

2009

# Goal Attainment on Long Tail Websites: An Information Foraging Approach

James A. McCart

*University of South Florida, jmccart@coba.usf.edu*

Donald J. Berndt

*University of South Florida, dberndt@coba.usf.edu*

Balaji Padmanabhan

*University of South Florida, bpadmana@coba.usf.edu*

Joni L. Jones

*University of South Florida, jjones@coba.usf.edu*

Follow this and additional works at: <http://aisel.aisnet.org/amcis2009>

## Recommended Citation

McCart, James A.; Berndt, Donald J.; Padmanabhan, Balaji; and Jones, Joni L., "Goal Attainment on Long Tail Websites: An Information Foraging Approach" (2009). *AMCIS 2009 Proceedings*. 556.

<http://aisel.aisnet.org/amcis2009/556>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Goal Attainment on Long Tail Websites: An Information Foraging Approach

**James A. McCart**

University of South Florida  
jmccart@coba.usf.edu

**Donald J. Berndt**

University of South Florida  
dberndt@coba.usf.edu

**Balaji Padmanabhan**

University of South Florida  
bpadmana@coba.usf.edu

**Joni L. Jones**

University of South Florida  
jjones@coba.usf.edu

## ABSTRACT

Information foraging theory (IFT) has emerged within the previous decade as a way of explaining the behavior of individuals as they hunt for information (Pirolli, 2007). In IFT, users forage for information using their metaphorical sense of smell which helps guides them through patchy areas of their environment. This preliminary research leverages IFT to build two versions of a clickstream model of information foraging that uses clickstream data to explain goal achievement. The goal being examined is the purchase of a product or submission of a contact form at long tail websites (i.e., sites with limited traffic). The first version of the model uses session-level panel data to examine across-website goal-seeking browsing patterns. Page-level data is used in the second version of the model to reason about browsing patterns within a website. The hypotheses and their related measures are presented for each version of the model.

## Keywords

Clickstream, information foraging, long tail, web behavior, data mining.

## 1. INTRODUCTION

Understanding the browsing behavior of users at websites has been the objective of much of the research employing data about individuals' Web usage (commonly known as "clickstream data"). Especially salient has been the investigation of factors relating to choice behavior, where choice is typically concerned with the purchase of a product (Bucklin et al., 2002). Besides having a general understanding of why users behave the way they do, such knowledge also forms the basis for developing mechanisms to influence choice. For example, to steer a visitor towards a purchase, dynamic on-the-fly changes may be made to a website in terms of its "...pages, link choices, promotional interventions, and prices and product assortments" (Bucklin et al., 2002, pg. 252).

Such a general understanding of factors affecting choice, however, has been difficult to obtain. In part, the difficulty arises because conceptual research focusing on the theories and ideas which provide an explanation of a user's behavior has been limited (Bucklin et al., 2002). This lack of a theoretical base hinders the ability of the results from clickstream research to be reconciled, synthesized, and thus provide a clearer picture of factors affecting choice.

However, finding an appropriate theory to use is challenging in light of the type of data available. Clickstream data provides information on the actions of a user (e.g., what pages were visited, how much time was spent at a site), but nothing else. A person's attitudes, emotions, intentions, and other such concepts are unknown. However, many theories examining an individual's behavior in information systems research rely on such concepts and thus are not appropriate to use. Therefore, a theory is needed which can (1) explain behavior based on a user's action and (2) be appropriately applied to the clickstream domain.

Within the last decade, a theory called Information Foraging Theory (IFT) has emerged which uses a production rule system to explain the searching behavior of individuals as they hunt for information (Pirolli and Card, 1999). The thesis of IFT is that an individual is driven by a metaphorical sense of smell that guides them through patches of information in their environment based on their information goal (i.e., what they are trying to accomplish) (Pirolli, 2007). As they "forage," individuals evaluate whether to continue browsing in their current patch of information or leave to hunt for another one. Central to this theory are the concepts of information patches and information scent. Information patches are areas of the search environment and information scent is what guides a forager to different patches.

The use of IFT in clickstream research requires conceptualizing the ideas of IFT in a non-production rule environment. In essence, this requires utilizing a visitor's actions to infer the cognitive process and thus the reasoning behind the observed behavior. To meet such an end this research will use the concepts of information patches and scent to build a clickstream model of information foraging. The model will rely on measures derived from clickstream data representing IFT concepts to explain goal achievement at "long tail" websites (i.e., sites with limited traffic). Goal achievement is from the perspective of the online firm and consists of something the firm would like to happen at their website (i.e., a choice). This research examines websites where the goal is the purchase of a product or the submission of a contact form.

The term "long tail" refers to a website that resides in the tail of a power law distribution (Anderson, 2006). Figure 1 shows a hypothetical power law distribution illustrating websites and their popularity in terms of the number of visits they received. The head of the curve (darkly shaded portion) represents the most popular websites such as Amazon.com and eBay.com. The long drawn-out tail of the curve (lightly shaded portion) extends to include all other websites. The targeting of a specific niche by long tail websites may explain their lack of traffic. For example, a website for a local medical malpractice law firm is likely only of interest to visitors seeking representation within that same geographical area.

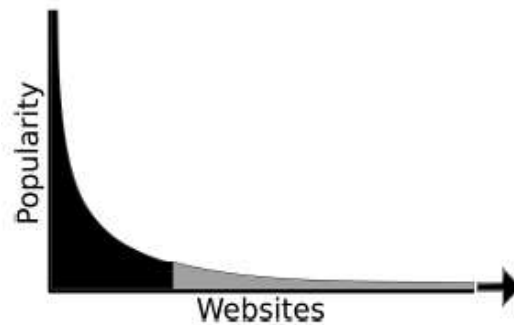


Figure 1. Power Law Distribution

The decision to analyze user behavior at long tail websites is motivated by the ability of IFT to guide analysis. Compared to sites in the head, long tail websites have significantly smaller amounts of data, which is precisely where theory can help guide analysis the most. Lacking theory, analysis would require large amounts of data to work well with commonly used techniques such as data mining. Such an exploratory approach is difficult at long tail websites due to their prohibitively small-sized datasets.

The remainder of this paper is organized as follows. First, an abbreviated literature review of clickstream research is presented in §2. Next, a description of information foraging theory is given in §3. The two versions of the clickstream model are introduced in §4 with the methodology outlined in §5. Finally, a discussion of the potential contributions and conclusion are given in §6.

## 2. LITERATURE REVIEW

Due to space constraints, only a small sampling of prior literature which used clickstream data to examine purchasing behavior is summarized in this section.

Sismeiro and Bucklin (2004) viewed the purchasing process as a series of sequential steps and found that a multi-step model outperformed single-step models in prediction accuracy. Model metrics also differed in effect sign, size, and significance between steps in the purchasing process, indicating some metrics were better predictors at some steps over others. Van den Poel and Buckinx (2005) explored how well different types of metrics predicted a purchase at an online wine seller. A group of detailed clickstream metrics, which were categorized according to the underlying content of the page (e.g., product information, community pages), were the most important predictors found for predicting purchase. Breaking a session down by the sequence of pages visited, Montgomery et al. (2004) looked at how prior path information could predict future paths and ultimately a purchase. Predicting a purchase from a path of one page and six pages viewed resulted in an accuracy of 10% and 21%, respectively.

Moe and Fader (2004) modeled the individual-level dynamic conversion behavior of visitors. The individual-level model contradicted aggregated conversion trends and found over time the overall purchase probability of a visitor decreased, repeat visits had less of an impact on purchasing, and visitor experience raised a person's purchasing threshold. Lastly,

Padmanabhan et al. (2001) determined the implications of using datasets solely from a single site (i.e., site-centric) compared to data capturing every site a user visited (i.e., user-centric) to predict purchases. Not surprisingly models using the user-centric datasets performed better than the site-centric models. More surprisingly, the use of site-centric data was found to lead to erroneous results since significant metrics in site-centric models were found insignificant in user-centric models.

**3. INFORMATION FORAGING THEORY**

Information Foraging Theory (IFT) explains the behavior of individuals as they search for information within an environment such as the Web (Pirolli, 2007). Central to this theory are the concepts of information patches and information scent. Information patches are distinct areas of the search environment which differ in their informational content. Information scent is the driving force behind why a person makes a navigational selection amongst a group of competing options. As foragers are assumed to be rational, scent is a mechanism by which foragers’ reduce their search costs by increasing their accuracy on which option leads to the information of value (Pirolli, 2007). This synthesis of behavior (e.g., information scent) and environment (e.g., information patches) provides for a rich theory of information foraging.

Prior research has used IFT to not only look at navigational patterns of foragers, but also how the information environment can be altered to facilitate foraging. IFT has been used to inform the design of graphical user interface controls (e.g., checkboxes, list boxes) which provide social activity visualizations as navigational cues (Willett et al., 2007); help interpret the effects of delay, familiarity, and breadth on users’ performance, attitude, and intentions at websites (Galletta et al., 2006); and analyze the role of scent in the decision to browse a menu as opposed to searching a website (Katz and Byrne, 2003).

IFT itself builds on more established theories such as Optimal Foraging Theory (OFT) (Stephens and Krebs, 1986) and the Adaptive Control of Thought-Rational Theory (ACT-R) (Anderson et al., 2004). OFT is an ecological theory concerned with explaining the foraging behavior of animals as they hunt for food. OFT assumes each animal goes through a search–encounter–decision process as they forage, with the goal being to maximize net energy gained. To maximize energy, the animal is faced with the decision of which prey to eat or how long to forage in a patch. OFT is used to explain the behavioral elements of people foraging for information.

ACT-R is a psychological theory of the human mind that includes the cognitive architecture and process by which cognition works. ACT-R is used to explain at a cognitive level why actions are performed. IFT uses a production rule system from ACT-R to determine probabilistically which action is selected based on its utility within the context of a user’s current information goal. For example, an action to click on a hyperlink may be chosen over backing up to a previously visited page because following the hyperlink may be more likely to lead to the information being sought.

Figures 2a and 2b show examples of production rules (Pirolli, 2007, pg. 97), which follow the form IF <condition(s)> THEN <action>. In situations with multiple production rules fulfilling their conditions, conflict resolution is undertaken where a rule is probabilistically chosen based on its utility (Equation 1) (Anderson et al., 2004). The utility of a production is based on its prior probability of success and prior cost spent when achieving a goal ( $P_i$  and  $C_i$ ), the expected gain from completing the goal ( $G$ ), and random noise ( $\epsilon$ ) (Anderson et al., 2004).

IF goal is Process-link  
 & there is a task description  
 & there is a browser  
 & there is an evaluated link  
 & the link has highest  
 activation  
 THEN click on the link

IF goal is Process-link  
 & there is a task description  
 & there is a browser  
 & there is an evaluated link  
 & the mean activation on page is  
 low  
 THEN click on the link

$$U_i = P_i G - C_i + \epsilon$$

**Figure 2a. Click-link Production Rule**

**Figure 2b. Process-link Production Rule**

**Equation 1. Production Utility**

**Information Patches**

An information patch is a grouping of similar information (Pirolli, 2007). A site-patch refers to an entire website as a patch, whereas a page-patch represents a webpage as a patch. Regardless of what a patch represents, a user will continue to forage within that patch until the “expected potential of that patch is less than the mean expected value of going to a new patch” (Pirolli, 2007, pg. 81). Stated mathematically, the patch-leaving rule (Charnov, 1976) is to forage in a patch as long as  $U(x) > \bar{U}$ , where  $U(x)$  is the utility of a forager in their current state and  $\bar{U}$  is the mean utility of other patches (Pirolli, 2007).

## Information Scent

Information scent is the use of cues obtained from the text and images associated with a hyperlink to provide information about distal content (Pirulli, 2007). As shown in the production rule in Figure 2a, links are followed based on their scent (i.e., utility) with respect to a user's current goal. The utility of link  $L$  given goal  $G$  is determined based on the sum activation of all features of the goal (i.e., words) plus random noise (Equation 2) (Anderson et al., 2004).  $A_i$  is the activation of each of the  $j$  cues of a link (i.e., words) with respect to goal feature  $i$ . Activation is determined by the base activation of the goal ( $B_i$ ); amount of attention paid to each link cue ( $W_j$ ); similarity between the link cue and feature of a goal ( $S_{ji}$ ); and random noise  $\varepsilon$  (Equation 3) (Anderson et al., 2004; Pirulli, 2007). In this model, an individual on a Web page may choose to follow the link with the strongest scent, operationalized as the one with the highest activation.

$$U_{LG} = \sum_{i \in G} A_i + \varepsilon$$

Equation 2. Link Utility

$$A_i = B_i + \sum_j W_j S_{ji} + \varepsilon$$

Equation 3. Goal Feature Activation

## 4. CLICKSTREAM MODEL OF INFORMATION FORAGING

To represent the concepts of information scent and patches using clickstream data, two versions of a clickstream model of information foraging are proposed. The user-centric (UC) model exploits user-centric data (Padmanabhan et al., 2001) about a forager's entire browsing behavior to explain goal achievement at a long tail website. This model compares a forager's behavior across multiple websites. However, due to user-centric data typically being aggregated at the session level, the model lacks depth at individual websites.

Since data about a user's entire clickstream over multiple sites is rarely available to an online firm, a site-centric (SC) version of the model employing site-centric data (Padmanabhan et al., 2001) is also developed. Page-level data makes the site-centric model capable of analyzing patches at all levels of analysis along with information scent at a website. However, since a forager's behavior across sites is unknown with site-centric data, the site-centric model compares a forager's behavior relative to users who had previously achieved a goal at that website (i.e., goal sessions).

The relative nature of comparisons is an important aspect to both models because the information goal of a forager may be complex. In such a situation comparing behavior against an absolute such as having no scent may fail to uncover goal-seeking behavior. For example, a user who visits the same page multiple times may appear to have low scent since the forager has increased their search cost with each repeated page visit. However, if that forager's scent were compared to what other users have exhibited at the same site, the scent may be relatively high.

### User-centric Clickstream Model of Information Foraging

The user-centric model contains four hypotheses which examine how browsing behavior can lead to goal achievement by considering the website as a patch (i.e., site-patch) and judging its value relative to other patches. Table 1 presents each of the hypotheses with the rationale provided below.

Hyp. #	Hypothesis
SITE-PATCH	
UC 1	Higher total duration spent at this site-patch relative to other site-patches within a user-session will be <i>positively associated</i> with achieving a goal on this long tail website.
UC 2	Higher number of pages viewed at this site-patch relative to other site-patches within a user-session will be <i>positively associated</i> with achieving a goal on this long tail website.
UC 3a	Returning to this site-patch during the <i>same user-session</i> will be <i>positively associated</i> with achieving a goal on this long tail website.
UC 3b	Returning to this site-patch during a <i>different user-session</i> will be <i>positively associated</i> with achieving a goal on this long tail website.

Table 1. User-centric Hypotheses

The first hypothesis recognizes that a forager has imperfect information and limited computational facilities. Therefore, a satisficing strategy (Reader and Payne, 2007) is employed such that a forager will continue to browse as long as information of value is being obtained (Pirolli, 2007). The second hypothesis notes that each page visited represents a conscious decision point where the user believes the value of continuing to browse at this site-patch is higher than what they expect to find elsewhere.

While foraging within a site-patch, a user forms a general opinion of the value of the website. When leaving one site-patch for another, a forager believes greater value may be found elsewhere. However, if a user returns after leaving, the forager was unable to find a more valuable site-patch. Therefore, the site-patch of interest is more likely than other site-patches to contain the information necessary to fulfill the user's goal, which leads to Hypotheses UC 3a and 3b.

### Site-centric Clickstream Model of Information Foraging

There are two main differences between the user-centric and site-centric models. First, since the site-centric model has no knowledge of browsing behavior at other websites, comparisons are made relative to users who achieved a goal at the site of interest. Deviations from known goal-achieving browsing behavior are assumed to indicate less scent or patch value and thus a lower probability of a goal being achieved.

Second, the ability to determine if a forager left the site and returned during the same session cannot be determined directly from the user's clickstream. However, the site-centric model can make use of the referring field (which many site-centric datasets have access to) which shows which URL a user came from. With those differences in mind, the hypotheses for the site-centric model are shown in Table 2 and described in more detail below.

Hyp. #	Hypothesis
SITE-PATCH	
SC 1	The <i>closer the total duration</i> spent relative to goal-achieving sessions at this site-patch, the more <i>positively associated</i> a session will be with achieving a goal on this long tail website.
SC 2	The <i>closer the number of pages</i> viewed relative to goal-achieving sessions at this site-patch, the more <i>positively associated</i> a session will be with achieving a goal on this long tail website.
SC 3a	<i>Returning</i> to this site-patch during the <i>same session</i> will be <i>positively associated</i> with achieving a goal on this long tail website.
SC 3b	<i>Returning</i> to this site-patch during a <i>different session</i> will be <i>positively associated</i> with achieving a goal on this long tail website.
VALUABLE-PATCHES	
SC 4	The <i>closer the visit incidence of goal patches</i> compared to non-goal patches relative to goal-achieving sessions at this site-patch, the more <i>positively associated</i> a session will be with achieving a goal at this long tail website.
SC 5	The <i>closer the average durations spent in goal patches</i> relative to goal-achieving sessions at this site-patch, the more <i>positively associated</i> a session will be with achieving a goal at this long tail website.
STRICT INFORMATION SCENT	
SC 6a	The <i>closer the proportion of repeatedly visited pages</i> is relative to goal-achieving sessions at this site-patch, the more <i>positively associated</i> a session will be with achieving a goal on this long tail website.
SC 6b	The <i>closer session complexity</i> , in terms of clickstream linearity, is relative to goal-achieving sessions at this site-patch, the more <i>positively associated</i> a session will be with achieving a goal on this long tail website.
RELAXED INFORMATION SCENT	
SC 7	The <i>closer the following incidence of goal scent trails</i> compared to non-goal scent trails relative to goal-achieving sessions at this site-patch, the more <i>positively associated</i> a session will be with achieving a goal on this long tail website.

Table 2. Site-centric Hypotheses

#### Information Patches

Hypotheses SC 1 and 2 are restated in the site-centric model to be compared relative to goal-achieving sessions at the website of interest. Hypotheses SC 3a and 3b are unchanged from the user-centric model.

Certain page-patches on a website may provide more useful information to a user than others. Grouped together, these pages represent a valuable patch (McCart et al., 2008), where each patch discovered may provide insight about the behavior of visitors to that website. When a valuable patch is predominately visited by goal-achieving foragers, it is known as a *goal patch*. When non-goal foragers are the majority visitors, then the valuable patch is a *non-goal patch*. Foragers who visit similar valuable goal and non-goal patches as goal sessions are likely to have related information goals. As the information goals are comparable and the past sessions resulted in a goal being achieved, this leads to Hypothesis SC 4.

Visitation of goal patches by itself does not necessarily indicate a forager obtained the same value from a patch as the goal sessions did. For example, a relatively short amount of time spent in a patch may indicate the user did not fully recognize the value of the patch due to divergent information goals. Thus, foragers who spend roughly the same amount of time as goal sessions did in goal patches are more likely to have similar information goals and thus achieve a goal, which leads to Hypothesis SC 5.

#### Information Scent

The remaining three hypotheses deal with information scent. The first two view information scent from a user's entire session. In addition, they both assume a strict viewpoint of scent where inefficiencies in a user's clickstream (e.g., backtracking) are indications of poor scent. The final hypothesis examines scent among different fragments of a user's session. In addition, a more relaxed characterization of scent is used which recognizes that complex sessions may still be of high scent even in the presence of some inefficiencies.

Hypothesis SC 6a assumes high scent is characterized by a lack of repeat visitation on webpages. Hypothesis SC 6b goes a step further by examining the linearity of a user's clickstream where the distinction between what pages are repeatedly visited can make a difference in determining scent.

The final hypothesis recognizes that a forager's information goal may change during a session. Thus a repeat visitation of a page may not be considered inefficient when viewed through the lens of a different information goal. Similar to the idea of valuable patch (McCart et al., 2008), certain fragments of a path through a website may be of more use to foragers than others. Foragers who follow similar fragments of these valuable paths as goal sessions are assumed to have comparable information goals and thus are more likely to achieve a goal, leading to Hypothesis SC 7.

## 5. METHODOLOGY

Tables 3 and 4 below present the measures to be used to test each of the hypotheses in the user-centric and site-centric model, respectively. In each table the hypothesis number is presented in the left-hand column with the measure in the right-hand column. The tables also contain the dependent variable (DV) and any control variables (CTRL) that will be used. Each model will be tested using logistic regression. The specifics for each model are provided in the following subsections.

### User-centric Model of Information Foraging

The user-centric model will be tested using panel data over a one year period, where the goal being examined is the purchase of a product. The sample will consist of the behavior of a forager at a long tail e-commerce website relative to what occurs at other websites of a user-session. Long tail e-commerce websites are those sites that are the 25% least popular e-commerce websites in the dataset. A user-session consists of a session at a long tail e-commerce site plus any additional sessions at other websites visited by the same user within 30 minutes of the beginning or end of the target session.

Hyp. #	Measure
UC 1	Relative total duration
UC 2	Relative number of pages
UC 3a	If another session was active during the target session
UC 3b	If user had visited the target session in the past
DV	Goal achieved during session at target site
CTRL	Website focused on
CTRL	Number of other sessions in user-session besides the target session

Table 3. User-centric Measures

### Site-centric Model of Information Foraging

The site-centric model will be tested using data over a one year period from a web hosting company, where the goal being examined is the submission of a contact form. Since a forager's behavior is compared relative to prior sessions at the website, only sites with a minimum of 20, 30, 40, or 50 goal and non-goal sessions will be included in the sample (a sensitivity analysis will be done). Since many of the measures are compared against previous sessions and the websites being examined are long-tail, the sparseness of data presents a challenge. Therefore, the measures for a forager will be calculated in a progressive manner. For example, the measures for a session occurring 30 days into the dataset will be calculated relative to all prior sessions at that point in time. A session at 60 days, will instead have its measures calculated relative to all 60 days worth of data.

Valuable patches are learned separately for each website following the methodology outlined in McCart et al. (2008). Goal and non-goal sessions for a website are separated into two datasets and then mined for frequent itemsets (i.e., potential patches). A patch becomes valuable if there is a significant difference between the proportion of goal and non-goal sessions visiting the discovered itemset. A similar methodology using sequential patterns instead of frequent itemsets is used for discovering valuable trails.

Many of the measure from Table 4 are relatively straightforward; however, three of the measures require further explanation. Patch visitation (SC 4) represents how many goal and non-goal patches were visited by a forager (weighted by their value in distinguishing between goal and non-goal sessions), relative to other goal sessions. Trail following (SC 7) is calculated in a similar manner as patch visitation, except valuable trails are examined instead of patches. Finally, clickstream linearity (SC 6b) measures how straight a session's path is (i.e., an absence of revisited webpages). Linearity, as determined from the path stratum metric, is calculated based on concepts from graph theory (McEneaney, 2002). Clickstream linearity is also calculated relative to prior goal sessions.

Hyp. #	Measure
SC 1	Relative total duration
SC 2	Relative number of pages
SC 3a	Referring URL from another domain
SC 3b	If user had visited the target session in the past
SC 4	Relative patch visitation (PV)
SC 5	Relative average durations spent in valuable goal patches
SC 6a	Relative percentage of unique pages viewed
SC 6b	Relative clickstream linearity (path stratum metric (McEneaney, 2002))
SC 7	Relative trail following (TF)
DV	Goal achieved during session
CTRL	Website visited

**Table 4. Site-centric Measures**

## 6. CONCLUSION

This research introduced a model which conceptualized information patches and scent in a non-production rule environment from readily available clickstream data over multiple long tail websites. In addition, two versions of the model were given for user-centric and site-centric datasets which leverages the strengths and recognizes the weaknesses of both types of data. Due to the presence of IFT guiding the analysis, long tail websites were able to be focused on even in light of their sparse datasets.

The next step for this research is to calculate the measures outlined in Tables 3 and 4 and then test each version of the model. Future research will test the model in other contexts, such as websites in the short-head of the power law distribution, to determine the model's generalizability. It is the hope that this early research will eventually lead to a model, or set of models, which theoretically explains goal-seeking behavior across a wide variety of websites. Future research will also examine how other theories, such as Social Exchange Theory (Blau, 1986), can be used to extend IFT by considering the cost-benefit analysis a visitor performs while browsing and their sunk information foraging costs.



## ACKNOWLEDGMENTS

We wish to thank the anonymous reviewers for their valuable suggestions to this paper.

## REFERENCES

1. Anderson, C. (2006) *The long tail: Why the future of business is selling less of more*, Hyperion, New York, NY.
2. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., and Qin y. (2004) An integrated theory of the mind, *Psychological Review*, 111, 4, 1036-1060.
3. Blau, P.M. (1986) *Exchange and power in social life*, Transaction Publishers, New Brunswick, NJ.
4. Bucklin, R.E., Lattin, J.M., Anwari, A., Gupta, S., Bell, D., Coupey, E., Little, J.D., Mela, C., Montgomery, A., and Steckel, J. (2002) Choice and the Internet: From clickstream to research stream, *Marketing Letters*, 13, 3, 245-258.
5. Charnov, E.L. (1976) Optimal foraging, the marginal value theorem, *Theoretical Population Biology*, 9, 2, 129-136.
6. Galletta, D., Henry, R., McCoy, S., and Polak, P. (2006) When the wait isn't so bad: The interacting effects of website delay, familiarity, and breadth, *Information Systems Research*, 17, 1, 20-37.
7. Katz, M.A. and Byren, M.D. (2003) Effects of scent and breadth on use of site-specific search on e-commerce web sites, *ACM Transactions on Computer-Human Interaction*, 10, 3, 198-220.
8. McCart, J.A., Jones, J.L., Padmanabhan, B., and Berndt, D.J. (2008) Identifying valuable information patches on long tail sites: An information foraging perspective, *The Eighteenth Annual Workshop on Information Technology and Systems (WITS)*, December 13-14, Paris, France.
9. Moe, W.W. and Fader, P.S. (2004) Dynamic conversion behavior at e-commerce sites, *Management Science*, 50, 3, 326-335.
10. Montgomery, A.L., Li, S., Srinivasan, K., and Liechty, J.C. (2004) Modeling online browsing and path analysis using clickstream data, *Marketing Science*, 23, 4, 579-595.
11. Padmanabhan, B., Zheng, Z., and Kimbrough, S.O. (2001) Personalization from incomplete data: What you don't know can hurt, *The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 26-29, San Francisco, CA, 154-163.
12. Pirolli, P.L. (2007) *Information foraging theory: Adaptive interaction with information*, Oxford University Press, New York, NY.
13. Pirolli, P.L. and Card, S. (1999) Information foraging, *Psychological Review*, 106, 4, 643-675.
14. Reader, W.R. and Payne, S.J. (2007) Allocating time across multiple texts: Sampling and satisficing, *Human Computer Interaction*, 22, 3, 263-298.
15. Sismeiro, C. and Bucklin, R.E. (2004) Modeling purchase behavior at an e-commerce web site: A task-completion approach, *Journal of Marketing Research*, 41, 3, 306-323.
16. Stephens, D.W. and Krebs, J.R. (1986) *Foraging theory*, Princeton University Press, Princeton, NJ.
17. Van den Poel, D. and Buckinx, W. (2005) Predicting online-purchasing behaviour, *European Journal of Operational Research*, 166, 2, 557-575.
18. Willett, W., Heer, J., and Agrawala, M. (2007) Scented widget: Improving navigation cues with embedded visualizations, *IEEE Transactions on Visualization and Computer Graphics*, 13, 6, 1129-1136.