Available online at www.sciencedirect.com
SCIENCE DIRECT®
Electronic Commerce Research and Applications 4 (2005) 49–65
Electronic Commerce Research and Applications
www.elsevier.com/locate/ecra

Drivers of Web portal use

Rahul Telang *, Tridas Mukhopadhyay

2107D Hamburg Hall, Carnegie Mellon University, Pittsburgh, PA 15217, USA

Received 20 September 2004; received in revised form 5 October 2004; accepted 11 October 2004
Available online 28 October 2004

Abstract

In this paper, we investigate three complementary measures of portal use: frequency of use, length of visits and repeat use. To examine these three measures of use, we first classify the services provided by portals into three categories: search, information, and personal services. We argue that these three different functions affect portal use in different ways. We primarily rely on the human computer interaction literature to develop our model of portal use. Our analysis is based on Internet navigation data of 102 demographically diverse users over a period of one year for six major portals. In total, we study 6321 distinct portal choices. Our results show strong repeat use for personal services followed by information services and search function. Our findings show that cumulative dissatisfaction with search results has a negative effect on future user choice decisions. Both information and personal services tend to extend the length of portal visits. As expected, search services tend to reduce the time spent as users move on to the search targets. But we also find that search function availability drives more traffic to portals than information or personal services. Of the three services, personal services use shows maximum (week-to-week) stability, information services use, on the other hand, shows least stability. Use of personal services leads to use of search and information services. We also find that demographic characteristics play some role in portal use.

1. Introduction

Internet portals often act as gatekeepers to the Internet. Users may begin their sessions on the Internet by visiting a portal, and obtain information like news, weather or stock quotes. They may move on to browse products, gather information or even make purchases only after the Web sites of interest have been located through the search process [14]. Portals also provide many personal communication services in the form of emails, message boards, etc. Moreover, most of these services are offered to users free-of-cost. It is no surprise then that portals are some of the most visited sites on the Internet [16].

* Corresponding author. Tel.: +412 268 1155.
E-mail addresses: rtelang@andrew.cmu.edu (R. Telang), tridas@andrew.cmu.edu (T. Mukhopadhyay).

1567-4223/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.elerap.2004.10.004

Driving traffic to their sites and making users stay for longer periods are important for portal firms because Internet-based advertising is their main source of revenue. [1] Generating repeat use of their services is also critical because it allows them to know their customers better and improve their software design, update their information and service offerings, etc. A repeat customer is worth more to Yahoo! than a new one because repeated interactions with the portal provide Yahoo! with a rich set of information about preferences and purchase patterns at the individual level. Individual-level data can then be used by Yahoo! directly or sold (or shared) with other companies to tailor product offerings and even pricing plans to identifiable and targetable consumers. It may also allow a portal to convert the customer into using several value-added subscription based services.

It is therefore safe to assume that it is a common goal for Web portals to develop a loyal user base that visits the site frequently, and spend sufficient time per visit. Do they succeed in doing so? What are the drivers of Web portal loyalty? What are the determinants of the frequency or length of portal visits? What roles do search, information or personal services play in user behavior? Does user behavior vary with demographic characteristics? Answers to these questions will not only enhance our knowledge of portal use, but also help understand the development and maintenance of this important class of software products.

While the portals are important players on the Internet, there is not much scientific research on different facets of their use. [2] We believe there are two hurdles in conducting such research. First, we have to rely on multiple disciplines to develop testable hypotheses. For example, we must consider the cognitive psychology and human computer interaction literature to understand the use process. We have to borrow from the marketing literature to analyze user loyalty to the portals. Second, it is hard to obtain reliable and rich data to empirically investigate the varied issues involved in this research. In particular, we not only need detailed use data of specific services (search, information, email, etc.) at the individual user level for an extended period of time (to track frequency and loyalty of use), but we also require sensitive user data to gauge the effects of demographic differences.

In this paper, we investigate three complementary measures of portal use: *frequency of use, length of visits* and *repeat use*. To examine these three measures of use, we first classify the services provided by portals into three categories: *search services, information services*, and *personal services*. We argue that these three different functions affect portal use in different ways. We also investigate the impact of demographic characteristics and control for exposure to banner ads that may lead a user to a particular portal. Finally, we consider user dissatisfaction with portal use. In particular, we observe whether users go to a portal to search for some information, and switch to another portal to continue the same search. We posit that the cumulative dissatisfaction of this nature weakens user loyalty.

Our analysis is based on the Internet navigation data of 102 demographically diverse users over a period of one year for six major portals. In total, we study 6,321 distinct portal choices. Our results show strong repeat use for personal services, followed by information services and the search function. We also find that Yahoo! and Excite have the strongest loyal base amongst the portals in our study. Our findings show that dissatisfaction with search results has a negative effect on future user choice decisions. Both information and personal services tend to extend the length of portal visits. We find that while the use of search services reduces the time spent by users on portals, search services are the key drivers of portal traffic. Of the three services, personal services use shows maximum (week-to-week) stability. Information services use, on the other hand, shows the least stability. Use of personal service also induces the use of the information and search services.

[1] The on-line ad revenues were close to US$6.5 billion in 2002 (Business Week, What's New in Online Ads: Improvement, October 31, 2002) and is expected to grow. Yahoo! earns close to 65% of its total revenues (almost US$1 billion) from ads alone.

[2] The research related to portal use has been generally confined to either technical domains or analytical models [3,17].

Demographic characteristics play some role in portal use. For example, high-income users tend to be more loyal, spend less time per visit, and their use of the search function is more stable. Gender differences show up in information and search services use; men tend to use more of these services compared to women. The use of portal services by adults and kids exhibits no difference. Finally, race differences also occur in portal use. While white users tend to be more loyal, they tend to use information and search functions less. But demographics have little impact on the frequency of personal services use.

The rest of this paper is organized as follows. In the next section, we discuss the prior research. In Section 3, we present the conceptual model underlying our analysis. Section 4 outlines the model estimation methods, while Section 5 describes the data collection procedures. Section 6 details the results of our estimation. Finally, Section 7 concludes our paper with some implications of our research and points to directions for future research.

2. Prior research

We rely on the cognitive psychology and human computer interaction (HCI) literature along with marketing literature to understand the drivers of Web portal use. People may repeatedly use a specific portal because of knowledge gained about the portal through usage experience. Users must spend time in learning to use a variety of search and other value-added features that impose switching costs. The literature in HCI and cognitive psychology has shown that there is considerable exploratory learning involved in using computer interfaces [6]. Users often develop mental models about the system and then attempt to apply these models to increase task efficiency [22]. For example, anecdotal evidence suggests that as people get used to a general "look and feel" on the Web, comprehension time reduces, and the ability to process information substantially increases [2]. Repeated interactions with the portal can lead to a comfort level with it and induce repeat use.

The marketing literature has also demonstrated that brand loyalty induces repeat use. A common explanation for brand loyalty is "inertia". Shugan [21] conjectures that this type of routinization reduces the "cost of thinking" and leads to repeat purchases. Another related explanation concerns consumer uncertainty about other brands. Economic costs arise from the possibility that the newly-purchased brand is not of expected quality, and therefore not worth the price paid for the item. Potential psychological costs include the discomfort associated with using an item that does not perform up to expectations (e.g., eating a poor tasting cereal).

However, there are two reasons that can weaken this switching cost. First, learning can be transferable, thus making it easy for the user to switch. Second, the search results may not be satisfactory or incomprehensive, leading the user to switch to other portals.

Portal companies also attempt to lock in users by adding new features to their products to enhance not only repeat use, but also the frequency and length of use. Yahoo!, for one, spends close 15% of its annual budget on product development [1]. Even Google, after years of sticking to a simple search offering, is moving to offer free email. Offering value-added features in software is not a new idea. Nault and Dexter [19] show how value-added features enable the software firms to charge a premium. Brynjolfsson et al. [5] study the spreadsheet features by employing hedonic price analysis. Portals typically include value-added features such as personalized news, email, and chat rooms. Studying the effects of portal features is particularly interesting because we believe that different features impose different switching costs, and affect the type of use in different ways.

Finally, we briefly note prior work on computer and Internet use. Kraut et al. [14] reported that demographic characteristics such as age, gender and race play an important role in Internet use. In another study, Kraut et al. [15] found that females used email more than males, but they used the Web less than males. An important public policy debate centers on the role of race and income in Internet access and use [11,12]. Given the importance of demographic characteristics, we include them in our models of portal use.

3. Model foundations

The Web, for the first time, enabled companies to develop software to be made available to the masses at no cost. Foremost among this new class of software is Web portals. A fundamental research question for the electronic commerce domain concerns the efficacy of both the process and outcome of software development. Our interest in this research involves the outcome measurement. Indeed, software outcome can be measured analogously in various ways such as extent of use, user satisfaction, business process improvement, etc. [8]. In the context of Web portals, we specifically focus on the first metric: extent of use. We demonstrate that a meaningful study of portal use should meet three conditions. First, a single metric of use is not sufficient. Second, portals do not provide a homogenous set of services; rather, their services belong to multiple classes. Third, we must use multiple literatures to understand the use of different classes of services available at Web Portals.

3.1. Measures of portal use

We recognize that the success of complex software products like Web portals ultimately depends on their ability to generate more revenue than costs. It is quite clear that portal use constitutes a necessary condition for this success. Our focus on portal use thus constitutes an important research question on its own merit. However, the selection of metrics of use in this context should take into account the dominant business model of the Web portals. Given that most portal services are given away for free, and users have access to multiple providers, we believe that a successful portal would require a loyal base of users who repeatedly come back to the same portal on a frequent basis for extended periods of use. We, therefore, use three complimentary measures of portal use that mirror the business goals of Web portals:

3.1.1. Repeat use

No portal can be financially viable without a significant fraction of its users coming back to the site repeatedly. Repeat users not only reduce the marketing costs of user acquisition, they also enable the company to understand user behavior, and accordingly change the offerings, and discover new service or business opportunities. However, *repeat use* or loyalty to a Web portal is not a certainty. [3] Users have many choices. They typically do not have to pay for portal services, nor do they incur much switching costs in going to a different portal. In this context, how much repeat use do we actually see for portal use?

3.1.2. Stickiness

Portal site operators should not simply be satisfied with frequent repeat users; they also want the users to spend more time per visit. We use the length of use or *stickiness* to measure this behavior. Ironically, while some portal services may increase stickiness, others may do just the opposite. Thus, for example, effective search services may actually reduce stickiness, but improve frequency of use.

3.1.3. Frequency of use

Repeat use in itself does not spell portal success; users must return to the site frequently to increase the viability of the portal business model. Why will users come to the portal site frequently? Will they come back for search, information or email? Will the use of one type of service (e.g., search) induce the user to try another service (e.g., information)? Will different user types (male vs. female, for example) exhibit different propensity to return to the site frequently?

3.2. Classification of portal services

We have made repeated references to the different types of services provided by portal sites. In classifying portal services we make use of two criteria. First, the development and maintenance of

[3] We will continue to use the term loyalty and repeat use interchangeably.

different service types should require different methods, and thus lead to different cost structures. Second, users derive different utilities from the services, and therefore exhibit distinct usage patterns. In addition, the propensity for switching may also vary across the services. For example, while switching may be relatively easy for search tasks, the use of information and personal services may raise barriers to switching. We divide portal services into three categories.

3.2.1. Personal services

These features generally require registration via entry of a username and password to access the services. They also let users customize their interactions with the site. All the portals we examined offer personalized services for emails, chat rooms, bulletin boards, messaging services, and personalized home pages, etc. In addition to meeting the aforementioned conditions, all of these features have interfaces of their own. For example, when a user accesses Yahoo! mail, she gets a completely different interface from the original portal interface.

3.2.2. Information services

As the name suggests, portals provide information services to users through these features. News, weather, and sports are some of the examples of information services. Generally, users can access these services directly from the portal, without entering any username and password. For example, users can often access current news by clicking the appropriate link on the portal site.

3.2.3. Search services

The emergence of the Web as the favorite destination for users to find information for practically anything has made the search function an indispensable tool for portals to lure users to their sites. The sheer size of the Web and its rapid growth make the search function a daunting task. At the same time, users demand quality results. To satisfy user requirements, portals follow a variety of strategies. While some provide homegrown solutions, others rely on outside solutions.

Next, we present our conceptual model of portal use in Fig. 1. This model has three components. The first model examines the effect of repeat use, demographics and other variables on portal choice, and is discussed in Section 3.1. The second model on stickiness concerns the length of user visits, and is described in Section 3.2. The last model examines the frequency of portal use, and is explained in Section 3.3. Each of these models is estimated at the aggregate portal level and disaggregated task level.

3.3. Model of repeat use [4]

One might argue that users need not show repeat use to any portal because they can easily switch to a competing portal without incurring any costs. In popular terms, another portal is only a mouse click away. However, as we saw in the previous section, there are many reasons for the users to keep choosing same portal including inertia, habit formation, and observable cumulative experience with the portal and switching costs. A positive experience with a portal might lead to repeated use even though other portals are available. To understand the repeat use of portals, we must consider two main drivers – *user learning* and *user satisfaction*.

The GOMS (goals, operators, methods and selection rules) Model [18] has been widely applied to quantify learning and transfer of learning. According to this model, a task can be decomposed into perceptual, cognitive, and motor activities. Perceptual activities include reading information and locating an icon; cognitive activities include remembering mapping instructions and icons, formulating goals, and developing an action plan. Motor activities include moving the mouse and clicking. Based on this theory, similar tasks make learning easily transferable and have lower switching costs. For the purpose of simple search tasks, all portals are quite similar in their

[4] Note that the measure *Repeat Use* is first created from the choice of portals by an individual. An individual selects one of the six portals at every visit. It is then regressed on the future choices to test whether it plays a significant role. *Repeat Use* therefore, appears as an independent variable in Fig. 1. Thus, the model of *Repeat Use*, in principle, is a multinomial choice model.

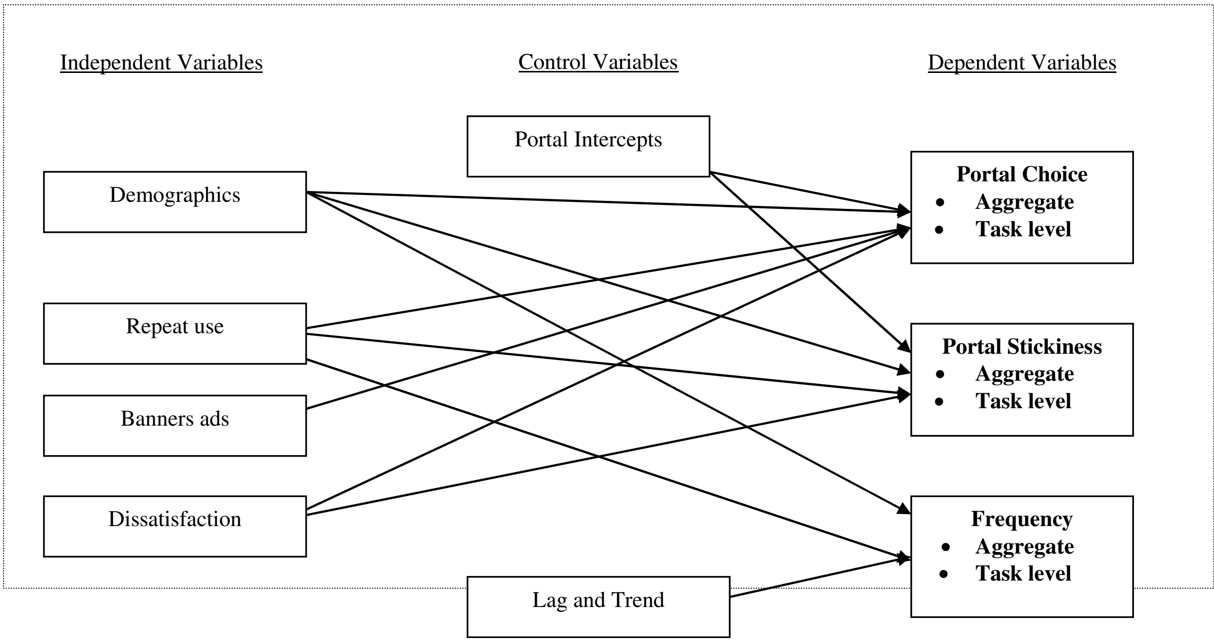

Fig. 1. A model of portal choice and stickiness.

designs. The search feature (which involves a text box and a search button) is easily visible, and easy to locate on the popular portals. All require the user to enter the search term(s) and click the "search" button. Clearly, this task is quite similar across portals in terms of perceptual, cognitive, and motor activities.

While learning information services features is cognitively simple as it merely involves looking up the feature and clicking it, it is perceptually difficult to transfer. Size, location, color, and presentation have significant impact on the perceptual ability of the user [7]. Similarly, time to identify and locate information is reduced significantly with familiarity [13]. Since portals display a great deal of information, it is difficult to locate the desired feature on different portals, if not to know whether they offer the feature at all! With repeated use, users not only become efficient, but also develop reasonable expectations about the results. They may find it easier to use the familiar features, which may lead to higher switching costs and hence more repeat use.

Finally, the personal services require the users to learn many rules (high cognitive load). For example, to use email, not only does a user have to locate and click it, but she also has to go to a different interface, and learn many rules about composing and sending the message. If the user *switches* to a different portal, she will have to learn many *new* rules. There is also a significant set-up cost when signing up for these features in terms of selecting an id, password, and enter some personal information. In addition, switching to a different email would also require the user to communicate the new id to her acquaintances. In sum, the switching cost for personal services is much higher. We, therefore, expect that personalized features will lead to higher repeat use. Thus, we hypothesize:

H1. (*Personal service hypothesis*). We expect repeat use will be a significant predictor of future choice and personal service use will generate the strongest repeat use.

While learning imposes switching cost, desire for comprehensive and complete search results may encourage users to switch to other portals. It has been well documented that search engines run the queries on a local database that contains only a fraction of documents available on the Web. Moreover, the engines run different algorithms for ranking the documents leading to vastly different results for similar search queries [4]. Therefore, users may not be totally satisfied with the results of one engine and may switch to another engine immediately to start the search process anew. Unlike other traditional goods, our server records of portal use can detect this phenomenon of immediate switching with a high degree of confidence. Whenever a user switches to another portal and continues the process with a similar search term, it suggests that the user was not entirely satisfied with the previous results. We argue that this user experience may lower her quality perception of the portal, and adversely affect the future user of the portal.

Choice sequence in Fig. 2 explains the concept of *cumulative experience*. The user has chosen seven portals in four search sessions. In the first search session, the user visits Yahoo!, performs a search and quits. Next time, she again starts her search session with Yahoo!, but switches to Excite to continue the same search. We conjecture that this immediate switching occurs due to unsatisfactory results from Yahoo! Here, Excite represents "intra-session switching" as it occurs during the same search session. The user visits Lycos for her third search session and finally, in the fourth search session, she starts with Excite and switches to Altavista to continue the same search. She is dissatisfied again with the Altavista results and switches to HotBot to continue the search (another example of "intra-session switching").

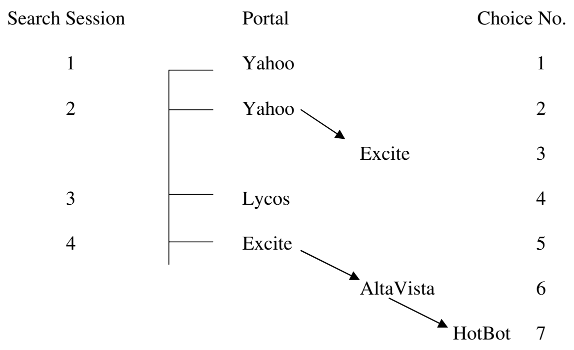

Fig. 2. Choice sequence for a user.

We argue that intra-session switching reflects the user's reaction to the quality of search results and affects her future use of the portal. We use the term *cumulative negative experience* (CNE) to capture the effect of dissatisfaction on future choices. This is the proportion of times the user made an intra-session switch away from an engine to the total number of times she used the same engine. In Fig. 2, the user chose Yahoo! twice (choice 1 and 2) and switched away once (choice 3). Therefore for a future choice (e.g., choice 4), CNE is equal to half for Yahoo!. It is clear that CNE for Yahoo! may change over time. We find that such intra-session switching occurs 22% of time for Infoseek and 15% for Yahoo!

H2. (*Search dissatisfaction hypothesis*). Higher dissatisfaction with search results, lowers the future use of the portal.

There is one more aspect of the user environment, namely *advertising exposure*, that may also affect the choice of a portal. We consider only exposure to online advertisements, generally known as banner advertisements [11,12]. Although the click-through rates are low, given the prevalence of advertising on the Internet, we have included these banner ad exposures in our choice model. We record whether the page visited immediately preceding a portal choice contained an advertisement for that portal. If it did, we consider the user to have been exposed to an advertisement. We believe advertisements that are seen immediately prior to the choice can have a positive impact on choice. We also include the demographic variables for the choice model. Finally, *portal intercepts* in Fig. 1 controls for portal-specific characteristics not included otherwise in our model.

3.4. Model of stickiness

The model of stickiness is based on user's willingness to spend time on the portal once the choice of the portal has been made. We expect the drivers of the stickiness model to be similar to those of the choice model discussed above. For example, higher repeat use is likely to increase stickiness but increased dissatisfaction is likely to have the opposite effect. Since the time spent on a portal is conditional on the choice decision, the banner ad is not a relevant here. We also include the portal specific dummies to control for different portals. Finally, we include the demographic variables for the stickiness model.

We argue that the type of task again plays a critical role here. Search tasks are goal-directed and can be accomplished quickly. Once the results are returned the users can quickly navigate to the referred sites. Therefore, we expect to observe a low level of stickiness for search tasks. On the other hand, the information service task is more browsing-oriented (for example, reading news or getting a weather report) and requires a long amount of time to be spent. Moreover, often users may choose some information services for leisure. For example, playing games or reading horoscopes requires user to spend a fair amount of time on the portal for these activities. Personal services also fall into a similar category. Participating in a message board, or writing and composing an email are time-consuming activities that may necessitate more time to be spent on the portals. Therefore:

H3. (*Task type hypothesis*). Type of task will play a crucial role in determining the stickiness of the portal. Search task will generate the least stickiness but personal and information service use should increase it.

We analyze this model in two steps. First, we examine stickiness at the portal level as a whole and then, we disaggregate across the three tasks: search, information and personal services.

3.5. Model of use frequency

The goal of the frequency model is to understand how frequent and stable portal use is over time. To accomplish this goal, we measure frequency as a count of the number of visits to all portals each week. Unlike the two previous models, we include time-specific drivers to explain the stability of portal use. Two specific drivers of this model are lag and trend as shown in Fig. 1. The lag variable captures the effect of prior use (previous week) on current use. The trend variable, on the other hand, examines how frequency of use changes over time. We also include repeat use and demographics as possible drivers. We expect repeat use to positively affect the frequency of use. Since CNE is specific to individual portals while our frequency measure is the aggregate weekly count, we do not include CNE in our frequency model.

As in the previous models, we disaggregate the frequency of use at the task level. This permits us to examine the stability as well as frequency of use for each task. We expect that the use of personal services will exhibit the highest stability. For example, we expect users to visit their mailboxes at regular intervals. However, their use of the search function will depend on their need for information, and can be uneven. Interestingly, the use of the search function can be more frequent, but less stable. Similarly, their use of information services may be subject to time availability, and may exhibit fluctuations. The task level model also allows us to investigate how the use of one service may lead to the use of the other two services. For example, we examine if the use of search function increases the use of information services.

H4. (*Frequency of use hypothesis*). We expect search services to be used more frequently but the use personal services should be more stable.

4. Model estimation

This section is divided into three subsections. Section 4.1 describes the model for repeat use. Section 4.2 describes the model for stickiness. Before we proceed, we first provide Table 1 which summarizes the independent and dependent variables of each model.

4.1. Repeat use model

The model for repeat use is basically a choice model. We first construct a measure for repeat use from the choices made by a user and estimate its impact on the subsequent choices of the portal. We test whether the parameter for repeat use is statistically significant. We use the commonly used multinomial logit for estimating the choice model.

Table 1
Independent and dependent variables in each model

	Dependent variable	Independent variables
1. Choice model		
Estimated using multinomial logit	Choice variable (1 for the portal visited, 0 for others)	Repeat use (RU), advertisement (Ad), cumulative negative experience (CNE), demographic variables (Demo), five portal level dummies
For the task level model		We create three RU variables for each task, RU_S (for search), RU_I (for information) and RU_PS (for Personal)
2. Stickiness model		
Estimated using fixed effect OLS	Number of minutes spent on a portal	Repeat use (RU), cumulative negative experience (CNE), demographic variables (Demo), six portal level dummies
For the task level model		We create three dummies variables (one for each task) and S (for search), I (for information) and PS (for personal)
3. Frequency model		
Estimated using negative binomial regression (NBD)	Count of number of visits to all portal in a week	Repeat use (RU), trend (T), previous week's count (Count_Lag), demographic variables (Demo)
For the task level (separate regression for each task)	Count of number of visits to all portals in a week for each task.	We add three lag variables, one for each task. Lag_S (Lag of search task), Lag_I (Lag of Information ask), Lag_PS (Lag of Personal task)

(See Ben Akiva 1989 for details.) But before that, we first outline our measure for repeat use.

By definition, repeat use is the use of same portal over time. While many different measures are possible, we use an exponentially weighted average of all previous use. [5] (See [10]; for details.) Therefore, we write repeat use as

$$\mathrm{RU}_{ikt} = \alpha \mathrm{RU}_{ik(t-1)} + (1-\alpha) d_{ik(t-1)}, \tag{1}$$

where α is the carry-over constant; $d_{ik(t-1)} = 1$, if the user i selects portal k at time $t-1$ and 0 otherwise.

Since the multinomial choice model is specified via random utility, we specify the utility of a portal as

$$U_{ikt} = V_{ikt} + \varepsilon_i,$$

where the deterministic component of utility is

$$V_{ikt} = \beta_{0k} + \beta_1 \mathrm{RU}_{ikt} + \beta_2 \mathrm{Ad}_{ikt} + \beta_3 \mathrm{CNE}_{ikt} + \beta_4 \mathrm{Demo}_i * \mathrm{RU}_{ikt}, \tag{2}$$

where β_{0k} is the portal type dummy for portal k, RU_{ikt} is the repeat use level of user i for portal k at time t, $\mathrm{Ad}_{ikt} = 1$ if user i was exposed to an online ad for portal k at time t and 0 otherwise; CNE_{ikt} is the proportion of time user i switched away from portal k until time t; and Demo_i is the demographic characteristics of user i.

Note that we use interaction of demographics with repeat use. There are two reasons for modeling demographics in this way. First, we want to investigate which demographic characteristic leads to higher repeat use. Second, since demographics are independent of the six portals, they cannot be estimated separately. We estimate the parameter β associated with each variable. Note that we can not identify all six portal dummies in a logit model. Therefore, we normalize the "Altavista" dummy to zero and estimate the other five. To estimate the carry-over constant α in our repeat use formulation and the loyalty coefficient β_1 in Eq. (2) we use the technique specified by [9].

To understand the impact of three tasks, we estimate another model at a task level. We create the loyalty variable, as in Eq. (1), but now it is specific to the each task. Therefore, the repeat use is split into: RU_S_{ikt}, RU_I_{ikt}, RU_PS_{ikt}, where (S, I and PS) indicate that the repeat use measure is for the specific task at portal k, by user i, at time t.

[5] This means that most recent past choices affect the current choices more than the distant past choices. Therefore, this measure fits well with the learning mechanism we propose.

4.2. Stickiness model

Once the choice of the portals has been made, the users decide to spend time on the portal, demonstrating what we have called stickiness. Stickiness is measured in the number of minutes spent on the portal. We use fixed effect OLS regression for estimation. As Fig. 1 indicates, there are three major independent variables affecting the level of stickiness of a portal. Therefore,

$$\text{Stickiness}_{itk} = \beta_{0k} + \beta_1 \text{RU}_{itk} + \beta_2 \text{CNE}_{itk} + \beta_3 \text{Demo}_i + \mu_i, \tag{3}$$

where β_{0k} is the portal type dummy for portal k. The rest of the variables have same interpretation as in Eq. (2).

For estimating the model, we use the log transformation of stickiness as it fits the model better. Moreover, instead of estimating the model for each portal, we pool the data. Since we pool the data, we can now estimate six separate dummies, without the constant of the regression included. The β_{0k} dummies indicate the portal type and we have six portals in our study.

To estimate the stickiness model at the task level, we add the additional task level dummies (S, I, and PS) in the stickiness model. The dummies S, I, and PS are '1' if a search task or an information task or a personal task, respectively, was performed at the portal and '0' otherwise.

4.3. Frequency model

For the frequency model, we calculate the number of times a user visited all portals within a week. Since frequency is represented by count data (which can never be less than zero), using OLS would bias our estimates. Therefore, we use the negative binomial regression instead of using simple regression. The model to estimate in frequency data is

$$\text{Count}_t = \beta_0 + \beta_1 \text{RU}_t + \beta_2 \text{Trend}_t + \beta_3 \text{demo} + \beta_4 \text{Count\_Lag}_{t-1} + \varepsilon. \tag{4}$$

Note that Trend represents week number and Lag corresponds to the frequency of the previous week. Since RU is also aggregated at weekly level, we take the maximum of RU generated for each portal during the week. Finally, to estimate the frequency model at the task level, we run three separate regressions for each task, search services, information services and personal services. Moreover, we also use three different lags (one each for each service) in these regressions (Lag_S, Lag_I and Lag_PS) rather than just one lag variable, Count Lag, as in (4).

5. Data

The data used for this study come from the HomeNet Project [15] that tracked Internet use at home. The details are in Appendix A.

We collected Internet navigation data for 102 demographically diverse Internet users from June 1998 to June 1999. The data used in this analysis were assembled from detailed usage records captured at the server and took four person months of coding time. In the interest of keeping our subsequent estimation tractable, we captured portal choices for the six most heavily visited (close to 90%) portals during the study period, Yahoo!, Lycos, Excite, InfoSeek, Altavista, and HotBot. This data set contained 6774 distinct portal choices. The minimum number of data points for a user was 7, and the maximum number was 367. The mean number of portal choices made by users over this period was nearly 67. The usual practice is to let the data initialize loyalty. So we used the observations from the month of June ($n = 453$) to initialize loyalty and the rest of the data ($N = 6321$) for model calibration.

Demographic characteristics of the users were measured using a survey. In Table 2A, we provide some statistics of the four important demographic variables (adult or kid, race, gender and income) we used in our study. While we have information on many different demographics (e.g., occupation and education, etc.), many of these are highly correlated. Therefore, we simply choose the four important and somewhat uncorrelated characteristics.

Table 2A
Demographic characteristics

Adult (1 for adult, 0 for child; age <18)	0.84 (0.34)
Income (on the scale of 1–7) 1 is <\$10,000 and 7 is >\$75,000	5.3 (1.42)
Race (1 is for white, 0 for others)	90% White
Gender (1 for male and 0 for female)	54% Males

SD is in parenthesis.

Table 2B
Summary statistics for the model

Portal	Repeat use level	Stickiness in minutes	Total frequency
Altavista	0.078	6.8	445
Excite	0.2263	13.98	1490
HotBot	0.0639	5.97	337
InfoSeek	0.1683	6.3	1056
Lycos	0.1495	7.0	986
Yahoo!	0.3135	9.86	2007

For the details on coding choices and creating CNE construct, see Appendix A. In Table 2B, we provide summary statistics for the dependant variables. The total frequency column has the total number of times users visit the portal during the observation time. Stickiness is the average number of minutes spent on each portal. Repeat use, similarly, is the average level of repeat use for each portal. Note that repeat use is calculated as in Eq. (1).

6. Results

We first estimate three models of choice, stickiness and frequency (Eqs. (2)–(4)) at the portal level data. We then re-estimate these models with the data at the task level.

6.1. Portal level analysis

We first present the results of the choice model and the stickiness model estimation in Tables 3A and 3B. For each parameter, we present the estimate and the t value. In both models, we test the hypothesis that repeat use is a significant factor. We also investigate the effect of negative experience and demographics in the portal choice and stickiness models.

Table 3A
Estimates for portal level choice[a]

Parameters	Estimates
Excite constant	0.400* (6.84)
Hotbot constant	−0.307* (−4.06)
Infoseek constant	0.332* (5.51)
Lycos constant	0.323* (5.29)
Yahoo! constant	0.468* (7.94)
RU	1.240* (7.64)
Ad	1.144* (14.7)
CNE	−1.394* (17.9)
Adult	−0.305 (−1.41)
Race	1.328* (4.31)
Income	1.202* (4.01)
Gender	−0.146 (−0.91)
LL (Log – likelihood) $N = 6321$	−7892

[a] t values are in the parentheses. *significant at 1% level.

Table 3B
Estimates for portal level stickiness[a]

Parameters	Estimates
Altavista constant	1.283* (3.75)
Excite constant	1.291* (3.84)
Hotbot constant	1.166* (3.34)
Infoseek constant	0.981* (2.90)
Lycos constant	1.363* (4.03)
Yahoo! constant	1.498* (4.44)
RU	0.588* (5.75)
CNE	−0.215 (1.67)
Adult	−0.022 (−0.14)
Race	−0.099 (0.40)
Income	−0.111* (−2.82)
Gender	−0.040 (−0.38)

[a] t values are in parentheses. *significant at 1% level. R^2 between = 0.17, R^2 within = 0.05, R^2 overall = 0.07.

The first five rows of Table 3A are the estimates for portal constants (β_{0k}). The portal constants show users' intrinsic preference for these portals relative to Altavista which is normalized to zero. Except for Hotbot, all other portal constants are significant and positive, suggesting that Hotbot is the least preferred portal followed by Altavista. On the other hand, Yahoo! and Excite are most preferred. The repeat use parameter is highly significant suggesting that portals are able to generate a strong level of repeat use among the users even at the aggregate level confirming H1. Interestingly, CNE (cumulative negative experience) is highly significant and negative. This means that as users experience more dissatisfaction with the portal, they tend to reduce their future visits to the portal confirming H2. Race and Income are positive and significant. Recall that we use interaction terms of demographics with repeat use in our model (2). Therefore, our results indicate that whites are more loyal than non-whites and higher income users are more apt to use the same portal than lower income users. It seems users with these characteristics are less apt to switch portals.

In Table 3B, we present the results of the stickiness model (4). As before, all the constants are significant. Note that in the fixed effect OLS model for stickiness (Table 3B) we can identify all six portal dummies (without the regression constant). A larger constant for a portal indicates that users spend more time at that portal, all else being equal. Yahoo! again has the largest coefficient. The repeat use parameter in the stickiness model is positive and significant. This means that loyal users also tend to spend longer time on the portal. Therefore loyal users provide two benefits. They tend to visit the same portal again and again and they also spend a longer amount of time on the portal. Interestingly, CNE is not significant although it is in the expected direction, suggesting that CNE affects the portal choice more than stickiness. Finally, income is negative and significant, suggesting that higher income users spend fewer minutes on the portal. Therefore higher income people, though more loyal, spend less time. Since we have panel data, we report the R^2 statistics for within group, between group, and the overall sample.

Next, we present our results from the frequency model in Table 3C. Recall that for each week starting from week 1, we count the total number of times a user went to all six portals. Therefore, we do not have portal specific constants in this model. (The constant shown in the table is simply the constant of regression.) Note that one could analyze frequency data at the individual portal level (Yahoo!, Lycos, etc.). However, a typical user is unlikely to visit all portals every week. Our data indeed show many zero counts at the individual portal frequency of use. Thus, disaggregating the data at portal level is not very informative.

Table 3C
Estimation of frequency model[a]

Parameter	Estimates
Constant	−0.47 (−1.71)
Gender	0.125 (1.46)
Adult	0.141 (1.17)
Race	−0.034 (−0.20)
Income	−0.012 (−0.40)
RU	1.124* (7.86)
Trend	−0.022* (−11.7)
Count_Lag	0.048* (15.4)
Log Likelihood (LL)	−5473
$N = 3455$	

[a] t values are in parentheses. *significant at 1% level.

Our results show that repeat use is again highly significant and positive indicating that loyal users also visit the portals more frequently. Note that we use a negative binomial regression for estimation, and therefore the coefficients should not be interpreted as linear effects. As expected, the coefficient of the Lag variable is positive and significant. This result signifies that users' visit to portals are fairly stable. We also notice a gradual decline in the frequency count in our sample as captured by the negative trend estimate. With respect to demographics, we find none to be significant. It is not surprising in this panel data because demographics do not change over time.

6.2. Estimation at task level

Now we disaggregate use to understand user behavior across the three tasks – search, information and personal services. We expect that the use of personal services will lead to a strong level of repeat use compared to both information services use and search services use. For the stickiness model, we expect a different pattern. User are expected to spend more time on information and personal services, whereas they will spend little time at the portal when they perform search queries. This is consistent with Google's strategy that users visit search engines many times, but spend little time for each visit [23]. To operationalize the repeat use across three tasks, we need at least five observations from a user for each task to construct a meaningful repeat use variable using Eq. (1). Following this rule, there are 35 users who used personalized features for a total of 861 times in our data set, which translates to about 25 uses per user. [6] Similarly, 65 users used the non-personalized features 1584 times, for an average of about 25 for non-personalized services like news, weather, etc. Finally, all users used the search 3629 times, for about 36 searches per user. We report the results in Tables 4A and 4B for the sample $N = 6074$.

The results of the task level choice model in Table 4A are consistent with our expectations. The coefficients of repeat use for all three tasks are significant and in the expected directions. However, note that the coefficient of personal services is much higher than both information and search services. In other words, the use of personal services makes users more loyal to a portal confirming H 1. The loyalty to the two other services seems equally strong. The remaining results are similar to the aggregate portal level results of Table 3A suggesting that our estimates are fairly robust. A smaller log likelihood number in Table 4A suggests that the task level model fits the data better.

The new results of the stickiness model in Table 4B also support our hypotheses. For example, we find that the coefficients of information and personal services are positive and significant, while that of search service is negative and significant. By the very nature of search service, users complete search tasks and leave the portal to go to other sites with the desired content. Therefore, the coefficient of this service is negative. The remaining estimates are similar to the aggregate portal level results of Table 3B indicating the robustness of our estimates. The larger R^2 numbers in Table 4B indicate a better fit, and highlight the importance of disaggregated analysis.

Table 4A
Estimates for task level choice

Parameters	Estimates
Excite constant	0.362* (6.15)
Hotbot constant	−0.310* (−4.11)
Infoseek constant	0.331* (5.49)
Lycos constant	0.343* (5.63)
Yahoo! constant	0.383* (6.33)
RU-S	1.925* (13.9)
RU-I	1.836* (11.2)
RU-PS	4.214* (15.3)
Ad	1.187* (15.0)
CNE	−1.231* (15.7)
Adult	−0.266 (−1.2)
Race	1.164* (3.67)
Income	0.851* (2.91)
Gender	−0.028 (−0.17)
Log likelihood (LL)	−7892

t values are in parentheses. *Significant at 1% level.

Table 4B
Estimates for task level stickiness

Parameters	Estimates
Altavista constant	1.330* (4.11)
Excite constant	1.104* (3.46)
Hotbot constant	1.154* (3.50)
Infoseek constant	1.052* (3.27)
Lycos constant	1.337* (4.19)
Yahoo! constant	1.136* (3.56)
S	−0.277* (−2.25)
I	0.693* (7.79)
PS	0.641* (5.80)
RU	0.584* (5.97)
CNE	0.025 (0.21)
Adult	−0.089 (−0.83)
Gender	−0.037 (−0.91)
Race	0.112 (0.69)
Income	−0.100* (−3.05)

We estimated only one parameter for repeat use in Tables 3A and 4A. But we expect that different portals to generate different levels of repeat use. We also observe in our data set that most of the information and personal services tasks were chosen on Yahoo! and Excite. Clearly, these two portals did an excellent job of encouraging the search users to use these other features. Therefore, we expect these two engines to have a much stronger loyal user base. To test our intuition, we calculated the elasticity of probabilities with respect to repeat

[6] If a user used the personalized services only 2 times then we cannot create a repeat use of personalized services because not enough history is available.

Table 4C
Elasticity of probabilities with respect to loyalty

	Altavista	Excite	HotBot	InfoSeek	Lycos	Yahoo!
Elasticity	0.27	**0.76**	0.19	0.56	0.54	**0.90**

use. The estimated β value of repeat use is not a slope parameter due to the non-linear logit model. We calculate the slope parameter for loyalty for six portals, and report the results in Table 4C. These estimates have a similar interpretation as the estimates in a linear regression. The results suggest that the marginal impact of loyalty on the probabilities of choosing Yahoo! or Excite is much higher than any other engine. Clearly, in our sample, users show the strongest loyalty to Yahoo!, followed by Excite.

Finally, we also estimate how frequency changes across three task levels. First we present the summary statistics. Table 4D shows that the highest visit frequency is for search service followed by information and personal services. This is in contrast to the stickiness results in Table 4B where we found that users spend least time for search service at the portal. Total average aggregate weekly frequency is about two visits per week.

Table 4D
Summary statistics for weekly usage frequency at the task level

Variable	Mean	SD	Min	Max
Search	1.15	2.33	0	25
Personalized	0.25	0.95	0	18
Information	0.47	0.19	0	19
Total frequency	1.88	3.42	0	56

The result at the task level is presented in Table 4E with t statistics in parentheses.

The frequency model in Table 4E now includes a lag variable for each of the three tasks to understand how the past use of a service affects the present use of all three services. As we disaggregate the data at the task level, some of the demographic variables become significant. In particular, we find that gender, income and race are significant for search service use. As before, we find repeat use and trend are both significant and are in the expected directions. We find that the uses of the three services are stable from week to week, with personal service being the most stable followed by search and information services. We also find that prior use of personal service positively affects the use of the other two services. On the other hand, prior use of information service has no impact on the use of the other two services. Finally, prior use of search service seems to affect information service use. On the whole, personal service is the most stable, and seems to lead the other services in increasing the frequency of portal use.

Table 4E
Frequency model at task level

Parameters	Personal service	Information service	Search service
Constant	0.104 (1.06)	0.274* (2.13)	0.280* (0.94)
Gender	0.016 (0.59)	0.041 (1.15)	0.188* (3.11)
Adult	−0.040 (−1.03)	0.056 (1.09)	−0.139 (−1.1)
Race	−0.016 (−0.26)	−0.283* (−3.44)	−0.360* (−2.9)
Income	−0.016 (−1.66)	−0.004 (−0.36)	0.054* (2.37)
RU	0.585* (7.99)	0.791* (8.25)	0.758* (8.48)
Trend	−0.005 (−5.71)	−0.008* (−6.18)	−0.016* (−6.06)
Lag_PS	0.497* (34.1)	0.168* (6.92)	0.122* (2.95)
Lag_I	0.022 (1.44)	0.309* (4.09)	0.098 (1.45)
Lag_S	−0.007 (−1.15)	0.033* (2.13)	0.386* (23.4)
Log likelihood	−6327	−6194	−6003
N = 3455			

t values are in parentheses. *Significant at 1% level. R^2: within = 0.09, between = 0.34, overall = 0.13.

7. Conclusion

As Internet use and electronic commerce continue to expand, it is important that we understand how consumers interact with Internet portal sites. These sites hold powerful positions in the marketplace, guiding users to sites of potential interest. This research is an attempt to explore how Internet users choose portals, a very common form of Internet portal site. A key contribution of this study is to model three metrics of use, namely repeat use, stickiness and frequency both at the portal level and at the three tasks level.

Our model is motivated by the notion that information goods such as portals are a different class of software products. We hypothesized that the ability to learn various portal features and the ease (or difficulty) of transferring this learning to other portals would determine loyalty in this context. We also explored why the user may exhibit differing levels of loyalty to search, information and personal services of the portals.

Overall, we find that users do develop loyalty for a given portal. If they have used a particular portal frequently in the past, they are much more likely to choose that portal again in the future. The quality of the results is also a strong predictor of user choices. If the user is dissatisfied with the portal results, then the probability that she will use the portal in the future, diminishes. A first-mover advantage can take the firm only so far. A poor quality portal cannot hope to develop a loyal base. We find that the search task alone does not develop strong repeat use. The loyalty becomes much stronger when the user starts using personalized features. This is a very important finding given the low barriers to entry in this industry and low economic costs of switching. Search portals should focus on developing features that can be personalized because such features can improve the value proposition to users. Moreover, personal services require additional learning, and thus increase the switching costs associated with patronizing another portal. Our results very clearly point out that personal services play an important role in building a strong loyal base, although encouraging users to use them may be a challenging task. Whereas Yahoo! and Excite seem to have done a good job in converting search users in our sample into using other features, other portals have not.

Interestingly, repeat use also has a favorable impact on both the stickiness and frequency of use. Use of information and personal services increases the time a users spends on a portal. However, search services are not sticky, as users move on to the search targets. But the search function is a key driver of traffic to the portals. Of the three services, personal services use shows maximum (week-to-week) stability. Information services use, on the other hand, shows the least stability. Personal services use also converts into using search or Information services. We also find that the effects of the three services on the three measures of use may conflict at times. For example, search services increase the frequency of use, but decrease the stickiness of use. Our results have implications for product development and portal strategies given that various portals follow different strategies in terms of stickiness, search quality and features offering.

Our results are based on actual usage data from the server. Yet our subject pool is diverse and compares well with the Internet population of today [20]. As mentioned previously, there are important advantages to using actual navigation data. These data, however, entail some restrictions. For example, we used a coding rule to infer the start of a new search session within a logon/logoff interval. To mitigate this restriction, we performed additional analysis but our results did not change appreciably when we altered the interval to an hour.

It is important for future research to more fully characterize the services provided by the portal sites. In our study, we classified these services into three broad categories. However, each of our service types can be further subdivided. For example, portals are offering various features for the search service such as advanced search, personalized search, etc. Similarly various new information and personal services are proliferating. Understanding user behavior at specific service level will generate new insights for portal sites. The "intra-session" switching behavior and sampling of portals observed in this study can also be generalized to other product categories. We might expect to see similar behavior for choices for shop bots, newspapers sites, etc. We believe that more research is necessary to explore this issue further.

Appendix A

The HomeNet project gave households a computer, modem, and an extra telephone line. Participants were also given three hours of training and free online support through a help newsgroup. HomeNet recorded a complete description of the Web sites visited by each user during each Internet session. Thus, for each user, we can determine which portals she visited, what she did, and to what advertising she was exposed to prior to a portal visit.

The most important characteristic of our data set is that it was collected in an unobtrusive and natural setting. Unlike other studies on the Internet where researchers have used either survey data or laboratory experiments, which may bias the results, our data set has no such weakness. Participants access the Internet from their home environment, not a laboratory setting. The data collected contain actual choices rather than elicited preferences. These features provide a uniquely appropriate data environment for studying Internet portal usage behavior.

To facilitate clear explanation, we also adopted the following definitions in collecting and coding the data:

A.1. Session

We use the following rule to define a search session. Within a given log-on/log-off interval, a new session begins anytime more than 30 min elapse between search-portal visits. Our coding scheme accounts for the possibility that a user may have multiple occasions to use a portal before exiting the Web. We later relax the assumption of the 30-min interval and examine the sensitivity of our results to this scheme.

A.2. Portal choice

From our data set we could easily distinguish the portal visited and the task performed by the user on that portal. We knew if the user performed a search and if so, what search term she entered. It is common to observe a navigation pattern in which a user chooses a portal, does some amount of query-based searching, visits some other sites, then returns to the portal (probably by using the "Back" button), and continues the search. The Web server used in this project captured this navigation pattern. Lest we overestimate repeat use level, we considered only unique visits in a search session. Thus, within a given session, a visit to the portal was considered a choice only the first time it was chosen. But if the user switched to another portal and then came back to the original portal and continued the search, we coded them as unique choices. For example, if a user visited Yahoo!, and Lycos, and then Yahoo! within half an hour to complete a search task, we coded these as unique choices. The navigation data were detailed enough so that we knew what search term users entered while using the portal. Therefore, if the user switched to another portal and continued with the similar search, then we could identify and count them to construct the variable CNE (cumulative negative experience) discussed previously.

Our records also identify whether the users used any of the information or personal services. For example, we can verify whether the choices were made for e-mail, chat room, news, etc., rather than search. For estimating our model at the task level, we construct the appropriate loyalty variable by coding the choices for search, information, and personal services. For the frequency analysis, we count the total number of times a user went to all six portals each week starting from week 1. However, some users dropped out of the HomeNet experiment before the end of our data collection period. Therefore, we do not have frequency data for these users for all 52 weeks. Finally, we also record the time users spend on the portal for each visit. The following table provides some demographics of the subject pool.

References

[1] Annual Report of Yahoo, 2002.

[2] D. Austin, Ford Motor Company: Maximizing the Business Value of Web Technologies, Harvard Business School Publishing, Boston, MA, 1997, pp. 9.

[3] Ben-Akiva, R.L. Steven, Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press, Cambridge-Mass, 1985.

[4] E. Bradlow, D. Schmittlein, The Little Portals that could: modeling the performance of World Wide Web portals, Marketing Science 19 (1) (1999) 43–62.

[5] E. Brynjolfsson, C.F. Kemerer, Network externalities in microcomputer software: an econometric analysis of the spreadsheet software, Management Science 42 (12) (1996) 1627–1647.

[6] J.M. Carroll, Interfacing Thought: Cognitive Aspects of Human–Computer Interaction, MIT Press, Cambridge, MA, 1987.

[7] M.C. Chuah, B.E. John, J. Pane, Analyzing graphic and textual layouts with GOMS: results of a preliminary analysis, Proceedings of the CHI '94 Conference Companion on Human Factors in Computing Systems (1994) 323–325.

[8] W.H. DeLone, E.R. McLean, The DeLone and McLean model of information systems success: A ten-year update 19 (4) (2003) 9.

[9] P.S. Fader, J.M. Lattin, J.D.C. Little, Estimating nonlinear parameters in the multinomial logit model, Marketing Science 11 (4) (1992) 372–385.

[10] P.M. Guadagni, J.D.C. Little, A logit model of brand choice calibrated on scanner data, Marketing Science 2 (3) (1983) 203–238.

[11] D.L. Hoffman, T.P. Novak, Bridging the racial divide on the Internet, Science (1996), April 17.

[12] D.L. Hoffman, T.P. Novak, A new marketing paradigm for electronic commerce, The Information Society 13 (Jan–Mar) (1996) 43–54.

[13] B. John, D.E. Kieras, The GOMS family of user interface analysis techniques: comparison and contrast, ACM Transactions on Human Computer Interactions 3 (4) (1996) 320–351.

[14] R. Kraut, W. Scherlis, T. Mukhopadhyay, J. Manning, S. Kiesler, The HomeNet field trial of residential Internet services, Communications of the ACM 39 (12) (1996) 55–63.

[15] R. Kraut, T. Mukhopadhyay, J. Szczypula, S. Kiesler, W. Scherlis, Information and communication: alternative uses of the internet in households, Information Systems Research 10 (4) (1999) 287–303.

[16] Media-Metrix. Available from: <www.jmm.com/xp/jmm/press/mediaMetrixTop50.xml>, October 2002.

[17] R. Telang, U. Rajan, T. Mukhopadhya, The market structure for Internet search engines, Journal of Management Information Systems 21 (2) (2004) 137–160.

[18] A. Newell, T.P. Moran, S.K. Card, The Psychology of Human Computer Interaction, L. Erlbaum Associates, Hillsdale, NJ, 1983.

[19] B.R. Nault, A.S. Dexter, Added value and pricing with information technology, MIS Quarterly 19 (4) (1995) 449–464.

[20] L. Rainie, D. Packel, More Online, Doing More, The Pew Internet and American Life Project, 2001 February.

[21] S. Shugan, The cost of thinking, Journal of Consumer Research 7 (2) (1980) 99–111.

[22] N. Staggers, A.F. Norcio, Mental models: concepts for human–computer interaction research, International Journal of Machine Studies 38 (1993) 587–605.

[23] The Interactive Wall Street Journal, Elevated notion of stickiness is discarded as hype fades, Thomas E. Weber, March 5, 2001.