

# Consumer Product Consideration and Choice at Purchase Time at Online Retailers

*Completed Research Paper*

**Bin Gu**

University of Texas at Austin  
McCombs School of Business,  
1 University Station, B6500  
Austin, TX 78712  
[bin.gu@mcombs.utexas.edu](mailto:bin.gu@mcombs.utexas.edu)

**Prabhudev Konana**

University of Texas at Austin  
McCombs School of Business,  
1 University Station, B6500  
Austin, TX 78712  
[prabhudev.konana@mcombs.utexas.edu](mailto:prabhudev.konana@mcombs.utexas.edu)

**Hsuanwei (Michelle) Chen**

University of San Francisco  
2130 Fulton Street  
San Francisco, CA 94117  
[hchen32@usfca.edu](mailto:hchen32@usfca.edu)

## Introduction

Research on electronic commerce has shown that information provided by online retailers, such as product recommendations and consumer reviews, have a significant impact on consumer product choice (Dellarocus 2006, Duan et al. 2008, Fleder and Hosnagar 2008, Lynch and Ariely 2000). However, few studies have considered the timing of the consideration and choice process. In this study, we analyze one aspect of this question – consumer consideration and choice at the time of purchase.

Understanding consumer consideration and choice at purchase time has both practical and theoretical values. From a practical perspective, it enriches retailers and product manufacturers' understanding of consumer decision process. In particular, it shows the degree to which consumers conduct information search and make purchase decision in the same online session. Such information could play an important role in real-time marketing strategies. From a theoretical perspective, prior studies on consumer consideration mainly focus on offline setting where consumer browsing behavior is not observable. As such, identifying consumer consideration and choice requires restrictive assumptions. In this study, we leverage a unique data set and show identification could be exact in some cases while close bounds can be obtained in other cases.

Our research is facilitated by the availability of an increasingly popular feature among online retailers - "What Do Customers Ultimately Buy After Viewing This Item?" Two of the largest online retailers, Amazon.com and Buy.com, offer this feature prominently on their product pages. We show that, using information from this feature and product sales data, we can exactly identify the size of different consumer segments when analyzing competition between two or three products and obtain a lower bounds on the size of consumer segments that consider no alternative product for analysis involving more than three products.

Our approach is simple and robust compared to prior models proposed in the marketing literature (Andews and Srinivasan1995, Nedungadi 1990, Roberts and Lattin 991). These models either require detailed micro level data or make restrictive assumptions. We apply our model on 7,000 unique products collected from Amazon's Electronics category. The results show that more than 78 percent of consumers who purchase at Amazon does not consider any other products. The finding suggests a majority of consumers have made the decision on their purchase before purchase time.

The remainder of this paper proceeds as follows. In Section 2, we review the prior literature as it relates to our research topic. In Section 3, we present the theories and methodologies to identify consumer

consideration sets. We present our data and modeling results along with discussions in Sections 4 and 5, respectively. We conclude our paper in Section 6.

## **Literature Review**

The prevalence of electronic commerce in recent years has inspired growing interests in product variety and consumer purchase decisions in online retailers. The first stream of related literature is concerned about the increased product variety introduced by online retailers (Anderson 2006, Brynjolfsson et al. 2003, Ostreicher-Singer and Sunararajan 2006). Studies suggest that online retailers are able to carry a greater variety of products than their physical counterparts (Brynjolfsson et al. 2006), thus expand the number of products considered by a consumer. In addition, lower search costs facilitated by product information and consumer recommendation in digital commerce further help consumers discover niche products and expand product consideration (Anderson 2006).

Our research also draws on the literature relating to the impact of electronic commerce on consumer search behavior. With the increase of marketing channels facilitated by the advance of technologies, media, and advertising activities, consumers have access to a great amount of information. The reduction in search costs enabled by the internet further facilitates consumer information search behavior. Studies suggest that consumers conduct information search in variety of contexts. Bloch et al. (1986) and Moe (2003) find that information search occurs both within and outside of the purchase process. Consumers often conduct exploratory information search for knowledge building, hedonic needs or recreational purpose.

The limitation of human cognitive and perceptual capability, however, restricts the amount of search conducted by a consumer. Literature on consumer search behavior finds that consumers often conduct limited amount of search online despite the low search costs (Johnson et al. 2004). Research also suggests that because of the limit in consumers' cognitive capability, bounded rationality (DeMarzo et al. 2003, Shocker et al. 1991) and information overload (Jones et al. 2004, Rogers and Agarwala-Rogers 1975) have forced online users to be more selective in processing information. As such, consumers often consider a fairly small set of products when making purchases among a wide range of product alternatives.

This study is also related to marketing literature on consumer consideration formation (Haubl and Trift 2002, Neungadi 1990, Roberts and Lattin 1991). These studies have shown that, consumers' consideration of products will significantly impact their ultimate choices of products. In particular, faced with cognitive limitations, complex choice tasks, and evaluation costs, consumers will resort to phased decision strategies (Gensch 1987). The phased consumer decision making process involves two stages. In the Consideration Stage, consumers consider a smaller set of products and form the choice set. In the Choice Stage, consumers evaluate every product in the choice set and purchase the one with the highest evaluation. In online environment, the consideration and choice stages are often separate. Consumers may conduct extensive information search through friends, third-party intermediaries (e.g. CNET) before they make the purchase at an online retailer (Gu et al. 2011). We complement this research stream by identifying consumers' choice set at the time of purchase.

## **Theories and Methodologies**

### ***Revealed Preferences in Online Retailers***

Online retailers provide a variety of product sales statistics to help consumers make better decisions. For example, Amazon provides the following statistics to their customers: 1) products "Frequently Bought Together", 2) "What Do Customers Ultimately Buy After Viewing This Item", 3) product sales rank, 4) "Customers Who Bought This Item Also Bought", and 5) Customer Reviews. These statistics reveal consumer preference in purchasing decisions and provide an important vehicle that allows researchers and practitioners to infer underlying purchase process. In this study, we show that two statistics - "What Do Customers Ultimately Buy After Viewing This Item" and product sales rank - can be used to identify the sizes of consumer segments with different consideration sets.

Figure 1 provides a screenshot of “What Do Customers Ultimately Buy After Viewing This Item”. The figure shows that Amazon provides the top 4 products ultimately bought by consumers after viewing a given product and the corresponding percentages of consumers who have done so.



We note that the percentages in Figure 1 represent conditional probabilities of consumers purchasing product Y after they have viewed X. We show below that consumer consideration sets and their choices can be exactly identified using these conditional probabilities and the Amazon sales data when analyzing competition between any two or three products. Further, we show that the information provides a lower bound of consumer segments that view only one product when analyzing competition among more than three products.

### Competition Between Two Products

Product competition often centers around two products that are close substitute in nature. In such cases, retailers and product manufacturers often focus on analyzing the competition between the two products while ignoring consumers who bought other products.

Consider a market with two products, X and Y. Each consumer purchases one and only one product. This assumption is due to the fact that Amazon’s statistics are reported on only consumers who made a purchase. Consumers who viewed a product but did not make any purchase subsequently are excluded from the statistics. Consumers make purchase decisions in two stages. In the *Consideration Stage*, they form a consideration set that could include one of the two products, or both. In the *Choice Stage*, they make purchase decisions. Given that the market contains two products, there exist three customer segments given their consideration sets at purchase time: i) those consider only X, ii) those consider only Y and iii) those consider both X and Y. We use  $N(X\bar{Y})$ ,  $N(\bar{X}Y)$ , and  $N(XY)$  to denote the three segments. To identify the size of each segment, we note that Amazon provides the following information:

- $x$  - number of consumers who purchased X (we use the lower case to indicate sales and the upper case to indicate views) as estimated from Amazon sales rank of product X
- $y$  - number of consumers who purchased Y as estimated from Amazon sales rank of product Y
- $P(x|X)$  – percentage of customers who purchased product X after viewing X.
- $P(y|X)$  – percentage of customers who purchased product Y after viewing X.
- $P(x|Y)$  – percentage of customers who purchased product X after viewing Y.
- $P(y|Y)$  – percentage of customers who purchased product Y after viewing Y.

It is useful to note that, given that the market has only two products and consumers purchase one and only one product,  $P(x|X) + P(y|X) = 1$  and  $P(x|Y) + P(y|Y) = 1$ . This indicates that only two of the four conditional probabilities carry unique information.

We further note that the probability of purchasing a product after viewing the product can be expressed as a ratio of product sales over the sum of all related customer segments that have viewed the product. In particular,

1.  $P(x|X) = \frac{x}{N(X\bar{Y})+N(XY)}$ . The equation suggests that the probability of consumers purchasing X after viewing X can be calculated as the sales of X divided by the sum of number of customers who viewed only X and number of customers who viewed both X and Y. Similarly, we have
2.  $P(y|Y) = \frac{y}{N(\bar{X}Y)+N(XY)}$

Finally, we note that, since consumers purchase one and only one product, total number of viewers equals to total number of customers.

$$3. x + y = N(X\bar{Y}) + N(\bar{X}Y) + N(XY)$$

It is straightforward to identify the unique solution for the above equations (1)-(3):

4.  $N(XY) = \frac{x}{P(x|X)} + \frac{y}{P(y|Y)} - x - y = \frac{P(y|X)}{P(x|X)}x + \frac{P(x|Y)}{P(y|Y)}y$
5.  $N(X\bar{Y}) = x + y - \frac{y}{P(y|Y)}$
6.  $N(\bar{X}Y) = x + y - \frac{x}{P(x|X)}$

Intuitively,  $\frac{x}{P(x|X)}$  identifies the total number of consumers who have viewed product X while  $\frac{y}{P(y|Y)}$  identifies the total number of consumers who have viewed product Y. Numbers of consumers who have viewed both products are counted twice in  $\frac{x}{P(x|X)} + \frac{y}{P(y|Y)}$ . We can thus identify these consumers using the difference between the sum of the two and the total number of consumers (i.e. equation 4). In the extreme case where every consumer views only one product,  $P(x|X)$  and  $P(y|Y)$  equal to 1 and,  $N(XY)$ ,  $N(X\bar{Y})$ ,  $N(\bar{X}Y)$  equal to 0,  $x$  and  $y$  respectively.

We can also identify consumer choice decisions between the two products. To identify the choice process of those who viewed both products, we note that the  $x$  customers who bought product X can be divided into two groups: those who viewed only product X and those who viewed both product X and Y. Since there are  $N(X\bar{Y})$  customers in the first group, the second group contains  $x - N(X\bar{Y})$  customers. The choice probability of  $x$  among customers that viewed both products is thus

$$P(x|XY) = \frac{x - N(X\bar{Y})}{N(XY)} = \frac{\frac{P(x|Y)}{P(y|Y)}y}{\frac{P(y|X)}{P(x|X)}x + \frac{P(x|Y)}{P(y|Y)}y}$$

Figure 2 illustrates our approach using sales rank data and conditional probability data on two popular software products: Adobe Photoshop Element 9 (PS 9) and Adobe Photoshop and Premiere Element 9 (Bundle 9). The former is popular photo editing software for amateurs, while the later is a bundle product that contains one copy of Adobe Photoshop Element 9 and one copy of Adobe Premiere Element 9, a popular video editing software for video enthusiasts. Data from Amazon indicates that fewer than 2 percent of consumers who have viewed either product end up purchasing something else. So the competition is mainly between the two products at purchase time. We thus rescale the data to remove consumers who ultimately bought other products and focus on consumers who bought either of the two products. For illustration, we assume that the relationship between sales rank and sales is known. The analysis shows that only 13% of the consumers consider both products. The remaining 87% of consumers consider only one product at purchase time. Among those who consider both products, 55% choose to purchase PS 9.

**I. Information from Amazon**

Bought Viewed	Photoshop Element 9	Bundle 9	Sales Rank
Photoshop Element 9	93%	7%	3
Bundle 9	23%	77%	21

**II. Derivation of Consumer Segments**

i. Viewed both products =  $\frac{x}{P(x|X)} + \frac{y}{P(y|Y)} - x - y = 264.03^1$

ii. Viewed only Adobe Element 9 =  $x + y - \frac{y}{P(y|Y)} = 1,447.46$

iii. Viewed only Adobe Element 9 + Adobe Premier Element 9 =  $N(\bar{X}Y) = x + y - \frac{x}{P(x|X)} = 363.05$ .

Probability of Viewing both Products Before Purchase = 12.73%.

**III. Derivation of Consumer Choice**

Probability of buying Adobe Photoshop Element 9 after viewing Both Product  $\frac{x - N(X\bar{Y})}{N(XY)} = 54.63\%$ .

**Figure 2. Illustration of Consumer Consideration and Choice of Two Products**

**Competition Between Three Products**

To identify consumer consideration and choice set for more products, we note that there exist  $2^{n-1}$  consumer segments for competition between  $n$  products given the permutation of consideration sets. In addition, to exactly identify consumer choice within all consideration sets, we need identification of  $\sum_{k=1}^n (k-1) \binom{n}{k}$  choice probabilities. In total,  $2^{n-1} + \sum_{k=1}^n (k-1) \binom{n}{k}$  variables need to be identified. Since Amazon provides  $n^2$  statistics<sup>2</sup>, such exact identification is only possible for  $n=2$ .

The number of variables required for the choice process can be significantly reduced if we use a discrete choice model to model the consumer choice process. A discrete choice model assumes that each product has a fixed utility in all consumer segments and a consumer's probability of purchasing the product equals to the ratio of the product utility over the sum of the utilities of all products considered by the consumer. This assumption reduces the number of variables required for the choice process to  $(n-1)$ . In total,  $2^n + n - 2$  variables need to be identified. Since Amazon provides  $n^2$  statistics, the identification is feasible for  $n = 3$ .

To identify consumer segments and choice process for three products, we use the following nine equations. The first set of three equations identifies the conditional probability of purchasing a product after viewing

<sup>1</sup> We assume the following power law relationship (Chevalier and Goolsbee 2003, Ghose et al. 2006) between product sales and sales rank:  $\text{Log Sales} = 8.046 - 0.613 \times \text{Log SalesRank}$ .

<sup>2</sup> Amazon provides  $n$  statistics on product sales, and  $n(n-1)$  statistics on the conditional probabilities. In total, we have  $n^2$  statistics.

the product. The second set of three equations identifies the conditional probability of purchasing a different product after viewing a given product. The final set of three equations identifies sales for each product. It is useful to note that the second and third sets of equations are non-linear and thus require numeric solution.

$$\begin{aligned}
 x|X &= \frac{x}{N(X\bar{Y}\bar{Z}) + N(XY\bar{Z}) + N(X\bar{Y}Z) + N(XYZ)} \\
 y|Y &= \frac{y}{N(\bar{X}Y\bar{Z}) + N(XY\bar{Z}) + N(\bar{X}YZ) + N(XYZ)} \\
 z|Z &= \frac{z}{N(\bar{X}\bar{Y}Z) + N(X\bar{Y}Z) + N(\bar{X}YZ) + N(XYZ)} \\
 y|X &= \frac{\frac{u_y}{(u_x + u_y)}N(XY\bar{Z}) + \frac{u_y}{(u_x + u_y + u_z)}N(XYZ)}{N(X\bar{Y}\bar{Z}) + N(XY\bar{Z}) + N(X\bar{Y}Z) + N(XYZ)} \\
 x|Y &= \frac{\frac{u_x}{(u_x + u_y)}N(XY\bar{Z}) + \frac{u_x}{(u_x + u_y + u_z)}N(XYZ)}{N(\bar{X}Y\bar{Z}) + N(XY\bar{Z}) + N(\bar{X}YZ) + N(XYZ)} \\
 x|Z &= \frac{\frac{u_x}{(u_x + u_z)}N(X\bar{Y}Z) + \frac{u_x}{(u_x + u_y + u_z)}N(XYZ)}{N(\bar{X}\bar{Y}Z) + N(X\bar{Y}Z) + N(\bar{X}YZ) + N(XYZ)} \\
 x &= N(X\bar{Y}\bar{Z}) + \frac{u_x}{(u_x + u_y)}N(XY\bar{Z}) + \frac{u_x}{(u_x + u_z)}N(X\bar{Y}Z) + \frac{u_x}{(u_x + u_y + u_z)}N(XYZ) \\
 y &= N(X\bar{Y}\bar{Z}) + \frac{u_y}{(u_x + u_y)}N(XY\bar{Z}) + \frac{u_y}{(u_y + u_z)}N(\bar{X}YZ) + \frac{u_y}{(u_x + u_y + u_z)}N(XYZ) \\
 z &= N(X\bar{Y}\bar{Z}) + \frac{u_z}{(u_x + u_z)}N(X\bar{Y}Z) + \frac{u_z}{(u_y + u_z)}N(\bar{X}YZ) + \frac{u_z}{(u_x + u_y + u_z)}N(XYZ)
 \end{aligned}$$

Figure 3 illustrates our approach by extending our earlier example to three software products: Adobe Photoshop Element 9 (PS 9) and Adobe Photoshop and Premiere Element 9 (Bundle 9), and Adobe Premier Element (PR 9). The Data from Amazon shows that while few customers who have viewed PS 9 or Bundle 9 choose PR 9, the reverse is not true. Half of the customers who viewed PR 9 end up with purchasing PS 9 or Bundle 9. The illustration shows that few consumers consider more than one product at purchase time. Our result reveals that 80.35% of the consumers consider only one product at purchase time. Our results also indicate PR 9 has a much lower utility compared with that for PS 9 or Bundle 9.

**I. Information from Amazon**

<b>Bought Viewed</b>	<b>Photoshop Element 9</b>	<b>Bundle 9</b>	<b>Premiere Element 9</b>	<b>Sales Rank</b>
<b>Photoshop Element 9</b>	92%	7%	1%	3
<b>Photoshop/Premiere Element Bundle 9</b>	23%	76%	1%	21
<b>Premiere Element 9</b>	34%	16%	50%	135

**II. Derivation of Consumer Segments**

The derivation is based on numeric solution to the nine non-linear equations outlined above.

- i. Viewed Only Adobe Photoshop Element 9 = 1,346.73
- ii. Viewed Only Adobe Photoshop / Premier Element Bundle 9 = 312.36
- iii. Viewed Only Adobe Premier Element 9 = 136.78
- iv. Viewed Both Adobe Photoshop Element 9 and Adobe Photoshop/Premier Element Bundle 9 = 267.23
- v. Viewed Both Adobe Photoshop Element 9 and Adobe Premier Element 9 = 116.13
- vi. Viewed Both Adobe Photoshop/Premier Element Bundle 9 and Adobe Premier Element 9 = 55.73
- vii. Viewed All Three Products = 0
- viii. The probability of PS 9 being considered is 78%.
- ix. The probability of Bundle 9 being considered is 29%
- x. The probability of PR 9 being considered is 14%.

**III. Derivation of Consumer Choice**

- i. Utility of Adobe Photoshop Element 9 = 0.517
- ii. Utility of Adobe Photoshop/Premier Element Bundle 9 = 0.428
- iii. Utility of Adobe Premier Element 9 = 0.055

The above estimated utilities suggest the following choice probability

- 1) Probability of buying Adobe Photoshop Element 9 after viewing both Adobe Photoshop Element 9 and Adobe Photoshop/Premier Element Bundle 9 is 54.68%
- 2) Probability of buying Adobe Photoshop Element 9 after viewing both Adobe Photoshop Element 9 and Adobe Premier Element 9 is 90.36%.
- 3) Probability of buying Adobe Photoshop/Premier Element Bundle 9 after viewing both Adobe Photoshop/Premier Element Bundle 9 and Adobe Premier Element 9 is 88.60%.

**Figure 3. Illustration of Consumer Consideration and Choice of Three Products**

### Competition Among Multiple Products

To consider competition among more than three products, we show below that, while exact identification of consumer segments are impossible, we can identify the lower bound of percentage of consumers who consider only one product at purchase time. We use  $D_i$  to denote the action of viewing product  $i$  and  $d_i$  to denote the number of consumers who purchase product  $i$ . Amazon provides the following information:

- a.  $d_i$  - number of consumers who purchased product  $i$ , for all  $i$ .
- b.  $P(d_j|D_i)$  – probability of consumers purchase  $j$  after viewing product  $i$ , for all  $i$  and  $j$ .

We again note that  $\sum_j P(d_j|D_i) = 1$ , for all  $i$ . This condition indicates that while (b) provides a total number of  $n^2$  statistics, only  $n(n-1)$  of them contain unique information.

Given that the conditional probability is between each pair of products but consumers may consider more than two products in this setting, we can not exactly identify all consumer segments for all permutation of consideration sets. However, the above information is sufficient to provide a lower bound of consumers who consider no alternative products at purchase time.

Note that the conditional probability of purchasing a product after viewing the product can be expressed as follow:

$$P(d_i|D_i) = \frac{d_i}{N(D_i\overline{D_{-1}}) + \sum_{j \neq i} N(D_i D_j \overline{D_{-1-j}}) + \dots + N(D_i D_{-1})}$$

In the above equation,  $N(D_i\overline{D_{-1}})$  refers to consumers who only view product  $i$  at purchase time.  $N(D_i D_j \overline{D_{-1-j}})$  refers to consumers who only view product  $i$  and  $j$  at purchase time and  $N(D_i D_{-1})$  refers to consumers who view all the products at purchase.

Switching the LHS and RHS, we have

$$N(D_i\overline{D_{-1}}) + \sum_{j \neq i} N(D_i D_j \overline{D_{-1-j}}) + \dots + N(D_i D_{-1}) = \frac{d_i}{P(d_i|D_i)}$$

Summing the LHS over all  $i$ , we have

$$\sum_i N(D_i\overline{D_{-1}}) + \sum_i \sum_{j \neq i} N(D_i D_j \overline{D_{-1-j}}) + \dots + \sum_i N(D_i D_{-1}) = \sum_i \frac{d_i}{P(d_i|D_i)}$$

Note that

$$\begin{aligned} \sum_i N(D_i D_{-1}) &= nN(D_i D_{-1}) \\ \sum_i \sum_{j \neq i} N(D_i D_j \overline{D_{-1-j}}) &= 2 \sum_{\text{all } i < j} N(D_i D_j \overline{D_{-1-j}}) \end{aligned}$$

We have

$$\sum_i N(D_i\overline{D_{-1}}) + 2 \sum_{\text{all } i < j} N(D_i D_j \overline{D_{-1-j}}) + \dots + nN(D_i D_{-1}) = \sum_i \frac{d_i}{P(d_i|D_i)}$$

We further note that the sum of all consumer segments equals to the number of total consumers, i.e.

$$\sum_i N(D_i\overline{D_{-1}}) + \sum_{\text{all } i < j} N(D_i D_j \overline{D_{-1-j}}) + \dots + N(D_i D_{-1}) = \sum_i d_i$$

We thus have

$$\sum_{\text{all } i < j} N(D_i D_j \overline{D_{-1-j}}) + \dots + (n - 1)N(D_i D_{-1}) = \sum_i \left( \frac{d_i}{P(d_i|D_i)} - d_i \right)$$

Note that the total number of customers who consider more than one product is:

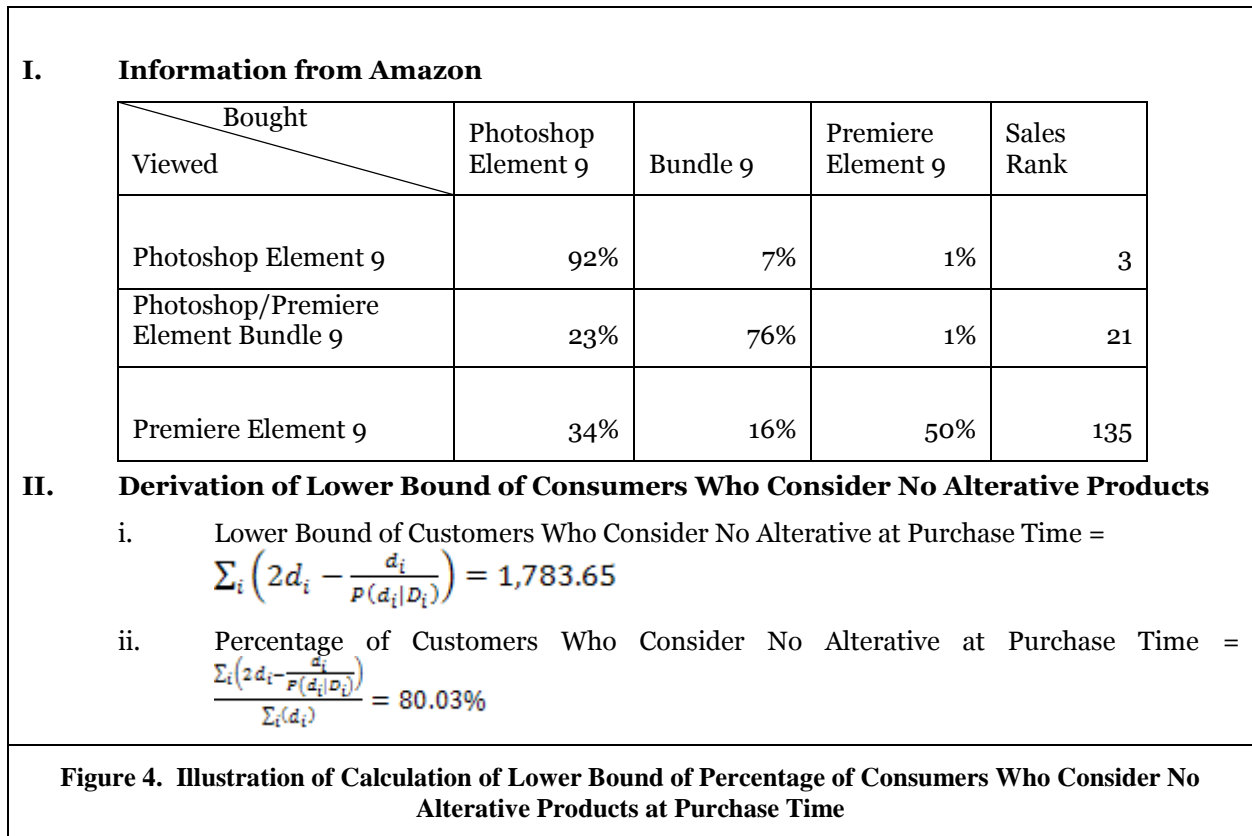


$$\sum_{\text{all } i < j} N(D_i D_j \overline{D_{-i-j}}) + \dots + N(D_i \overline{D_{-i}}) < \sum_{\text{all } i < j} N(D_i D_j \overline{D_{-i-j}}) + \dots + (n - 1)N(D_i \overline{D_{-i}})$$

$$= \sum_i \left( \frac{d_i}{P(d_i|D_i)} - d_i \right)$$

$\sum_i \left( \frac{d_i}{P(d_i|D_i)} - d_i \right)$  thus identifies the upper bound of number of customers who consider more than one product. Alternatively, we can express the lower bound of consumers who consider no alternative product as  $\sum_i \left( 2d_i - \frac{d_i}{P(d_i|D_i)} \right)$ . This bound is tight if few customers consider three products or more at purchase time. In addition, one advantage of this bound is that it is derived without any assumption on the choice process.

Figure 4 uses our earlier example of three Adobe products to identify the lower bound of customers who consider no alternative products at purchase time. The result shows that the bound is 80.03%, very close to the percentage (80.35%) identified in Figure 3.



## Data

The data we collected are from publicly available statistics of purchase propensity at Amazon’s “What Do Customers Ultimately Buy After Viewing This Item?” (see Figure 1). The data are extracted using automated scripts to access and parse HTML pages from the retailer. We collected 7,000 unique products under Electronics category, as illustrated in Table 1. Each product is accompanied with the conditional percentages of consumers buying one product after viewing other top three alternative products. For every electronic product, the following information is included: conditional purchase propensity, sales rank, and selling price.

Our data were collected on May 15th, 2008. Table 1 lists summary statistics for our data. Note that in Table 1, C1 refers to the conditional purchase percentage for the most purchased product after viewing the product page. In most cases, the most purchased product is the product being viewed. Similarly, C2

refers to the conditional purchase percentage of the second most purchased product after viewing the product page%. C3 and C4 follow the same rationales for the third and the fourth most purchased products, respectively.

**Table 1: Summary Statistics**

<b>Variable</b>	<b>Obs.</b>	<b>Mean</b>	<b>Std dev</b>	<b>Min</b>	<b>Max</b>
Sales Rank	7,000	345.49	3,337.75	1	249,242
Sale Price	7,000	116.40	209.59	0.01	2,799.77
Rank 1 Purchase Propensity (C1)	7,000	69.90%	0.14	25%	99%
Rank 2 Purchase Propensity (C2)	7,000	12.28%	0.11	0.5%	45%
Rank 3 Purchase Propensity (C3)	7,000	5.42%	0.05	0.5%	32%
Rank 4 Purchase Propensity (C4)	7,000	3.28%	0.03	0.5%	17%

## Empirical Results

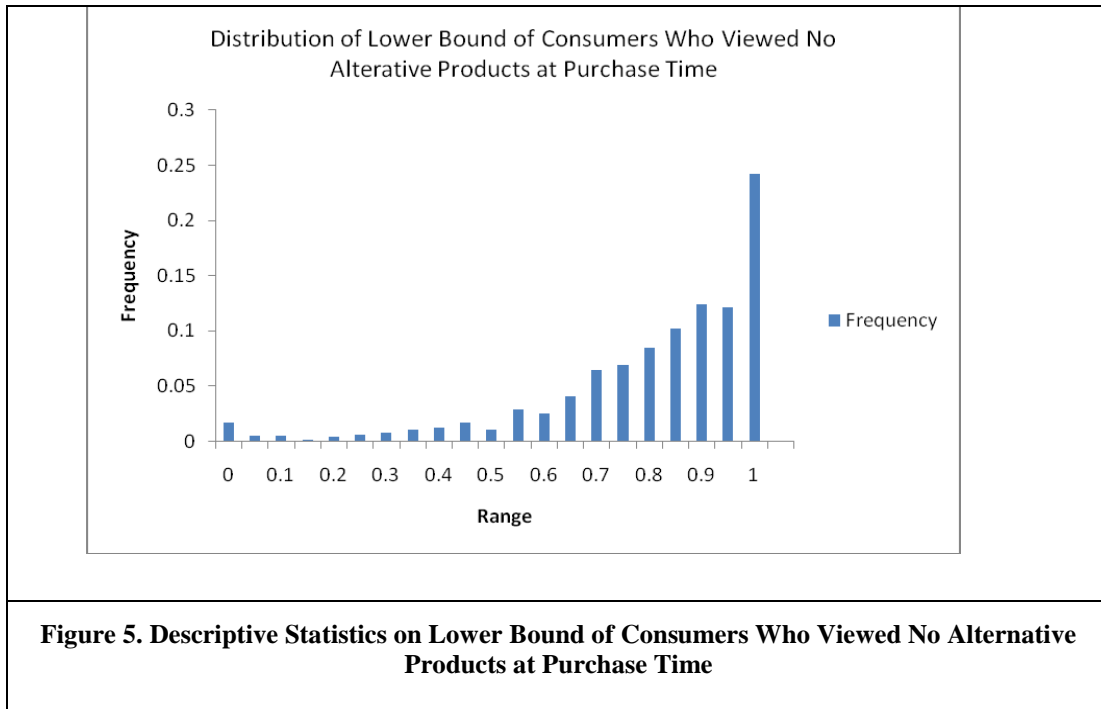
### Consideration Set

We start our analysis by estimating the lower bound of number of customers who view no alternative products. We conduct the above analysis for each group of four products listed by Amazon. Figure 5 shows the mean, standard deviation and the histogram of the estimation. The results show that, on average, 78% of customers viewed no alternative products. The histogram further shows that, for a large number of the groups, over 90% of consumers made purchase without viewing any other products.

**Panel A: Summary Statistics**

<i>Mean</i>	<i>Std Dev</i>	<i>Min</i>	<i>Max</i>
78.21%	21.73%	0.00%	98.46%

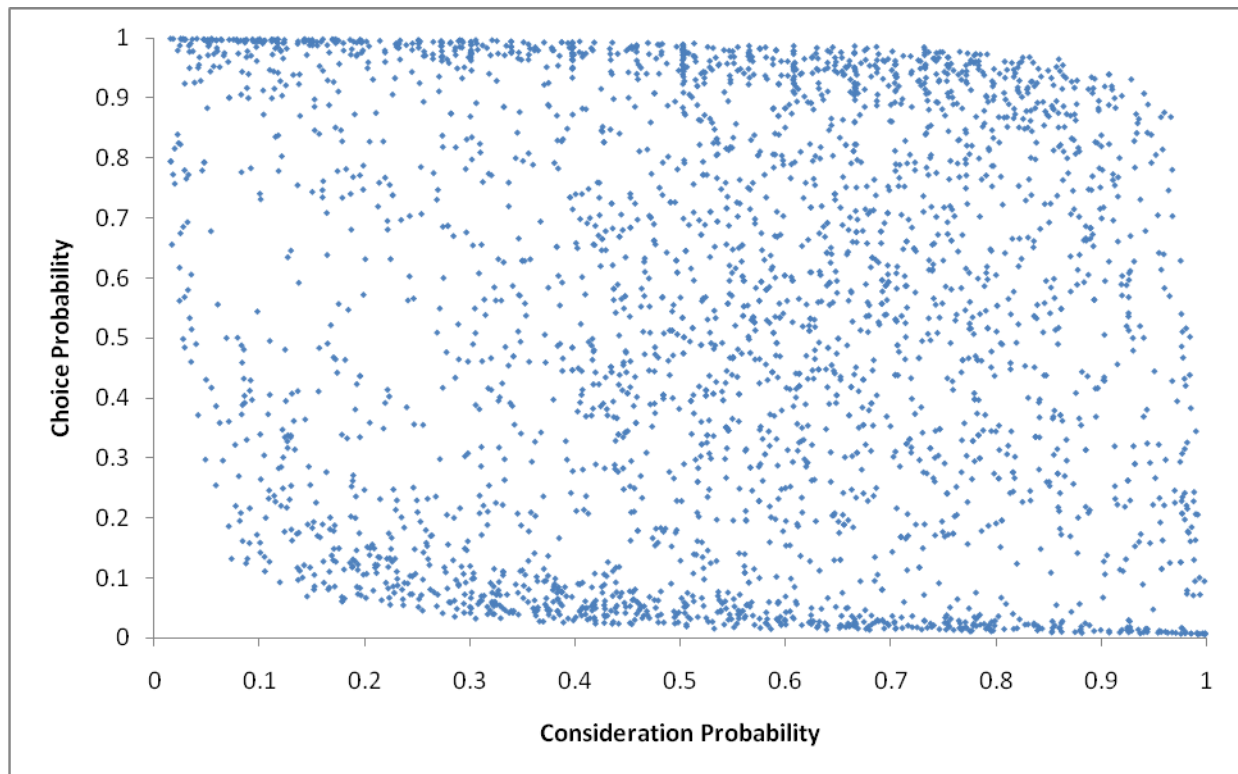
**Panel B: Histogram**



The above analysis shows that few consumers consider more than one product at purchase time. We also note from Table 1 that most of the consideration is between the two leading products in each group. We thus focus on analyzing two-product competition below using the approach outlined earlier.

Panel A in Figure 5 shows the summary statistics of percentage of consumers who considered a given product in each group and percentage of consumers who purchase a given product after viewing both products. The probability of a given product being considered by a consumer (“consideration probability” from here on) is 57%. Note that if each consumer considers only one product, the average consideration probability in a two-product group shall be 50%, while if every consumer considers both product, the probability shall be 100%. The result provides another indication that few consumers consider more than one product at purchase time. Panel A also shows that there is a negative correlation between a product’s consideration probability and its probability of being purchased conditional on being considered by a consumer (“choice probability” from here on). The negative correlation indicates that products that are more likely to be considered often provide lower utility. Panel B of Figure 6 shows the scatterplot of the relationship between consideration probability and choice probability. The x-axis represents the probability of a product being considered and the y-axis represents the probability of the product being chosen if compared side by side against the competing product. The plot indicates that the relationship between the two probabilities vary substantially. In some groups, a product’s consideration probability is significantly higher than its choice probability, indicating that product sales are mainly driven by a product being frequently considered by consumers. In other groups, a product’s choice probability is significantly higher than its consideration probability, indicating that, while the product is less known among consumers, it offers higher utility compared to its direct competitor.

<b>Panel A: Summary Statistics and Correlation</b>						
Variable	Mean	Std Dev	Min	Max	Consideration Probability	Choice Probability
Consideration Probability	57.26%	28.09%	0.02%	99.94%	1.00	-0.22
Choice Probability	50.00%	34.91%	0.00%	99.94%	-0.22	1.00

**Panel B: Scatter Plot****Figure 6. Scatter Plot of Consideration Probability vs. Choice Probability**

### ***The Effects of Consideration Probability and Choice Probability on Purchase***

To compare the relative impact of consideration and choice on product sales, we note there is a non-linear relationship between consideration, choice and product sales. As such, we cannot use linear regression or ANOVA for the analysis. Instead, to demonstrate the effect of consideration and choice on product purchase, we do so some separately for each factor. We first assess the influence of consideration by removing the influence of choice with the assumption that consumers have equal probability of choosing either product in the consideration set. We then calculate predicted product sales and report the summary statistics of predicted product sales and its correlation with actual sales in Table 2. Since the percentage of variation explained in product sales equals to the square of the correlation, the result suggests that consideration alone explains 98% of the variation. We conduct the same analysis for the choice probability by assuming consumers give equal consideration to products. The result in Table 2 suggests that choice alone explains only 8.4% of the variation.

**Table 2: Predicted Sales Based on Consideration Probability and Choice Probability**

Variable	Mean	Std dev	Min	Max	Corr with Actual Sales
Predicted Sales Based on Consideration Probability	865.75	945.17	0	4234	0.99
Predicted Sales Based on Choice Probability	681.50	837.30	1.13	5894	0.29

## Conclusion

In this study, we develop a methodology to identify consumer consideration and choice at purchase time. We show that most consumers consider only one product at the purchase time. There are two possible explanations of the result: consumers either do not conduct product search before purchase, or they engage in pre-purchase information and product comparison elsewhere before visiting online retailers to make the purchase. Given our analysis is conducted in the electronic category, where products are relatively expensive; we believe the second explanation is likely to be true. This finding is also consistent with recent studies that show third-party infomediaries have a significant impact on product sales at online retailer.

This research suggests that most consumers have narrowed down to one or two products when making final purchase decisions. Our analysis also reveals that the majority of variations in product demand are due to products' propensity of being considered. Product utility has limited influence at the final stage. Our analysis has significant implications for product, price, and marketing strategies for online retailers and manufacturers. The finding that consumers only consider a small set of products when making final purchase decisions suggest that consumer purchase decision is likely to be a multi-stage process. Consumers often conduct pre-purchase information search and narrow down product choices elsewhere before visiting online retailers to make the final purchase.

By developing a model to measure product consideration and choice using only publicly available aggregate statistics from online retailers, we also contribute to the literature from a methodological perspective. While prior two-stage consideration choice model requires micro-level data that is usually not available to manufacturers, we develop a methodology that shows consideration and choice can be exactly identified under some conditions using only aggregated information in online markets. This is particularly important for electronic commerce research given that micro-level transactions are not generally accessible.

Our study has a number of limitations. First, Amazon statistics are based on consumer behavior within a purchase session. It doesn't consider consumers' previous browse and search behavior in non-purchase sessions. As such, the consideration set identified in this study is the product choice set at the time that consumers make product purchase. While there is significant value in understanding the size of the final choice set, our results does not imply that consumers consider few products in the overall information search process. Second, our model requires modeling the relationship between Amazon sales rank and product sales. The relationship has been studied extensively for books, but few studies have considered it for electronic products. Sensitivity analysis will be needed to assess the robustness of our approach that uses the function form derived from book sales for electronic products. Third, our methodology can only be applied to substitute products, but not to complementary products. Complementary consumption requires different modeling techniques about consumers' consideration and choice (Gentzkow 2007), and the issues are not covered in this study. Finally, it would be valuable for future studies to obtain retailer server log data to validate our approach and assess the accuracy of the low bounds derived in this study.

## Reference

- C. Anderson. 2006. *The Long Tail*, New York, NY: Hyperion.
- R. L. Andrews, T. C. Srinivasan. 1995. "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data," *Journal of Marketing Research* (32:1) pp. 30-41.
- A. D. Baddeley. 2001. "Is Working Memory Still Working?" *American Psychologist* (56:11) pp. 851-64.

- M. R. Baye, J. Morgan, P. Scholten. 2004. "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site," *Journal of Industrial Economics* (52:4) pp. 463-496.
- S. Berry, J. Levinsohn, A. Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy* (112:1) pp. 68-105.
- P. H. Bloch, D. L. Sherrell, N. M. Ridgway. 1986. "Consumer Search: An Extended Framework," *Journal of Consumer Research* (13: 1) pp. 119-126
- J. R. Brown, A. Goolsbee. 2002. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy* (110:3) pp. 481-507.
- E. Brynjolfsson, Y. J. Hu, M. D. Smith. 2003. "Consumer Surplus in the Digital Economy: Estimating The Value of Increased Product Variety at Online Booksellers," *Management Science* (49:11) pp. 1580-1596.
- E. Brynjolfsson, Y. J. Hu, M. D. Smith. 2006. "From Niches to Riches: The Anatomy of the Long Tail," *Sloan Management Review* (47:4) pp. 67-71.
- P.-Y. Chen, L. M. Hitt. 2002. "Measuring Switching Costs and the Determinants of Consumer Retention in Internet-Enabled Businesses: A Study of The Online Brokerage Industry," *Information Systems Research* (13:3) pp. 255-274.
- J. Chevalier, A. Goolsbee. 2003. "Measuring Prices and Price Competition Online: Amazon and Barnes & Noble," *Quantitative Marketing Economics* (1:2) 203-222.
- J. Chevalier, J., A. Goolsbee. 2005. "Are Durable Goods Consumers Forward Looking? Evidence from College Textbooks," *Yale School of Management Working Paper*, New Haven, CT.
- K. Clay, R. Krishnan, E. Wolff. 2001. "Price Strategies on the Web: Evidence from the Online Book Industry," *Journal of Industrial Economics* (49:4) pp. 521-540.
- E. K. Clemons, G. G. Gao, L.M. Hitt. 2006. "When Online Reviews Meet Hyperdifferentiation," *Journal of Management Information Systems* (23:2) pp. 149-171.
- C. Dellarocus. 2006. "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science* (52:10) pp. 1577-1593.
- P. M. DeMarzo, D. Vayanos, J. Zwiebel. 2003. "Persuasion Bias, Social Influence, And Unidimensional Opinions," *Quarterly Journal of Economics* (118:3) pp. 909-968.
- W. Duan, B. Gu, A. Whinston. 2008. "The Dynamics of Online Word-of-Mouth and Product Sales – An Empirical Investigation of the Movie Industry," *Journal of Retailing* (84:2) pp. 233-242.
- D. M. Fleder, K. Hosanagar. 2008. "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity," *NET Institute Working Paper* No. #07-10. Available at SSRN: <http://ssrn.com/abstract=955984>
- D. H. Gensch. 1987. "A Two-Stage Disaggregate Attribute Choice Model," *Marketing Science* (6:3) pp. 223-239.
- M. Gentzkow. 2007. "Valuing New Goods in a Model With Complementarity: Online Newspapers," *The American Economic Review* (97:3) pp. 713-744.
- W. H. Greene. 2008. *Econometric Analysis*, New York, NY: Prentice Hall.
- A. Ghose, M. D. Smith, R. Telang. 2006. "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," *Information Systems Research* (17:1) pp. 3-19.
- B. Gu, J. Park, P. Konana. 2011. "The Impact of External Word-of-Mouth Sources on Retailer Sales for High Involvement Products," *Information Systems Research*, forthcoming. Available at <http://isr.journal.informs.org/cgi/content/abstract/isre.1100.0343v1>
- G. Häubl, V. Trifts. 2002. "Consumer Decision Making in Online Shopping Environment: The Effects of Interactive Decision Aids," *Marketing Science* (19:1) pp. 4-21.
- J. R. Hauser, B. Wernerfelt. 1990. "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research* (16:3) pp. 393-408.
- W. D. Hoyer, S. P. Brown. 1990. "Effects of Brand Awareness on Choice for A Common, Repeat-Purchase Product," *Journal of Consumer Research* (17:2) pp. 141-148.
- E. J. Johnson, W. W. Moe, P. S. Fader, S. Bellman, G. L. Lohse. 2004. "On the Depth and Dynamics of Online Search Behavior," *Management Science* (50:3) pp. 299-308
- Q. Jones, G. Ravid, S. Rafaeli. 2004. "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration," *Information Systems Research* (15:2) pp. 194-210.
- A. Lee. 2002. "Effects of Implicit Memory on Memory-Based Versus Stimulus-Based Brand Choice," *Journal of Marketing Research* (39:4) pp. 440-454.

- G. Lilien, A. Rangaswamy. 2004. *Marketing Engineering*, 2nd Edition, Bloomington, IN: Trafford Publishing.
- J. G. Lynch, D. Ariely. 2000. "Wine Online: Search Costs Affect Competition on Price, Quality, And Distribution," *Marketing Science* (19:1) pp. 83-103.
- H. W. McMahon. 1980. "TV Loses the 'Name Game' but Wins Big in Personality," *Advertising Age*, December 1, 1980, pp. 54.
- G. A. Miller. 1956. "The Magical Number Seven, Plus Or Minus Two: Some Limits On Our Capacity For Processing Information," *Psychological Review* (63:2) pp. 81-97.
- W. W. Moe. 2003. "Buying, Searching, or Browsing: Differentiating between Online Shoppers Using In-store Navigational Clickstream," *Journal of Consumer Psychology*, (13:1-2), pp. 29-39.
- R. J. Mooney, L. Roy. 2000. "Content-Based Book Recommending Using Learning for Text Categorization." *The Fifth ACM Conference on Digital Libraries*, pp. 195-204.
- P. Nedungadi. 1990. "Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations," *Journal of Consumer Research* (17:12) pp. 263-76.
- G. Ostreicher-Singer, A. Sundararajan. 2006. "Recommendation Networks and The Long Tail of Electronic Commerce," *2nd Annual Statistical Challenges in Electronic Commerce Research Symposium*.
- J. H. Roberts, J. M. Lattin. 1991. "Development and Testing of a Model of Consideration Set Composition." *Journal Marketing Research* (28:11) pp. 429-40.
- E. M. Rogers, R. Agarwala-Rogers. 1975. "Organizational Communication," in *Communication and Behavior*, G. J. Hanneman & W. J. McEwen (eds.), Reading, MA, pp. 218-236.
- A. D. Shocker, M. Ben-Akiva, B. Boccara, P. Nedungadi. 1991. "Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions," *Marketing Letters* (2:3) pp. 181-197.
- H. A. Simon. 1957. *Models of Man, Social and Rational*, New York, NY: John Wiley & Sons.
- M. Smith. 2002. "The Impact Of Shopbots On Electronic Markets," *Journal of Academy of Marketing Science* (30:4) pp. 442-450.
- M. Smith, M., E. Brynjolfsson. 2001. "Customer Decision Making at an Internet Shopbot: Brand Still Matters," *Journal of Industrial Economics* (49:4) pp. 541-558.