ASIM ANSARI and CARL F. MELA*

Customized communications have the potential to reduce information overload and aid customer decisions, and the highly relevant products that result from customization can form the cornerstone of enduring customer relationships. In spite of such potential benefits, few models exist in the marketing literature to exploit the Internet's unique ability to design communications or marketing programs at the individual level. The authors develop a statistical and optimization approach for customization of information on the Internet. The authors use clickstream data from users at one of the top ten most trafficked Web sites to estimate the model and optimize the design and content of such communications for each user. The authors apply the model to the context of permission-based e-mail marketing, in which the objective is to customize the design and content of the e-mail to increase Web site traffic. The analysis suggests that the content-targeting approach can potentially increase the expected number of click-throughs by 62%.

E-Customization

Marketers have long realized the value of targeting and customization. Customized products and communications attract customer attention and foster customer loyalty and lock-in. Targeted communications aid customer decisions and reduce information overload, and highly relevant products yield satisfied customers. The customer loyalty that results from such personalization and targeting can translate into increased cash inflows and enhanced profitability. Customized marketing solutions are useful for both customer acquisition and retention and can engender successful, long-term relationships. However, customization has often proved difficult because of implementation challenges, insufficient customer information, and other factors.

The Web makes mass customization eminently possible. Firms can exploit the capabilities afforded by digitization and networking to provide unique content of direct relevance to each customer. Moreover, such tailoring of information can be done quickly and at low cost. Customization is possible in part because of the interactivity afforded by the Web. Firms can collect and update preference information of customers from on-site surveys and from the traces customers leave as they navigate through a Web site. This knowledge can then be seamlessly integrated with algorithms and software to customize content automatically for individual consumers. Indeed, customized design (serving different variants of content to different users at different points in time) represents one of the key features that differentiate the Web from more traditional media.

In the online world, content sites such as C-net and Yahoo can leverage a loyal customer base to increase readership and therefore advertising revenues.[1] Various surveys have extolled the rapid growth in advertising dollars on the Web. Given the large magnitude of expected revenue involved, content providers are increasingly turning toward customization strategies to increase their share of advertising income. Similarly, e-commerce sites such as Amazon.com and Dell can customize content (e.g., information, digital products such as software, advertising, promotions, and other incentives) to increase repeat purchases and cross-selling.[2]

In spite of this potential, few models exist to help firms implement one-to-one marketing on the Internet. Therefore, given the substantial potential arising from e-customization, it is our objective to develop a statistical and optimization approach for customization of information on the Internet. A secondary goal is to model clicking behavior on the Internet. Our procedures enable firms to customize both the content (what and how much information) and the design (rendition)

*Asim Ansari is an associate professor, Graduate School of Business, Columbia Business School (e-mail: maa48@columbia.edu). Carl F. Mela is Associate Professor of Marketing, The Fuqua School of Business, Duke University (e-mail: mela@duke.edu). The authors are listed alphabetically. The authors thank Bill Boulding, Andrew Gershoff, Don Lehmann, Kamel Jedidi, John Lynch, Richard Staelin, Seenu Srinivasan, and the seminar participants at Columbia and Duke Universities for their helpful comments. Special thanks to Rajeev Kohli for many insightful discussions and help on the optimization. Jeffrey Inman served as guest editor for this article.

[1]Lehman Brothers (Becker, Anmuth, and Leichter 2002) forecasts that online advertising will grow with a 12% compound annual growth rate from $5.3 billion in 2002 to $7.7 billion in 2005.

[2]The online market now exceeds $171 billion and is growing rapidly.

Journal of Marketing Research
Vol. XL (May 2003), 131–145

to their users. Specifically, we develop an approach that enables Web sites to customize permission-based e-mail communications to increase Web site traffic (though the approach can be more broadly applied to the issue of Web site customization). As such, we contribute to the growing literature on one-to-one marketing (e.g., Rossi, McCullouch, and Allenby 1996; Shaffer and Zhang 1995).

However, e-customization differs from some other contexts inasmuch as (1) the "product" (the e-mail or Web site itself) is complex and requires explicit optimization for customization; (2) the product can be constructed dynamically, thereby making customization truly possible; and (3) the media are highly addressable, thereby facilitating targeting. Moreover, although it is possible to target products or marketing tactics in other contexts, the costs of doing so (e.g., printing individual catalogs, the cost of point-of-sale coupons) can often be exceedingly high.

E-customization requires knowledge of individual-level preferences, and it is therefore important to accommodate unobserved sources of preference heterogeneity (by allowing model parameters to vary across consumers). Moreover, the efficacy of content and design is likely to differ across implementations. Although it is possible to identify and account for some of the factors affecting the response to content and design in a typical application, it is unlikely that a Web marketer can enumerate all the variables that affect consumer response. Accordingly, our approach also models sources of unobserved contextual heterogeneity across e-mail content and design. This enables Web marketers to better predict potential responses to a particular type of content on a particular e-mail (or, in the case of an on-site targeting strategy, a particular type of content on a particular Web page design).

In the marketing literature, sources of heterogeneity are typically modeled by means of a random-coefficients approach. The computer science literature on Web customization suggests the use of collaborative filtering approaches to model heterogeneity (Breese, Heckerman, and Kadie 1998). Collaborative filtering systems use data from users with similar preferences to recommend new items. Such systems form the basis for most commercial recommendation systems (e.g., Netperceptions, Macromedia). We show how model-based collaborative filtering can be implemented using a Bayesian semiparametric model. Specifically, we show how a mixture of Dirichlet process (MDP) probit model can be used to perform collaborative filtering and to flexibly accommodate different sources of unobserved heterogeneity. The Bayesian literature in marketing has predominantly used normal distribution to capture differences across customers (Allenby and Rossi 1999; Ansari, Essegaier, and Kohli 2000; Rossi, McCulloch, and Allenby 1996). The normal distribution has limited flexibility as it is unimodal, has thin tails, and does not accommodate skewness. These sources of inflexibility of the normal distribution could result in misleading inferences and inaccurate individual-level estimates (Escobar 1994). The MDP model, in contrast, is flexible enough to accommodate deviations from normality and, depending on the date, can automatically adjust to mimic either a finite mixture of support points or a continuous distribution for heterogeneity, whichever is appropriate. Moreover, because of its discrete representation of heterogeneity, it can mimic a collaborative filtering representation. In this article, we explicitly compare results from a MDP specification with those obtained from models that use normal distributions to capture heterogeneity.

Given a set of parameter estimates, customization then requires the construction of customized e-mails for each consumer. We therefore develop an optimization procedure for customized e-mail design. The optimization procedure uses as input the individual-level parameter estimates from the hierarchical Bayesian statistical model and enables firms to (1) select relevant information to include in an e-mail and (2) configure the content to enhance the probability of site visits. Although the permutations in design can be quite large, we provide an exact solution to the design problem. Our design problem shows how the promise of targeting that is offered by hierarchical Bayesian models can be brought to fruition.

The rest of the article proceeds as follows: In the next section, we detail various approaches to customization on the Internet. We then describe our statistical model for clicking behavior and provide the details of our application. Next, we overview the data, specify the model, discuss parameter estimates, and perform model comparisons. Then, we present the e-mail design optimization problem and the optimization approach for obtaining optimal configurations. Finally, we offer conclusions and suggest future research directions.

CUSTOMIZATION APPROACHES

Web sites can use a combination of on-site and external customization approaches to manage customer relationships. Both approaches are useful in enhancing site loyalty because they increase switching costs for users. When users are faced with a decision to switch sites, it is quite feasible that they will be reticent to invest the time to begin "training" another firm. As Alba and colleagues (1997) note, consumers might expect to experience switching costs and a decrease in customer service were they to switch to another site.

On-Site Customization

In this approach, companies either customize the Web site to appeal to users or enable the users themselves to customize the content. For example, portal sites such as Netscape and Altavista enable users to self-customize the site. Users of such sites can specify keywords of interest to filter news stories, can provide lists of stocks for which they require regular information, or can manipulate the page views themselves. Such user-initiated customization has obvious advantages, as it elicits user preferences and gives control to users in defining what they want.

However, in many cases, such a user-initiated approach may not be completely successful, as users may not be able to fully or accurately self-explicate their preferences. Many novice users may not feel confident about performing such customization actions. Moreover, preferences are dynamic and users may be reluctant to provide information continually (or may find it cumbersome to do so). In such situations, company-initiated customization based on revealed preferences data may be more useful.

On many Web sites, company-initiated on-site customization occurs in the form of recommendation systems. For example, firms such as Amazon.com and Kraft use recommendation systems to suggest products (Ansari, Essegaier, and Kohli 2000; Gershoff and West 1998) or content to customers.[3] Most commercial recommendation systems (e.g., Netperceptions, Macromedia) use techniques such as collaborative or content filtering on customer ratings data to determine customers' product preferences. Recommendations can also be made with attribute-based approaches. For example, Ansari, Essegaier, and Kohli (2000) show that hierarchical Bayesian models are best suited for recommendation systems because they incorporate different sources of heterogeneity and provide individual-level estimates even in sparse data environments.

These recommendation systems have typically been oriented toward suggesting a new product (e.g., a movie) or service rather than designing Web pages or e-mails. Although these approaches could be adapted to the customization of content (by extending them to consider multiple, concurrent recommendations), they are more difficult to adapt to issues of customized design, because design is a large-scale (many control variables) optimization problem (e.g., how to design a new product as opposed to recommending an existing one). Our approach therefore generalizes these previous works.

External Customization

An external customization approach is intended to bring users to a Web site. Typically, e-mails, banner advertisements, affiliate sites, or other communication media herald site content that may be of interest to site users. For example, companies such as Amazon.com, Morningstar, and *The New York Times* regularly send e-mails containing hypertext links to the content of their Web sites to registered users. E-mails intended to attract customers to a site typically contain (1) brief summaries of editorial content and (2) a link (or a set of links) to the site, where more detailed information can be found. After reading the summaries, users can click on the link listed in the e-mail. By learning user preferences from clicking histories and demographics, Web sites can tailor the messages in the e-mail to the user's interests. The greater the history of users' information, the more likely a firm can learn (and thus match) users' interests.

E-mail is one of the most popular Internet applications and is rapidly being adopted for e-commerce. Between 1999 and 2000, e-mail revenues increased 270% to $342 million and are expected to grow to $1 billion by 2003 (Aberdeen Group 2001). As click-through rates on banner advertisements continue to drop, e-mail is becoming the instrument of choice for business-to-consumer communication. Customization of e-mails to suit the preferences of users is therefore of paramount importance. Many companies, such as Doubleclick, Clickaction.com, Netperceptions, and Macromedia, have developed e-mail marketing systems to assist companies in outbound e-mail marketing. The specific details of their implementations, unfortunately, are not publicly available.

We focus on an external customization application for several reasons (our particular application involves personalizing permission-based e-mail design and content to attract the e-mail recipients to a Web site). First, external customization is relatively easy to implement, as firms do not need to create a significant number of alternative Web site designs (according to a Gartner Group study [Satterthwaite 1999], an average site costs $1 million and five months of implementation time). Second, internal customization strategies rely on the visitor to come to the site, which can take some time. In contrast, external customization strategies can be effective immediately, as communications are sent directly to the user by e-mail, post, or advertising. Direct communications, such as e-mails, enable firms to entice users more actively to the site and are therefore useful for both acquisition and retention activities. Third, on-site customization of the Web site can be risky if users have become familiar with the interface; changes to the home page can confuse loyal users. Fourth, as noted previously, e-mail targeting is an important and growing application in its own right.

Note that the application of our algorithm to e-mail does not preclude its use on Web page customization. Indeed, often e-mails are sent as Web pages that are opened directly by a Web browser. Yet porting our analysis to the redesign of Web sites would require careful deliberation. First, dynamic Web sites may confuse users, which suggests that greater benefit may accrue to varying content than design. Second, successful Web site customization is incumbent on reliably identifying site visitors and may thus be of limited use when dynamic Internet protocol (IP) addresses manifest or cookies are disabled. Third, as e-mails are served daily (or less frequently), there is ample time to estimate models and serve new content between points of contact with the user. Applications such as ours are not likely to scale well to on-the-fly customization, and more research is needed in that area. Nonetheless, it is possible to update Web sites on a daily or weekly basis, and substantial benefits might obtain even at this frequency of customization.

MODELING APPROACH

Using parameter estimates to custom design e-mails, we develop an individual-level model for estimating the probability of clicking on links within e-mails. Given that our objective is to customize content and design, our modeling approach consists of two phases:

Phase 1: In the first phase, we specify and estimate a probability model that correlates content and design characteristics to individual clicking likelihoods. The input to this model is individual-level clickstream data of past responses to content links included in e-mails. The output is a probability function and a set of individual-level and e-mail–level parameters that represent the preference structures of the users and the differences across e-mails.

[3]In addition to content recommendations, some companies provide navigational aids to help customers interact with the information provided on the Web sites. Perkowitz and Etzioni (1997) show how Web sites can use log-file analysis to suggest index pages of links to users. Such link prediction can also be used to fetch documents in advance while the user is reading a page. Alternatively, Web sites can suggest navigational shortcuts predicated on most popular navigational patterns. Sarukkai (2000) describes such a tour generation procedure based on Markov chains.

Phase 2: In the second phase, we use the probability model and the individual-level preferences as input to an optimization model. The optimization model recommends the optimal e-mail configuration (content and layout) for each recipient on each occasion.

As the optimization model is predicated on the results obtained from the econometric model, we first describe our statistical model and its results.

In the statistical model, we use an attribute-based approach and model the customer responses in terms of e-mail design attributes, content descriptors, and user characteristics. The database contains click-through responses of users for content links delivered in different e-mails. Let i = 1 to I index users, j = 1 to J represent e-mails, and k = 1 to K indicate the distinct links for which user response data are available. Each customer i provides binary responses y_{ijk} for n_{i2} links over n_{i1} e-mails. Let $E_i = \{j_1, j_2, \ldots, j_{ni1}\}$ be the index set of e-mails sent to user i and let $L_i = \{k_1, k_2, \ldots, k_{ni1}\}$ be the index set of links for which user i's responses are available. Users differ in the number of observations (links), and similarly e-mails differ in the number of links they contain, thus yielding a highly unbalanced data set. The total number of observations in the data set is given by $N = \Sigma n_{i2}$.

The observed binary responses (clicks), y_{ijk}, can be modeled through a random-utility framework. Users click on a particular link when the utility for exploring the content associated with the link exceeds a threshold. The relation between the observed response and the latent utility of clicking can be written as

$$(1) \qquad y_{ijk} = \begin{cases} 0 & \text{if } u_{ijk} \leq 0 \\ 1 & \text{if } u_{ijk} > 0 \end{cases}.$$

We model the latent utility u_{ijk} for link k of e-mail j for user i as the function of observed and unobserved e-mail and link characteristics. Specifically, the utility function can be written as

$$(2) \qquad u_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\mu} + \mathbf{z}'_{jk}\boldsymbol{\lambda}_i + \mathbf{w}'_{ik}\boldsymbol{\theta}_j + \gamma_k + e_{ijk},$$

where $e_{ijk} \sim N(0, 1)$. The vector $\mathbf{x}_{ijk}$ contains observed user-, e-mail- (design variables), and link-level (content) variables. The coefficients in $\boldsymbol{\mu}$ contain the "fixed effects" and describe the population-level impacts of the independent variables. The remaining terms specify the random effects and are used for capturing different sources of heterogeneity. Note that the error specification in Equation 2 assumes that errors are independent across the different links. Although it is possible to relax this assumption, we refrain from doing so for several reasons. First, such a solution would not be scalable, especially when the number of content categories is large. The introduction of dependencies through a multivariate probit specification leads to a substantial increase in computational complexity of the estimation and optimization algorithms, making broader implementation of our model difficult. Second, the introduction of covariance terms between the utilities of links makes customized design difficult, as customization would require the knowledge of the correlations between each new link and all others, as well as the evaluation of a multivariate integral. Third, our model predicts response well even though we assume independence (we predict 649 e-mail clicks in our data and observe 639). Therefore, in our view, the costs of this approach outweigh the benefits.

It is unlikely that all factors that affect responses can be isolated in any given application. It is therefore crucial to allow for multiple sources of heterogeneity. For example, users may differ in their preferences for content and in their propensities to click on links in different portions of the e-mail. Moreover, these differences in preferences may be unrelated to user demographics and other observed user variables. To allow for such unobserved preference heterogeneity across users, we introduce the term $\mathbf{z}'_{jk}\boldsymbol{\lambda}_i$ in the utility function. The vector $\mathbf{z}_{jk}$ can contain a subset of the variables in $\mathbf{x}_{ijk}$ such as the content descriptors and design variables pertaining to the link k and the e-mail j. The individual-specific coefficients in $\boldsymbol{\lambda}_i$ then indicate how the content and design preferences for individual i differ from the population average, $\boldsymbol{\mu}$.

In addition to modeling preference heterogeneity, in applications involving responses to information products, it is also important to model contextual heterogeneity. E-mails, being information products, are complex entities and therefore are not completely amenable to simplistic feature-based renditions. For example, some e-mails may be better designed than others, and the design features may interact in intricate patterns, which makes it difficult to code an e-mail on the basis of few observable attributes. Thus, e-mails may differ in terms of both observed and unobserved attributes. It is therefore desirable to account for the contextual influences with a random-effects approach. Accordingly, we use the term $\mathbf{w}'_{ik}\boldsymbol{\theta}_j$ to capture the differential impact of e-mail j on the utility function. The vector $\mathbf{w}'_{ik}$ can contain link-level and individual-level variables. The coefficients in $\boldsymbol{\theta}_j$ indicate how e-mail j differs from the average e-mail in terms of the impact of link-level and individual-level variables on click-through.

A broad categorization of the content of a link is possible based on a few features. However, it is likely that the content remains only partially explained in terms of the observed variables. For example, although a news item can be broadly classified as a "business news" item, there can still be considerable variation in content among all business news items. We therefore use a random effect γ_k to accommodate unobserved content heterogeneity.

Dirichlet Process Priors

In this section, we show how semiparametric distributional assumptions on the population distribution of the random effects can yield a principled approach for model-based collaborative filtering. The random effects are assumed to come from a population distribution. The marketing literature has used either finite mixture distributions (Wedel and Kamakura 2000) or continuous distributions (Ansari, Jedidi, and Jagpal 2000) to represent population heterogeneity. Finite mixtures allow flexibility, but in complex models, it is difficult to determine the appropriate number of components. Moreover, it is not straightforward to incorporate multiple sources of heterogeneity in finite mixture models. Alternatively, continuous distributions are used as part of hierarchical Bayesian models to capture heterogeneity. In this case, typically a normal population distribution is used to represent the variation in random effects. Although the choice of the normal distribution is made for tractability and conjugacy reasons, this assumption may not necessarily hold in reality. The normal distribution provides limited flexibility because it is unimodal, has thin tails, and does not accommodate skewness. If the population distribution is not normal, misleading inferences about the magnitude of effects and the nature of heterogeneity are possible. Researchers have used finite mixtures of normal components (Allenby, Arora, and Ginter 1998) to circumvent these problems, but in such models the difficulty of determining the number of components remains.

In this article, we show how an MDP model (Escobar 1994; MacEachern 1994) can be used to model heterogeneity in a flexible yet structured manner. The MDP model enables us to capture the uncertainty about the functional form of the population distribution using a semiparametric approach. This model avoids the typical assumption of a parametric population distribution such as the normal and instead uses an unknown distribution F to model heterogeneity. As this population distribution is assumed to be random, in the MDP model, a Dirichlet process prior (Blackwell and MacQueen 1973; Ferguson 1973, 1974) is placed on the population distribution F. The Dirichlet process provides a mechanism of placing a probability distribution on the space of distributions.

The Dirichlet process prior $F \sim D(F|F_0, \alpha)$ is described by two parameters: F_0 is a parametric baseline distribution that defines the "location" of the Dirichlet process prior, and α is a positive scalar precision parameter that determines the concentration of the prior for F about the baseline distribution F_0. The baseline distribution F_0 can be considered a prior "guess" for the population distribution. In this article, we use a normal distribution as the baseline distribution. The precision parameter α determines how close the nonparametric distribution F is to the baseline distribution F_0. When α is large, a randomly sampled population distribution F is very similar to F_0. Therefore, if the baseline distribution F_0 is normal and the precision parameter $\alpha \to \infty$, then the population distribution is a discrete distribution that mimics a normal distribution. In contrast, when α is small ($\alpha \to 0$), the sampled population distribution has its mass concentrated on a few points and is therefore similar to a finite mixture distribution.[4]

As the precision parameter is inferred from the data, the MDP specification allows flexible incorporation of heterogeneity. If the nature of the heterogeneity is consistent with a normal distribution, the precision parameter is automatically adjusted to be large and the MDP yields a population distribution that mimics a normal. Conversely, if the data come from a nonnormal population distribution, the MDP model allows enough flexibility because of its semiparametric nature and, as do finite mixture models, accommodates deviations from normality. In addition, the number of "segments" is automatically determined by the MDP algorithm as outlined subsequently. Thus, the MDP model places fewer restrictions on the shape of the population distribution, and as the population distribution is well approximated, it can have beneficial consequences for the accuracy of individual-level estimates (see Escobar 1994).

[4]Escobar (1994) indicates how sample size interacts with α.

In the context of our model, we assume that the user-, e-mail-, and link-specific random effects come from population distributions that arise from different Dirichlet process priors. Specifically, we use

$$\text{(3)}\qquad \begin{aligned} \lambda_i &\sim F_1 \\ F_1 &\sim D\big[F_1|N(0, \mathbf{\Lambda}), \alpha_1\big] \end{aligned}$$

to model the user-specific random effects. Equation 3 assumes that the user-specific random effects come from an unknown population distribution F_1. The population distribution in turn comes from a Dirichlet process prior with a multivariate normal baseline distribution that has a mean $\mathbf{0}$ and an unknown covariance matrix $\mathbf{\Lambda}$. The precision parameter of the Dirichlet process prior, α_1, controls how close the sampled population distribution is to the baseline normal distribution.

Similarly, the e-mail random effects can be modeled as

$$\text{(4)}\qquad \begin{aligned} \boldsymbol{\theta}_j &\sim F_2 \\ F_2 &\sim D\big[N(0, \mathbf{\Theta}), \alpha_2\big], \end{aligned}$$

where the baseline distribution is a normal with $\mathbf{\Theta}$ as the covariance matrix. Finally, the link-specific random effects can be modeled as

$$\text{(5)}\qquad \begin{aligned} \gamma_k &\sim F_3 \\ F_3 &\sim D\big[N(0, \tau), \alpha_3\big], \end{aligned}$$

where τ represents the variance of the associated univariate baseline distribution.

A Bayesian approach is needed for inference regarding the unknown parameters of the MDP model. The unknown quantities in our model include $\{\{u\}, \boldsymbol{\mu}, \{\boldsymbol{\lambda}_i\}, \{\boldsymbol{\theta}_j\}, \{\gamma_k\}, \mathbf{\Lambda}, \Theta, \tau, \alpha_1, \alpha_2, \alpha_3\}$. The joint posterior distribution cannot be written in closed form, and therefore Markov chain Monte Carlo (MCMC) methods (see Bush and MacEachern 1996; Doss 1994; West, Muller, and Escobar 1994) are needed to sample from the posterior distribution. The MCMC methods involve sampling iteratively from the full-conditional distributions.

To understand further how the MDP model implements model-based collaborative filtering and to explicate how the MDP model differs from the popular approach of using normal population distributions, we contrast here the full-conditional distribution of the user-specific random effects λ_i obtained from these models. The full-conditional expresses the uncertainty about λ_i given the values of the other unknowns. In contrasting these full-conditional distributions, let $\tilde{u}_{ijk\lambda} = u_{ijk} - \mathbf{x}'_{ijk}\boldsymbol{\mu} - \mathbf{w}'_{ik}\boldsymbol{\theta}_j - \gamma_k$ represent the adjusted utility for an observation. Then $\tilde{u}_{ijk\lambda} \sim N(\mathbf{z}'_{jk}\boldsymbol{\lambda}_i, 1)$. Form the vector $\tilde{\mathbf{u}}_i$ by stacking the adjusted utilities $\tilde{u}_{ijk\lambda}$ for all the observations of the user, and form the matrix $\mathbf{Z}_i$ by stacking row by row all the row vectors $\mathbf{z}'_{jk}$ for the observations belonging to user i. When $\boldsymbol{\lambda}_i$ is assumed to be distributed normal $N(0, \mathbf{\Lambda})$ (as is the case for one of our null models), it is well known that the full-conditional distribution for the random effects $\boldsymbol{\lambda}_i$ is multivariate normal and can be written as

$$\text{(6)}\qquad p(\boldsymbol{\lambda}_i|\{u_{ijk}\}, \boldsymbol{\mu}, \{\boldsymbol{\theta}_j\}, \{\gamma_k\}, \mathbf{\Lambda}) \sim N(\hat{\boldsymbol{\lambda}}_i, \mathbf{V}_i),$$

where the posterior precision is given by $\mathbf{V}_i^{-1} = \mathbf{\Lambda}^{-1} + \mathbf{Z}'_i\mathbf{Z}_i$, and the posterior mean is given by $\hat{\boldsymbol{\lambda}}_i = \mathbf{V}_i + \mathbf{Z}'_i\tilde{\mathbf{u}}_i$.

In contrast, for the MDP model, the full conditional for $\boldsymbol{\lambda}_i$ is given by the following mixture of a normal distribution and a mass point distribution:

$$(7)\qquad p(\boldsymbol{\lambda}_i|\{\boldsymbol{\lambda}_k, k \neq i\}, \{u_{ijk}\}, \boldsymbol{\mu}, \{\boldsymbol{\theta}_j\}, \{\gamma_k\}, \boldsymbol{\Lambda}) \sim q_{p0}F_b(\boldsymbol{\lambda}_i|.) + \sum_{l \neq i} q_{pl}\delta_{\lambda_l},$$

where

- $F_b(\boldsymbol{\lambda}_i|.)$ is the baseline posterior distribution given in Equation 6.
- The weight associated with the normal component is $q_{p0} \propto \alpha_1 f_i$, where f_i is the marginal density of the adjusted utilities for user i under the multivariate normal baseline prior density $N(0, \boldsymbol{\Lambda})$. The marginal density is obtained by integrating out the user random effects—that is, $\int f(\tilde{u}_i|\boldsymbol{\lambda}_i, \boldsymbol{\mu}, \{\boldsymbol{\theta}_j\}, \{\gamma_k\})N(0, \boldsymbol{\Lambda})d\boldsymbol{\lambda}_i)$ and is multivariate normal $N(\mathbf{0}, \mathbf{Z}_i\boldsymbol{\Lambda}\mathbf{Z}_i' + \mathbf{I}_{n_i})$, where $\mathbf{I}_{n_i}$ is the identity matrix. The quantity f_i is obtained by evaluating at $\tilde{\mathbf{u}}_i$, the marginal density.
- $q_{pl} \propto f(\tilde{\mathbf{u}}_i|\boldsymbol{\lambda}_l, \boldsymbol{\mu}, \{\boldsymbol{\theta}_j\}, \{\gamma_k\})$, the normal density of the utilities for user i evaluated by means of user *l*'s parameters; that is, each q_{pl} is proportional to the multivariate normal density of $\tilde{\mathbf{u}}_i \sim N(\mathbf{Z}_i\boldsymbol{\lambda}_l, \mathbf{I}_{n_i})$.

The weights q_{pk} are standardized to sum to 1. The parameter δ_s indicates a degenerate distribution with point mass at s. Therefore, in Equation 7, with probability proportional to q_{p0} we sample $\boldsymbol{\lambda}_i$ from the full-conditional under the baseline population distribution; that is, using Equation 6 and probability proportional to q_{pl}, we select from the degenerate distribution $\delta_{\lambda l}$, which means that we set $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_l$ (i.e., we set person i's parameters to be the same as person *l*'s). This results in a mixture with one component being a normal distribution, and all other components are point masses on the parameters of other users.

Intuitively, this mixing scheme implies that in any iteration of the MCMC scheme, if the likelihood of observing user i's data is relatively large using user *l*'s parameters, then the random effect $\boldsymbol{\lambda}_l$ for person *l* is more likely to be chosen as user i's random effect. In this case, users i and *l* would be in the same cluster. Conversely, if the likelihood of observing i's data is relatively low when user *l*'s random effect is used for user i, the more likely it is that the i's random effect is a new value (generated from the baseline distribution or from a different user, *l*′, whose parameters generate a higher likelihood for i's data). This mixing scheme results in a clustering of the random effects because users share common random-effect parameters on any given iteration. However, the number of "clusters" and the allocation of users to the clusters change from one iteration to another. Thus, this mechanism for generating the user-specific random effects is similar to what can be called model-based collaborative filtering on parameter space, as information from similar users is used to predict the users' preferences. In contrast to the standard model, in which the estimates for a user depend on the data for other users only through the population mean and variance, here, the posterior of $\boldsymbol{\lambda}_i$ heavily depends on the data for the user and the data of "nearest neighbors."

APPLICATION

Data

The data for this project are furnished by one of the Web's leading Internet sites. As a condition for their use, the sponsoring firm wishes to remain anonymous, and any identifying aspects of the data are therefore disguised. The organization derives the majority of its revenue from selling advertising space on its Web site to client firms and is therefore highly interested in increasing site usage. To accomplish that, it sends e-mails to registered users inviting them to visit the site. These e-mails include a synopsis of several articles on the site. Below each article summary is a link to the article (the link is a Web address that readers can click on to go to the site). It is the response to these links that forms our dependent measure.

To clarify our exposition of the site and its content further, we depict the site as an automotive news and reviews site (though the content is not automotive). We also categorize the site's content (including the links it sends in its e-mails) much like an automotive site can be categorized as cars or trucks. These categorizations are denoted content areas and are established by the sponsoring firm on the basis of its knowledge of the industry. On this site, a particular content area (e.g., the car portion of the site) sends a daily e-mail to users who register to receive the e-mail. These e-mail links pertain to content types within the content area. For example, the link types for a car area may include car reviews, car pricing, car specifications, automotive news, and so forth. Upon receiving this e-mail, users might click on a particular type of link. If they do, this information is recorded. There are two key components to the data set provided by this firm: (1) the user log files that record usage history for a given respondent and (2) the e-mail files that provide the date, content, and design of the e-mails.

User log files. Each time the visitor to the site clicks on an e-mail or Web site link, a record is generated and stored. The record contains the time, the origin of the click (the IP address), and information about the link or Web page that was clicked. These records, collected between June and August 1999, form one of the two key portions of the data we analyze. The origin of the click (the clicker's identity) is determined by "cookies," or records placed on the visitor's (clicker's) hard drive, as well as the user's IP address. When the user visits the site again, the content provider makes note of the cookie on the user's computer to ascertain whether he or she visited the site before.

E-mail files. Visitors to the content provider's site can request to receive e-mails pertaining to information on the site. Only people who are registered receive e-mails; therefore, all recipients had registered.[5] In addition to the e-mail addresses of the users, the e-mail files contain the dates of the e-mails, the links listed on the e-mails, and a synopsis of information contained on those links. This information is used to code the design and content of the links provided in the e-mail and to infer which links were not clicked. The

[5]Note that our model applies equally well to contexts in which users are not registered. In this case, e-mail addresses are obtained from the purchase of e-mail lists from third-party vendors.

design of the e-mail includes (1) the amount (number) of links listed in the e-mail, (2) the order of the links, and (3) the e-mail type (html or text). The link content in our data is coded into content categories provided by the firm. For example, an article reviewing an automobile would be coded "review," and automotive news would be coded "news." All in all, there are 12 of these content categories. Finally, information regarding the links in the e-mails can be merged with the database on users' clicking histories to determine which links (if any) the user clicked.

Sample size. All totaled, there are three months of e-mail file data and two months of log file data. The number of e-mails averaged about five per week, and there are 1048 users in our sample.[6] The number of links per e-mail averaged approximately 5.6 and ranged from 2 to 8. The average response rate across links is approximately 7%. We used data from a random sample of 100 users to estimate the models and assigned 60% of the e-mails (11,436 observations) into an estimation sample and 40% of the e-mails into validation sets.

We created two such validation data sets. In the first, we randomly held out observations across all e-mails, users, and links, so the validation set included extant e-mails and extant links. Therefore, predictions regarding a user's likelihood of clicking on a given link could be made with information regarding how others clicked on that link. The size of this validation set, which we denote "Extant," was 3735 observations. In the second validation data set, the sample was composed entirely of new links and e-mails. Thus, only information regarding users' past behavior could be used to forecast the likelihood of clicking on a link. This validation data set, which we denote "Novel," consisted of 3900 observations.

Ansari, Essegaier, and Kohli (2000) outline the benefits of creating multiple validation data sets. In contexts in which extant communications are to be targeted to additional consumers who have yet to receive them, the Extant data set is more relevant. Such may be the case after a test e-mail or when an e-mail is sent to additional customers of a site. In contexts in which new content is to be targeted, the Novel data set may be more relevant. Such may be the case when new editorial content is generated, as in the case of daily news.

Model Specification

Previously, we developed a general random-effects model that can be applied across a number of contexts. Here we describe the specific instantiation for our application. We include several observed link variables, e-mail design variables, and user variables in our specification. In addition to these observed effects, we allow for different sources of unobserved heterogeneity across users, e-mails, and links. Our specification can be described as follows:

$$
\begin{aligned}
(8)\quad u_{ijk} = {} & \mu_1 + \mu_2\text{Content}_1 + \cdots + \mu_{12}\text{Content}_{11} + \mu_{13}\text{Position} \\
& + \mu_{14}\text{Num-Items} + \mu_{15}\text{Text} + \mu_{16}\text{Since} + \lambda_{i1} \\
& + \lambda_{i2}\text{Content}_1 + \cdots + \lambda_{i,12}\text{Content}_{11} + \lambda_{i,13}\text{Position} \\
& + \lambda_{i,14}\text{Num-Items} + \lambda_{i,15}\text{Since} + \theta_{j1} + \theta_{j2}\text{Content}_1 \\
& + \cdots + \theta_{j,12}\text{Content}_{11} + \theta_{j,13}\text{Position} + \theta_{j,14}\text{Since} \\
& + \gamma_k + e_{ijk}.
\end{aligned}
$$

Link variables. The link variables characterize both the editorial content of the link and its position within the e-mail. The likelihood of clicking on a link will be a function of its content (e.g., news versus reviews) because users are expected to differ in their preference for content categories. The editorial content within the e-mails can be categorized into 12 categories (or content types). These content types are included by means of a set of 11 dummy variables (Content 1–Content 11).

As is the case with traditional media, the placement of the link may affect response (click-through). Hanssens and Weitz (1980) find that the later the advertisement is presented in a magazine, the less likely it is to be seen or read. Hoque and Lhose (1999) have replicated this result in electronic media. They argue that the impact of placement is magnified in electronic media because it is more difficult to read online and because of the effort involved in scrolling. The ordinal position of the link within the e-mail is represented by a variable (Position).

E-mail variables. We specify two e-mail design variables. The first represents the number of links within the e-mail (Num-Items). Houston and Scott (1984) show that increasing the number of advertisements within a print environment lowers the readership of any particular advertisement on a page. We expect a similar result in electronic settings as an increase in the number of links exacerbates clutter and therefore increases the cognitive costs of perusing the e-mail. For some users, the cost of needing to scan a large number of links may exceed the potential value of reading the e-mail. The second e-mail–level variable (Text) characterizes whether the e-mail is a text e-mail or an html e-mail.[7]

Person variable. As the duration of time increases since the last link clicked, the likelihood of a subsequent click may change. Therefore, we include a covariate for days since the last click (Since). We use one observation to initialize this variable. In customer relationship management applications, an increase in time since the last purchase leads to a decrease in subsequent purchase likelihood (Schmittlein, Morrison, and Colombo 1987). Therefore, we expect that the coefficient for Since will typically be negative, as users who have not clicked in some time are less likely to be active. The descriptive information for the link, e-mail, and person variables is presented in Table 1.

User heterogeneity. We model unobserved user heterogeneity by specifying a random-effects model for both the intercepts and the slopes in our model. Thus, we specify (1) user-specific random intercepts (which capture differences across users in their propensities to click on links) and (2)

[6]Only users who responded to at least one e-mail per month are included in the sample. People who never or rarely respond represent 24% of the sample but only 2% of the clicks. Because of their lack of response, there is little information available to improve this number. Therefore, customization is likely to be inefficacious for these people, so we omit them from our analysis. We note that the practice of including only respondents has an analog in scanner data research pertaining to category incidence (Manchanda, Ansari, and Gupta 1999; Mela, Jedidi, and Bowman 1998).

[7]We tested for an interaction between Num-Items and Position. This did not enhance model fit. We also tested for nonlinear effects for these variables and found none.

Table 1
VARIABLE DESCRIPTORS

	Calibration Data		*Extant Validation Data*		*Novel Validation Data*	
Variable	*Mean*	*Standard Deviation*	*Mean*	*Standard Deviation*	*Mean*	*Standard Deviation*
Link Click	.075	.263	.077	.250	.087	.282
E-mail Click	.307	.461	.278	.448	.350	.477
Text	.635	.481	.623	.485	.621	.485
Num-Items	5.609	.784	5.620	.803	5.784	1.021
Position	3.305	1.656	3.310	1.663	3.392	1.747
Since	7.197	9.372	8.111	10.204	6.160	8.011
Content1	.181	.385	.180	.384	.179	.384
Content2	.182	.386	.182	.386	.179	.384
Content3	.182	.386	.182	.386	.179	.384
Content4	.069	.255	.065	.246	.068	.252
Content5	.062	.242	.064	.244	.065	.246
Content6	.061	.239	.064	.245	.052	.221
Content7	.036	.187	.038	.191	.035	.184
Content8	.036	.187	.037	.189	.033	.179
Content9	.022	.148	.023	.150	.033	.178
Content10	.037	.188	.036	.187	.036	.185
Content11	.032	.177	.035	.184	.031	.173

user-specific random slopes (which capture differences across users in their responses to link content, e-mail variables, and time since last click). We note that user and random effects are identified only for variables that are not fixed across users (similarly, e-mail random effects are identified only for variables that are not fixed across e-mails).

E-mail heterogeneity. We capture e-mail heterogeneity by allowing (1) an e-mail–specific random intercept that captures e-mail "attractiveness" and (2) e-mail–specific random slopes that capture differences across e-mails. The unobserved e-mail variables (captured by $\boldsymbol{\theta}_j$ in Equation 8) interact with the link-level variables such as the content of the link and the position of the link within the e-mail. These interactions enable us to model contextual heterogeneity. As it is not possible to completely describe e-mails using observed attributes, incorporating e-mail–level heterogeneity is crucial for modeling click-through probabilities.[8]

Link heterogeneity. For parsimony, we capture link heterogeneity by including a single link-specific random effect. This term captures the impact of unobserved link effects—for example, the particular editorial (news or review) content of a link type on a given day—that is left unexplained by the content variables describing the link.

RESULTS

We estimated three models on the data. The first model is a simple model (Model S) that includes no heterogeneity. The second model (Model N) includes all observed variables and accounts for user-, e-mail-, and link-specific unobserved sources of heterogeneity, using normal population distributions for the random effects. The third model (Model DP) uses Dirichlet process priors to account for user, e-mail, and link heterogeneity.

We estimated the models using MCMC methods. The full-conditional distributions used in MCMC sampling for the DP model are available on request from the authors. For each model, we ran the chain for 30,000 iterations. The results reported are based on a sample of 20,000 draws from the posterior distribution, after we discarded 10,000 burn-in draws. We ensured convergence by monitoring the time series of draws.

Model Selection

Model fit. We use the pseudo-Bayes factor (PsBF) (Geisser and Eddy 1979; Gelfand 1996) to compare different models. The PsBF is based on the cross-validation predictive density, which we can conveniently compute for our models using the MCMC draws. Let $\mathbf{y}$ be the observed data, y_{il} represent the lth observation for user i, and $\mathbf{y}_{(il)}$ represent the data with the observation l for user i deleted. The cross-validation predictive density can be written as

$$(9) \qquad \pi(y_{il}|\mathbf{y}_{(il)}) = \int \pi(y_{il}|\boldsymbol{\beta}, \mathbf{y}_{(il)})\, \pi(\boldsymbol{\beta}|\mathbf{y}_{(il)})d\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is the vector of all parameters in the model. The PsBF for comparing two models (M1 and M2) is expressed in terms of the product of cross-validation predictive densities and can be written as

$$(10) \qquad \text{PsBF} = \prod_{i=1}^{I} \prod_{l=1}^{n_i} \frac{\pi[(y_{il})|\mathbf{y}_{(il)}, \text{M1}]}{\pi[(y_{il})|\mathbf{y}_{(il)}, \text{M2}]}.$$

The PsBF summarizes the evidence provided by the data for M1 against M2, and its value can be interpreted as the number of times model M1 is more (or less) probable than model M2.

The PsBF for our model can be calculated easily from a sample of d MCMC draws $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d\}$. As $\boldsymbol{\beta}$ is the vector of all parameters, including all the random effects, the binary responses y_{il}, i = 1 to I, l = 1 to n_i, are conditionally independent given $\boldsymbol{\beta}$. Thus, a Monte Carlo estimate of $\pi(y_{il}|\mathbf{y}_{(il)})$ can be obtained as

$$(11) \qquad \hat{\pi}(y_{il}|\mathbf{y}_{(il)}) = \left[\frac{1}{d}\sum_{t=1}^{d} \frac{1}{\left(p_{il}^{(t)}\right)^{y_{il}}\left(1 - p_{il}^{(t)}\right)^{1-y_{il}}}\right]^{-1},$$

[8]We thank the guest editor for noting that e-mail heterogeneity may also capture potential confounds between content and position.

where $p_{il}^{(t)}$ is the probability $\Pr(y_{il} = 1|\beta^{(t)}) = 1 - \Phi(u_{il}^{(t)} - \mathbf{x}'_{il}\boldsymbol{\mu}^{(t)} - \mathbf{z}'_{j(l)k(l)}\boldsymbol{\lambda}_i^{(t)} - \mathbf{w}'_{ik(l)}\boldsymbol{\theta}_{j(l)}^{(t)} - \gamma_{k(l)}^{(t)}$, where the superscript t represents the tth draw of the MCMC sampler and j(l) denotes the e-mail associated with the lth observation and k(l) represents the link associated with that observation. Gelfand (1996) provides the derivation for Equation 11 for the cross-validation predictive distribution. The estimates from Equation 11 can be used to calculate the logarithms of the numerator and denominator of the PsBF. These quantities can be considered a surrogate for the log-marginal data likelihoods for the two competing models. On the basis of the MCMC output, the log-marginal data likelihood for Model DP is –2290.25, for the normal model (Model N) is –2340.45, and for the nonheterogeneous model is –2807.22. Accordingly, the PsBF implies an exp(50.2) improvement for Model DP over Model N and an exp(517.0) improvement for Model DP over Model S. Thus, the PsBFs provide support for the Dirichlet process model.

Model Predictions

The predictive ability of the three competing models can be assessed through the validation data. The link-level probabilities for the links within an e-mail can be used to compute an e-mail–level probability of at least one click-through from the e-mail. To compute this probability, we first use the observed data to determine the likelihood of clicking on each link in the e-mail, $\hat{p}_{ijk}$. The complement of these predictions yields the likelihood that the link was not clicked, $1 - \hat{p}_{ijk}$. Under the assumption of independence, the product of these link-level nonclick likelihoods yields the joint probability that the e-mail was not clicked. The complement of this probability, $\hat{p}_{ij} = [1 - \Pi_{k=1}^{K}(1 - \hat{p}_{ijk})]$, yields the probability that at least one link was clicked. The independence assumption seems plausible, as it predicts 649 e-mail clicks compared with 639 e-mail clicks in our data. The link-level estimated probabilities of click-through can be compared with a classification threshold to predict the response on any given observation in the validation data. For example, a classification threshold of .5 means that if the estimated clicking probability for an observation is greater than .5, the observation is classified as a click; otherwise, it is classified as a nonclick. Similarly, the e-mail–level probabilities can be compared with a classification threshold to predict whether a given e-mail will elicit a click from a given user.

We use receiver operating characteristic (ROC) curves to compare the predictive performance of our models for a range of thresholds. An ROC analysis is used in psychology and medical statistics for signal detection purposes (e.g., Egan 1975; Metz 1986; Swets 1979). It illustrates the tradeoffs between two types of errors. For any threshold that is used for predicting observations as clicks or nonclicks, two types of errors can occur. False negatives are errors that occur when actual clicks are predicted as nonclicks. False positives are errors that occur when nonresponses (i.e., nonclicks) are predicted as responses (i.e., clicks). The proportion of actual click-throughs that are predicted as nonclicks is called the false negative fraction (FNF) and the proportion of actual nonclicks that are predicted as clicks is called the false positive fraction (FPF).

An ROC curve (also called a Lorenz diagram) is constructed by plotting the true positive fraction (1 – FPF) against the FPF for a range of possible classification thresholds. The resulting plot is represented over a unit square. Different points on the ROC curve correspond to different classification threshold values used for prediction. The area under the ROC curve A_z provides a summary measure of the quality of the model. A model with an ROC curve that tracks the 45-degree line would be worthless, as it would not separate the two classes of observations at all (i.e., there would be as many false positives as true positives). Such a curve has $A_z = .5$. In contrast, a perfect model would have an ROC curve that follows the two axes and would have $A_z = 1$. The ROC curve, as it spans across all possible thresholds, yields a more complete picture of predictive accuracy than measures predicated on a single classification threshold (e.g., a hit rate using a 50% classification threshold).

Figure 1 compares the ROC curves for the three models using the Extant data set. The top panel shows the ROC curves based on the link-level probability estimates, and the bottom panel shows the curves obtained from the e-mail–level probabilities. In both panels, we observe that the two models with heterogeneity (Model DP and Model N) have similar predictive abilities. Moreover, the ROC curves for the two heterogeneous models dominate the ROC curve for the nonheterogeneous model (Model S) at all values of the classification threshold.

Comparing the areas under the link-level ROC curves, we find that $A_z = .85$ for Model DP, $A_z = .84$ for Model N, and $A_z = .76$ for Model S. For the e-mail–level ROC curves, $A_z = .81$ for Model DP, $A_z = .80$ for Model N, and $A_z = .74$ for Model S. These results suggest that for our data set, Model DP performs slightly better than Model N and that both substantially outperform a model with no heterogeneity.

For the Novel data set, the conclusions are similar. We find that $A_z = .77$ for Model DP, $A_z = .75$ for Model N, and $A_z = .68$ for Model S. The e-mail–level ROC curves also provide similar conclusions. Specifically, $A_z = .69$ for Model DP, $A_z = .66$ for Model N, and $A_z = .56$ for Model S. As with the Extant data set, Model DP predicts slightly better than Model N, and both models are superior to the nonheterogeneous model. These findings reinforce the importance of modeling heterogeneity for customization purposes.

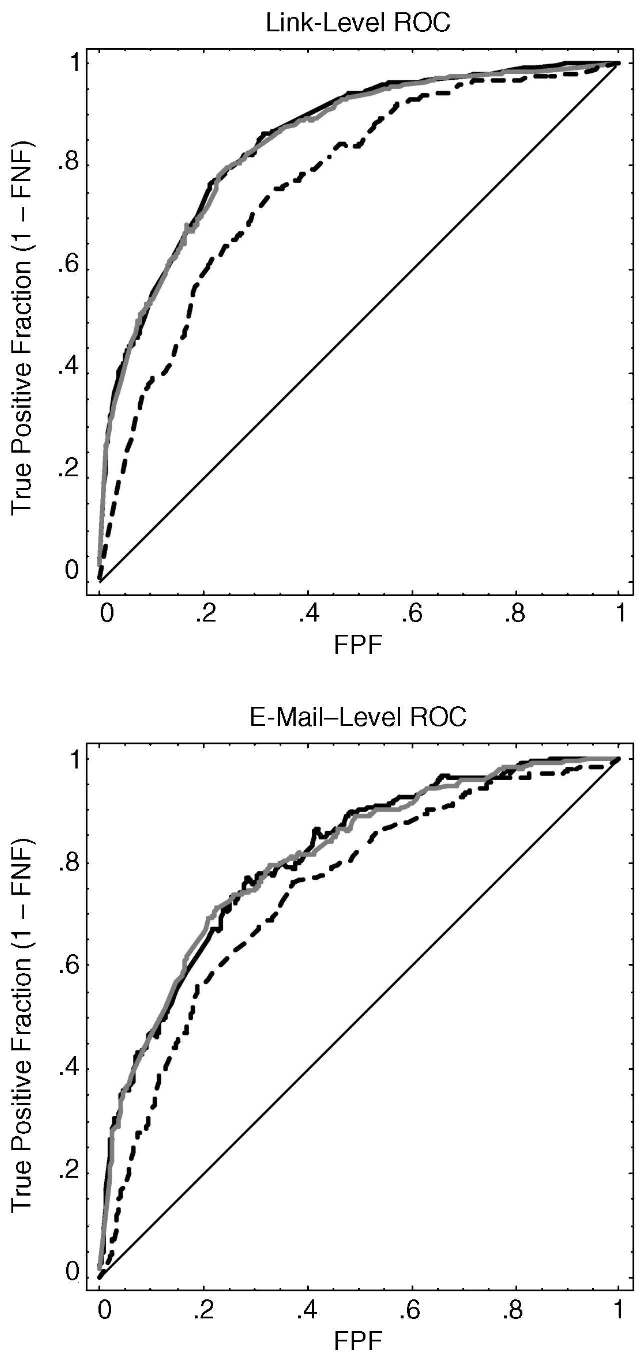
In summary, the PsBF and the ROC analysis indicate that the models that account for heterogeneity are preferable to the model that ignores sources of difference in parameters. That unobserved heterogeneity matters indicates that there could be sufficient gains from customization. Moreover, we find that Model DP is superior to Model N, on the basis of PsBF, and evidences slightly better predictive performance on the validation data.

Parameter Estimates

We now report the parameter estimates for Model DP. As reported previously, these estimates are based on a post–burn-in sample of 20,000 MCMC draws. Table 2 reports the parameter estimates, which we discuss next.

Design variables. The response rates for links within text e-mails appear to be no greater than the response rates for links within html e-mails. As anticipated, the effect of link order is negative, indicating that the effectiveness of links decreases as the link appears later in the e-mail. In contrast to our expectations, we observe no population effect for the number of links within an e-mail. In our data, the number of links never exceeded eight, and this may have been too few to generate a negative effect of clutter. We find that there is

Figure 1
ROC CURVES FOR LINK-LEVEL AND E-MAIL–LEVEL PROBABILITIES



Notes: The dark solid curve pertains to Model DP, Model N is represented by the gray curve, and the nonheterogeneous model is depicted by the dashed curve.

considerable design heterogeneity; users react differently to different designs. Thus, the number of items and the link position can influence the responses of at least some consumers in our sample. The magnitude of the e-mail heterogeneity also indicates that some e-mails evidence a greater effect of link order than others. The finding that some respondents are more likely to respond to text and others respond better to html may reflect the influence of bandwidth. Because the html is bandwidth-intensive, recipients with lesser bandwidth may prefer text-based e-mails.

Content variables. At the population level, content types differ in their ability to elicit click-throughs. Moreover, the standard deviations associated with the normal baseline distribution for the unobserved user random effects clearly reveal that there is considerable user preference heterogeneity in the data. People differ in their preferences for different content types. There is a sizable degree of heterogeneity in content preference across e-mails. This heterogeneity presumably arises from editorial and design differences across the e-mails (e.g., a review in one e-mail may be of more interest than a review in another e-mail because of how it interacts with unobserved contextual variables). When the fixed and random effects are into account, it is clear that the content variables play an important role in predicting click-throughs.

Heterogeneity. The Since variable has a negative impact on response. This suggests that the greater the duration since the previous click, the less likely it is that a user will click on a link within the e-mail. The precision parameters α_1, α_2, and α_3 associated with the Dirichlet process priors suggest that there is greater clustering in the random effects than is evidenced by normal population distributions. For example, $\alpha_1 \approx 103$ implies an average of 61 clusters for the users. Similarly, $\alpha_2 \approx 115$ and $\alpha_3 \approx 383$ imply on average 65 clusters for the e-mails and 383 clusters for the link random effects. This, coupled with the PsBF favoring the DP model, indicates that the population distributions deviate from normality and justifies the need for a semiparametric approach.

It is informative to compare the different sources of heterogeneity. A variance decomposition of the random terms in the utility function shows that user heterogeneity accounts for 28.37%, the e-mail heterogeneity accounts for 38.99%, the link heterogeneity accounts for 1.85%, and the residual errors account for 31.77% of the total random variation. This implies that substantial improvements in model performance can be realized if multiple sources of heterogeneity are modeled.

CUSTOMIZED E-MAIL DESIGN

The customer- and e-mail–specific parameter vectors obtained from the statistical model are used to forecast potential customer reactions to proposed changes in e-mail content and configuration. These forecasts can be used to determine the optimal design and content of an e-mail for each customer. Given an objective function, combinatorial optimization is used to customize the e-mail design for each e-mail and each user.

In e-mail marketing situations, management's primary goal is to maximize the expected number of click-throughs within e-mails. A secondary objective might be to maximize the probability that at least one link is clicked within an e-mail. Given the objective function, design optimization involves (1) selecting from the set of available content the specific content (links) to be included for each person and each e-mail and (2) configuring the e-mail layout to maximize objectives. Note that in our sample, even though the number of links within the e-mail (Num-Items) does not have an effect at the population level, there is considerable

Table 2
PARAMETER ESTIMATES FOR MODEL DP

Variables	*Fixed Effects* μ	*95% Probability Interval*	*Standard of Base Distribution Across Users*	*Standard of Base Distribution Across E-Mails*
Intercept	–1.47	(–2.97, –.59)	.51	.45
Content1	.25	(–.48, .79)	.97	.44
Content2	1.07	(.57, 1.47)	.34	.52
Content3	.21	(–.54, .71)	.65	.47
Content4	.29	(–.56, .72)	.35	.71
Content5	1.21	(.21, 1.62)	.93	.52
Content6	.79	(.13, 1.25)	.38	.57
Content7	–.67	(–2.45, .19)	.49	.72
Content8	.80	(–.11, 1.43)	.38	.52
Content9	.28	(–.95, .93)	.45	.56
Content10	.32	(–.59, .89)	.49	.62
Content11	–1.20	(–3.38, .27)	.54	1.45
Position	–.37	(–.59, –.19)	.23	.27
Since	–.24	(–.59, –.19)	.17	.28
Num-Items	.01	(–.17, .13)	.18	
Text	.29	(–.33, .65)		
α_1	103.32	(69.37, 130.87)		
α_2	114.25	(77.26, 144.88)		
α_3	383.07	(319.93, 431.89)		

Notes: The parentheses contain the 2.5th and the 97.5th percentiles.

heterogeneity at the user level, and therefore content selection can be important for at least some users.

Let n be the total number of content links available to be included in a particular e-mail. For the aforementioned objective functions, the design problem of selecting from the n available links and then ordering the included links can be solved in two stages. In the first stage, n linear assignment subproblems are solved. We index each of these n subproblems by k and let $k \in \{1, \ldots, n\}$. In the second stage, the solution for the first-stage subproblem that yields the largest objective function is chosen as the optimal solution for the original problem.

In the first stage, let us consider the kth subproblem of assigning the n available links to k contiguous positions within an e-mail. Let x_{ij} be a binary variable that is equal to 1 when content link i is present in ordinal position j (j = 1 for the first position and k for the lowest position) within an e-mail, and $x_{ij} = 0$ otherwise. Similarly, let $p_{ij|k}$ be the probability that the user clicks on link i if it is placed in position j, when the total number of included items is k. When the interest is to maximize the expected number of click-throughs, the problem of assigning n links to k positions is given by the following linear assignment specification:

$$\text{Maximize} \sum_{i=1}^{n} \sum_{j=1}^{k} p_{ij|k} x_{ij}$$

subject to

$$\sum_{j=1}^{k} x_{ij} \leq 1, \text{ for } i = 1, 2, \ldots, n$$

$$\sum_{i=1}^{n} x_{ij} = 1, \text{ for } j = 1, 2, \ldots, k.$$

The first constraint in this specification ensures that each content link i can be in at most one of the k destinations. The second constraint ensures that each of the k destinations (ordinal position) can have at most one link. Each of the n subproblems, corresponding to k = 1 to n can be solved by means of the Hungarian method (Foulds 1984, pp. 72–76), which provides an efficient and simple approach. The solution to the kth subproblem yields an assignment x_k^* of the content links to k contiguous positions and the optimal value of the objective function V_k. In each problem, where $k \neq n$, n – k links are left unassigned and are therefore not included in the e-mail.

In the second stage, the solutions to the n subproblems are compared to determine the subproblem that yields the maximum value for the objective function. That is, the subproblem m is determined such that $V_m = \max\{V_1, \ldots, V_n\}$. The solution x_m^* is chosen as the solution of the original problem. This yields the optimal content and configuration of the e-mail.

Similarly, when the interest is to maximize the probability of at least one click from an e-mail, the kth second stage subproblem of assigning n links to k positions can be written as

$$\text{Minimize} \sum_{i=1}^{n} \sum_{j=1}^{k} \log(1 - p_{ij|k}) x_{ij}$$

subject to

$$\sum_{j=1}^{k} x_{ij} \leq 1, \text{ for } i = 1, 2, \ldots, n$$

$$\sum_{i=1}^{n} x_{ij} = 1, \text{ for } j = 1, 2, \ldots, k.$$

Minimizing the objective function in this specification is the same as maximizing the probability that at least one link is clicked, $1 - \Pi_{ij}(1 - p_{ij|k})^{x_{ij}}$, and therefore the problem is of the linear assignment type. The optimal solution can be

Table 3
EXTANT OPTIMIZATION RESULTS

	Probability (At Least One Click)		*Expected Number of Clicks*	
Configurations	*Model DP*	*Model N*	*Model DP*	*Model N*
Original	.23	.23	.34	.32
Greedy	.34	.36	.51	.51
Ordering	.35	.36	.53	.52
Optimal	.36	.39	.55	.57

ascertained as previously, by comparing the n first-stage subproblems in terms of the objective function.

Optimization Results

In this section, we report the optimization results, which are based on the validation data. In particular, we compare and contrast the results from the optimization procedure detailed previously (Optimal) with those obtained using two suboptimal but simpler procedures. The first suboptimal procedure (Ordering) uses a single linear assignment algorithm (instead of n) for each user merely to reorder the existing links within an e-mail. This procurement therefore ignores the content-selection aspect of the optimization and is expected to do well when clutter does not significantly influence the probability of click-through. The second suboptimal procedure (Greedy) uses a greedy heuristic to reorder the links. In this heuristic, links are assigned sequentially from the first position within the e-mail to the last position. To determine which link resides in which position for a given user, the link having the highest probability of click-through (among the set of unassigned links) is assigned to the highest (uppermost) remaining position. Each algorithm yields a customized solution for each user and e-mail, predicated on the link-level clicking likelihoods. We report the results for each validation set described in the "Data" section.

Extant validation data. Table 3 reports the optimization results for the two objective functions and the three optimization procedures using the estimates from Model DP and Model N. Columns 1 and 2 display the results when the objective is to maximize the probability of at least one click from an e-mail. The entries in the first two columns contain the mean probability of at least one click across all the e-mails in the validation sample. Columns 3 and 4 report the results when the objective is to maximize the expected number of click-throughs in an e-mail. The entries in these columns give the mean of the expected number of clicks across all the e-mails in the validation sample. Columns 1 and 3 present the results for the Model DP, and Columns 2 and 4 indicate the Model N results. As the Model DP and Model N results are similar, we discuss only the Model DP results.

The first row of Table 3 gives the predicted results when the original configuration within the data is e-mailed and therefore serves as a benchmark for assessing the performance of the various optimization approaches. For example, Table 3 shows that mean probability of at least one click from an e-mail is .23 when the original e-mails (as designed by the site) are sent. The second row gives the predicted results for the Greedy procedure, the third row gives the results arising from the Ordering procedure, and the last row gives the results using the Optimal two-stage procedure. Thus, Table 3 indicates that, for Model DP, the mean probability of at least one click can increase to .36 if the optimal e-mails are sent.

Comparing the entries in the first column, we find that Optimal procedure can yield a 56% improvement over the original configuration in the mean probability of at least one click-through. In contrast, the improvement for the Ordering procedure is 52%. Decomposing the total potential improvement into the portion arising from reordering and the portion arising from content selection suggests that reordering results in 92% of the total improvement, and content selection constitutes the balance. Further analysis suggests that the Optimal algorithm improves the likelihood of at least one click over the Ordering algorithm for 43% of the e-mails sent out in the validation sample. A majority of the gains are small and arise from the users who react adversely to clutter (i.e., have a negative coefficient for the variable number of items). Were users more averse to clutter, content selection would matter even more.

The Greedy procedure results in a 48% increase in predicted e-mail click rates (from .23 to .34). Thus, the algorithm performs nearly as well as the Ordering algorithm. Nonetheless, we are wary of predicting similar improvements in different data. In particular, when subjects have a positive coefficient for the positive variable (i.e., they tend to scroll to the bottom of the e-mail and then click on one of the links at the bottom), the Greedy algorithm we use will not do well.

The entries in the second column show that the Optimal optimization procedure can yield a 62% improvement over the original configuration in the expected number of clicks. In contrast, the improvements for the Ordering procedure and the Greedy algorithm are 53% and 50%, respectively. Furthermore, for approximately 42% of the e-mails sent out in the validation sample, Optimal made an improvement over Ordering. Similarly, for 58% of the e-mails, Optimal was better than Greedy, though the magnitude of improvement was small in most cases.

The Optimal procedure is better than the other two procedures and leads to improvements in response rates for e-mails, especially when clutter adversely affects the probability of clicking. In addition, the linear assignment problem that forms the basis for the optimal procedure can be solved quickly and efficiently, even for large problems. The Greedy solution, though simpler, performs poorly for users who have a propensity to click at the bottom of the e-mail. In contrast, the Ordering algorithm can do relatively well in this situation yet performs poorly when clutter matters. In general, the Ordering procedure and the Greedy procedure are marginally less demanding computationally. The choice

Table 4

NOVEL OPTIMIZATION RESULTS

	Probability (At Least One Click)		*Expected Number of Clicks*	
Configurations	*Model DP*	*Model N*	*Model DP*	*Model N*
Original	.45	.45	.61	.60
Greedy	.53	.53	.75	.74
Ordering	.53	.53	.75	.74
Optimal	.54	.54	.75	.75

between these depends on the balance among accuracy, speed, and complexity desired by managers.

When applying a similar analysis using the parameters from Model S (the nonheterogeneous model), the optimal two-stage optimization procedure yields a predicted improvement of only 12.5% in the mean probability of at least one click-through and an improvement of 15.4% for the second objective function. This relative lack of improvement predicted by the simpler model compared with the heterogeneous models arises because the simpler model has limited flexibility in differentiating among users and e-mails. It therefore generates e-mail configurations that are optimal only for the average user or e-mail.

Novel validation data. Table 4 reports the optimization results for the Novel validation data set. Column 2 indicates that the mean probability of at least one click increases 20% from .45 when the original e-mails are sent to .54 when the optimal e-mails are sent. The improvement for the Ordering procedure is similar, 18%. Decomposing the total potential improvement into the portion arising from reordering and the portion arising from content selection suggests that reordering results in 90% of the total improvement and content selection constitutes the balance.

For the second objective function, the results obtained from the DP model estimates in the fourth column show that the Optimal optimization procedure can yield a 23% improvement over the original configuration in the expected number of clicks. The improvements for the Ordering procedure and the Greedy algorithm are also 23%. The results for the normal model are similar and are shown in the fifth column. For both objective functions, the optimization results for the Novel data set further suggest that any of the three optimization approaches work equally well in these data.

Model S yields a predicted improvement in the mean probability of at least one click-through of 14.2% and an improvement of 15.8% for expected clicks. However, given that heterogeneous models predict significantly better than the simple model with no heterogeneity, we place greater credence on the optimization results from the heterogeneous models.

Finally, we note that the improvement in hit rates using the Novel data, though substantial, is not as great as the improvement indicated by the Extant data. The difference arises because the information regarding the clicking behavior of other users on a particular link is informative about the targeting of links. When possible, content providers should seek to "test market" information, as this enhances targeting efficacy.

CONCLUSION

The advent of the Internet has enhanced the ability of marketers to personalize communications and engender relationships with consumers. By enabling the right content to reach the right person at the right time, the Web can yield substantial dividends to Web marketers and can enhance the quality of service to consumers. However, this promise is contingent on learning more about consumer preferences and developing techniques that enable marketers to fulfill these preferences. Our objective has been to facilitate that task.

Accordingly, we describe an approach to harness the potential afforded by the Web to determine individual-level preferences, and then we develop an algorithm to customize content predicated on those preferences. In the context of targeting and customizing e-mails that herald content in a Web magazine, we develop a customization system that uses an MDP probit model coupled with an optimization model to personalize communications on the Internet.

Our adaptation of the Dirichlet process model is unique from previous implementations in that it incorporates multiple sources of heterogeneity, uses the probit framework, and is applied to a large-scale data application. In contrast to the normal model, the MDP model predicates a user's choice behavior on that of the user's "nearest neighbors." As such, a comparison of the MDP and normal models provides some insight into the additional predictive power of model-based collaborative filtering. Our model comparison results indicate that the MDP model is preferable to the normal model on the basis of the PsBF. The predictive performance of the MDP model is slightly better than that of the normal model. We leave a thorough comparison of these alternative approaches for further research, but we note that our approach is tailored to the problem, as opposed to the data. By virtue of its flexibility, there may be cases in which the Dirichlet process model substantially outperforms the normal model (i.e., nonnormal heterogeneity). The converse is unlikely to be true, as normal models do not adjust well to nonnormal heterogeneity.

Given that the additional programming demands inherent in the Dirichlet process model (over that of the normal model) are negligible, the trade-off between the approaches is an issue of flexibility and scalability. The computational demands of the normal model are simpler, and therefore we recommend the normal model when scalability of the model is a major concern. Moreover, the scalability of either model to the demands of sending e-mails for many users requires a careful decomposition of the overall requirements into offline and online (i.e., real-time) components. For example, aggregate features of the models, such as the population distribution, can be estimated offline on the basis of a sample of users. Moreover, the population distribution can be updated periodically (say, weekly) as new data arrives. When the population distribution is known, obtaining the estimates for a particular user is not computationally intensive and can be done relatively quickly. The optimization component is quick, as it is based on the linear assignment problem that has quick and exact solutions.

Our approach adds to the targeting and customization literature in marketing by integrating heterogeneous choice models with optimization techniques to personalize content. Specifically, we describe an optimization algorithm based on the assignment algorithm to optimize the design and content of electronic communications. We believe that such a general approach (combining choice models with optimization models; Rossi, McCulloch, and Allenby 1996; Tellis and Zufryden 1995) has utility beyond e-mail customization and can be used in the design of tailored services, custom-designed catalogs, and bundling of goods.

The results of our model indicate that the design of the e-mail is crucial in affecting click-through probabilities. For example, we find that the order of content matters and that there exists a great deal of heterogeneity across users in their preferences and across links and e-mails in terms of their effectiveness in design and content. Capitalizing on these results, we demonstrate that design and content can indeed be optimized. We find that response rates (expected click-throughs) could be increased by 62% if the e-mail's design is customized.

Finally, we propose that our analysis be extended along two dimensions—modeling other behaviors in the Internet environment and using our methodological approach in other contexts. With regard to other Internet applications, it would be desirable to customize Web content in an effort to increase the frequency of site visits and clicks per visit. Similarly, our general approach could be adapted to the design of e-commerce sites and personal agents. It is also possible to use our underlying methodology to target advertising content. Another pressing problem centers around optimal contact strategies for e-mail communications. Excessive contact, or "over-touching," can lead to unsubscribe decisions. Infrequent contact could lead to few responses. Moreover, these effects could vary by user. With respect to other contexts, the proposed model could be extended beyond e-commerce models to more traditional models of targeting purchase opportunities, such as direct mail marketing or product customization. Our design approach could be adapted to conjoint tasks. The conjoint domain may prove especially promising, as affective measures can be integrated with behavioral ones to enhance the predictive capability of the model, and the researcher may have more latitude in the design of the stimulus set. We hope that this analysis will encourage further research along these dimensions.

REFERENCES

Aberdeen Group (2001), "e-Mail Marketing: Relevancy, Retention, and ROI," working paper, [available at http:www.aberdeen.com/abcompany/hottopics/emailmarketing/default.htm].

Alba, Joseph, John Lynch, Barton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer, and Stacy Wood (1997), "Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces," *Journal of Marketing*, 61 (July), 38–53.

Allenby, Greg M., N. Arora, and J.L. Ginter (1998), "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35 (August), 384–89.

——— and P. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89 (1–2), 57–79.

Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000), "Internet Recommender Systems," *Journal of Marketing Research*, 37 (August), 363–75.

———, K. Jedidi, and S. Jagpal (2000), "A Hierarchical Approach for Modeling Heterogeneity in Structural Equation Models," *Marketing Science*, 19 (4), 328–47.

Becker, Holly, Douglas Anmuth, and Stephanie Leichter (2002), "Online Advertising," *Internet & Media Global Equity Research*. New York: Lehman Brothers.

Blackwell, D. and J.B. MacQueen (1973), "Ferguson Distribution via Polya Urn Schemes," *The Annals of Statistics*, 1, 353–55.

Breese, John S., David Heckerman, and Carl Kadie (1998), "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Technical Report MSR-TR-98-12, Microsoft Research (October).

Bush, C.A. and S.N. MacEachern (1996), "A Semi-parametric Bayesian Model for Randomized Block Designs," *Biometrika*, 83 (2), 275–85.

Doss, H. (1994) "Bayesian Nonparametric Estimation for Incomplete Data via Successive Substitution Sampling," *The Annals of Statistics*, 22 (December), 1763–86.

Egan, J.P. (1975), *Signal Detection Theory and ROC Analysis*. New York: Academic Press.

Escobar, M.D. (1994), "Estimating Normal Means with a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89 (March), 268–77.

Ferguson, T.S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–30.

——— (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–29.

Foulds, L.R. (1984) *Combinational Optimization for Undergraduates*. New York: Springer-Verlag.

Geisser, S. and W. Eddy (1979), "A Predictive Approach to Model Selection," *Journal of the American Statistical Association*, 74 (March), 153–60.

Gelfand, A.E. (1996), "Model Determination Using Sampling-Based Models," in *Monte Carlo Markov Chain in Practice*, W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, eds. London: Chapman and Hall, 145–61.

Gershoff, Andrew and Patricia M. West (1998), "Using a Community of Knowledge to Build Intelligent Agents," *Marketing Letters*, 9 (1), 79–91.

Hanssens, Dominique M. and Barton A. Weitz (1980), "The Effectiveness of Industrial Print Advertisements Across Product Categories," *Journal of Marketing Research*, 17 (August), 294–306.

Hoque, Abeer Y. and Gerald L. Lohse (1999), "An Information Search Cost Perspective for Designing Interfaces for Electronic Commerce," *Journal of Marketing Research*, 36 (August), 387–94.

Houston, F.S. and Carol Scott (1984), "The Determinants of Advertising Page Exposure," *Journal of Advertising*, 13 (2), 27–33.

MacEachern, S.N. (1994), "Estimating Normal Means with a Conjugate Style Dirichlet Process Prior," *Communications in Statistics: Simulation and Computation*, 23 (3), 727–41.

Manchanda, P., A. Ansari, and S. Gupta (1999), "The Shopping Basket: A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18 (2), 95–114.

Mela, Carl F., K. Jedidi, and D. Bowman (1998), "The Long-Term Impact of Promotions on Consumer Stockpiling Behavior," *Journal of Marketing Research*, 35 (May), 250–62.

Metz, C.E. (1986), "ROC Methodology in Radiologic Imaging," *Investigative Radiology*, 21 (9), 720–33.

Perkowitz, M. and Oren Etzioni (1997), "Adaptive Web Sites: Automatically Learning from User Access Patterns," in *Proceedings of 6th International WWW Conference*. Seattle: University of Washington.

Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15 (4), 321–40.

Sarukkai, Ramesh (2000), "Link Prediction and Path Analysis Using Markov Chains," *Computer Networks*, 33 (1–6), 377–86.

Satterthwaite, R. (1999), "Survey Results: The Real Cost of E-commerce Sites," Gartner Group Research Note M-07-8931. Stamford, CT: Gartner Group (May 7).

Schmittlein, D.C., D.G. Morrison, and R. Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, 33 (1), 1–24.

Shaffer, G. and Z. John Zhang (1995), "Competitive Coupon Targeting," *Marketing Science*, 14 (4), 395–416.

Swets, J.A. (1979), "ROC Analysis Applied to the Evaluation of Medical Imaging Techniques," *Investigative Radiology*, 14 (2), 109–21.

Tellis, Gerard J. and Fred S. Zufryden (1995), "Tackling the Retailer Decision Maze: Which Brands to Discount, How Much, and When?" *Marketing Science*, 14 (3), 271–99.

Wedel, M. and W. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2d ed. Boston: Kluwer.

West, M., P. Muller, and M.D. Escobar (1994), "Hierarchical Priors and Mixture Models with Applications in Regression and Density Estimation," in *Aspects of Uncertainty: A Tribute to D.V. Lindley*, A.F.M. Smith and P.R. Freeman, eds. London: John Wiley & Sons, 363–86.

Copyright © 2003 EBSCO Publishing