minor findings, we also mirror the results of Lohse et al. (1999, 2000) showing income predicting purchase amounts and maleness being positively associated with online shopping. The results involving the demographic variable of age, and gender as well, were interesting because they illustrate the dangers of confounding among predictive variables. Given that gender can only take on two values, it would be instructive to see if gender was an important predictor in terms of classifying people who purchase on the Internet from those who do not. Perhaps the leap of faith required to make that first Internet purchase has an association with age or gender.

In conclusion, our findings generally support those of Lohse et al. (1999, 2000), and Swaminathan et al. (2000). Discrepancies that exist are very probably the result of two factors. Lohse et al. (1999, 2000) were predicting the presence or absence of Internet purchasing and we were predicting amounts of purchase and frequency of Internet purchases. Differences with the results of Swaminathan et al. (2000) are more likely attributed to the different measures used in both studies, especially the absence of experience variables in their work. But the differences also might be due to the nature of the samples. Lohse et al.

used a large (10,180 in year one and 9,738 panelists in year two) random sample that closely matched the US online population. Swaminathan et al. used secondary data based on an email survey that respondents were invited to participate in through newsgroup and mailing list announcements, and banner ads. Based on the demographic information for this sample, it was clearly not projectable to the US population. The UCLA study was a random projectable sample.

References

Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A., and Wood, S. (1997). Interactive home shopping: consumer, retailer and manufacturer incentives to participate in electronic marketplaces. *Journal of Marketing*, 61(3), 38–54.

Bauer, R. A. (1960). Consumer behavior as risk taking. In D. F. Cox (ed.), *Risk taking and information handling in consumer behavior* (pp. 22–3). Cambridge, MA: Harvard University Press.

Bellenger, D. N. and Kargaonkar, P. K. (1980). Profiling the recreational shopper. *Journal of Retailing*, 56, 77–82.

Bloom, P. N., Milne, G. R., and Adler, R. (1994). Avoiding misuse of new information technologies: legal and societal considerations. *Journal of Marketing*, 58, 98–110.

BMRB International (1999, October 11). *Six different types of UK web shoppers*. Available online at: http://www.nua.ie/surveys/ index.cgi?f=VS&art_id=905355333&rel=true

Camp, L. J. (2000). *Trust and risk in Internet commerce*. Cambridge, MA: MIT Press.

Cranor, L. F., Reagle, J., and Ackerman, M. S. (1999). *Beyond concern: understanding net users' attitudes about online privacy*. AT&T Labs-Research Technical Report TR 99.4.4. Available online at: http://www.research.att.com/ projects/privacystudy/

Emeagwali, P. (2000). *History of the Internet: Emeagwali's recollections from 27 years*. Available online at: http://emeagwali.com/history/Internet/index.html

Feldman, D. L. (2000). Public confidence in cybersystems: issues and implications for sustainability. *International Political Science Review*, 21, 23.

Fox, S. (2000). *Trust and privacy online: why Americans want to rewrite the rules.* Pew Internet and American Life Project. Available online at: http://www.pewInternet.org/reports/toc.asp?Report=19

Friedman, B. and Kahn, P. H., Jr (2000). Trust online. *Communications of the ACM*, 43(12), 34–40.

Gefen, D. (2000). E-commerce: the role of familiarity and trust. *Omega*, 28(6), 725–37.

Gehrt, K. C., Yale, L. J., and Lawson, D. A. (1996). The convenience of catalog shopping: is there more to it than time? *Journal of Direct Marketing*, 10(4), 19–28.

Gillett, P. L. (1976). In-home shoppers: an overview. *Journal of Marketing*, 40, 81–8.

Grossman, W. (2000). Circles of trust. *Scientific American*, 283, 34.

Harris Interactive (June 8, 2000). *Six types of online shopper identified*. Available online at: http://www.nua.ie/surveys/index.cgi?f+VSandart_id=905355830andrel=true

Hoffman, D. L., Kalsbeek, W. D., and Novak, T. P. (1999). Building consumer trust online. *Communications of the ACM*, 42(4), 80–5.

Lohse, G. L., Bellman, S., and Johnson, E. J. (1999). Predictors of online buying behavior. *Comunications of the ACM*, 42(12), 32–8.

Lohse, G. L., Bellman, S., and Johnson, E. J. (2000). Consumer buying behavior on the Internet: findings from panel data. *Journal of Interactive Marketing*, 14, 15–29.

Luedi, A. F. (1997). Personalize or perish. *Electronic Markets*, 7(3), 22–5.

Milne, G. R. and Gordon, M. E. (1993). Direct mail privacy–efficiency trade-offs within an implied social contract network. *Journal of Public Policy and Marketing*, 12(2), 206–15.

Minahan, T. (1997). Internet buying is a tough sell. *Purchasing*, 122(1), 113.

Moorman, C., Deshpande, R., and Zaltman, G. (1993). Factors affecting trust in marketing relationships. *Journal of Marketing*, 57, 81–101.

Morgan, R. M. and Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58, 20–38.

Olson, J. S. and Olson, G. M. (2000). Trust in e-commerce. *Communications of the ACM*, 43, 41.

Palmer, J. W. (1997). Retailing on the WWW. *Electronic Markets*, 7(3), 6–9.

Reynolds, F. D. (1974). An analysis of catalog buying behavior. *Journal of Marketing*, 38(3), 47–51.

Rohm, A. J. and Milne, G. R. (1999). Consumers' privacy concerns about direct marketers' use of personal medical information. In J. F. Hair, Jr (ed.), *Proceedings of the 1999 Association for Health Care Research Conference* (pp. 27–37). Breckenridge, CO.

Sabo, J. T. (1997). Lessons learned: providing government Internet services with public trust. *Public Manager*, 26(3), 55–8.

Salnoske, K. (1998). Building trust in electronic commerce. *Business Credit*, 100(1), 24–5.

Schoder, D. and P.-L. Yin (2000). Building firm trust online. *Communications of the ACM*, 43(12), 73–9.

Sheth, J. N. and Parvatiyar, A. (1995). Relationship marketing in consumer markets: antecedents and consequences. *Journal of the Academy of Marketing Science*, 23, 255–71.

Sklar, D. (2001). Building trust in an Internet economy. *Strategic Finance*, 82(10), 22–5.

Swaminathan, V., Lepkowska-White, E., and Rao, B. P. (2000). Browsers or buyers in cyberspace? An investigation of factors influencing electronic exchange. *Journal of Computer Mediated Communication*, 5(2). Available online at: http://www.ascusc.org/jcmc/vol5/issue2/swaminathan.htm

UCLA Center for Communication Policy (2000). *The UCLA Internet Report*. Available online at: http://ccp.ucla.edu/pages/Internet-report.asp

UCLA Center for Communication Policy (2001). *The UCLA Internet Report*. Available online at: http://ccp.ucla.edu/pages/Internet-report.asp

Urban, G. L. and Sultan, F. (2000). Placing trust at the center of your Internet strategy. *Sloan Management Review*, 42(1), 39–48.

US Department of Commerce (1999). *The emerging digital economy II*. Available online at: http://www.ecommerce.gov/ede/

Wachter, R. M. (1999). The effect of gender and communication mode on conflict resolution. *Computers in Human Behavior*, 15, 763–82.

Zellweger, P. (1997). Web-based sales: defining the cognitive buyer. *Electronic Markets*, 7(3), 10–16.