The Role of the Management Sciences in Research on Personalization

B. P. S. Murthi • Sumit Sarkar
School of Management, The University of Texas at Dallas, Richardson, Texas 75080
murthi@utdallas.edu • sumit@utdallas.edu

We present a review of research studies that deal with personalization and synthesize current knowledge about these areas. We identify issues that we envision will be of interest to researchers working in the management sciences, taking an interdisciplinary approach that spans the areas of economics, marketing, information technology (IT), and operations research. We present a framework for personalization that allows us to identify key players in the personalization process as well as key stages of personalization. The framework enables us to examine the strategic role of personalization in the interactions between a firm and other key players in the firm’s value system. We conceptualize the personalization process as consisting of three stages: (1) learning about consumer preferences, (2) matching offerings to customers, and (3) evaluation of the learning and matching processes. This review focuses on the learning stage, with an emphasis on utility-based approaches to estimate preference functions using data on customer interactions with a firm.

(Customization; Choice Models; Internet Marketing; Online Tracking; Learning Consumer Preferences; Recommendation Systems)

1. Introduction

The process of using a customer’s information to deliver a targeted solution to that customer is known as personalization or one-to-one marketing (Peppers and Rogers 1997). A common form is the use of recommendation systems that make product recommendations to a customer based on her prior interactions with the firm. For example, Amazon.com uses several diverse techniques like collaborative filtering and association rule mining to recommend books and gifts to their customers. Another personalization technique targets the right communication message and promotion to the customer. For instance, DoubleClick uses visitor profiles to target appropriate banner advertisements on their clients’ sites, while YesMail specializes in targeting and sending personalized e-mails.

Personalization has become important in Internet-based applications for a number of reasons. First, it can be an important source of competitive advantage in some areas (e.g., through differentiation). Second, because there is an explosion in the number of choices that are available to customers on the Internet, firms can add value by providing appropriate information to simplify the customer’s decision process. Third, the drastic reduction in information technology (IT) costs, coupled with the development of database technologies, has significantly changed the economics of collection, storage, and processing of customers data. The low cost of personalization substantially enhances the ability of firms to deliver customized products and, even more so, for digital products.

In this paper, we present a review of research studies that deal with personalization. We find that the research on personalization is addressed in relative isolation in different fields. We synthesize current knowledge about these areas, and identify issues of interest to researchers working in the management sciences. To bring appropriate context to these issues,
we take an interdisciplinary approach that spans the areas of economics, marketing, IT, and operations research.

The primary motivation for this paper is to identify research opportunities in the context of online personalization, although many of these issues also apply to personalization in traditional brick-and-mortar environments. This paper focuses on personalization as it applies to end consumers, and not to businesses. While the business-to-business segment is large, it is outside the scope of this study.

We begin by presenting, in §2, a framework for personalization that allows us to identify key players and important stages in the personalization process. We examine the strategic role of personalization in the interactions between a firm and other key players in the firm’s value system in §3. In §4, we examine how a firm can learn a customer’s preferences, which is one of the key components of the personalization process. Section 5 presents concluding remarks.

2. A Framework for Personalization
The framework, a modification of Brandenburger and Nalebuff’s (1995) Value Net approach, serves two purposes. First, it enables us to examine the strategic effects of personalization in the interactions between a firm and other key players. We review the literature on the strategic behavior of firms, and discuss opportunities for analytical and empirical research in this regard. Second, it helps us to examine the important stages in the personalization process and understand future research issues in implementing personalization (see Figure 1).

In the enhanced Value Net approach, a firm interacts with customers and suppliers in the vertical dimension, and with competitors and complementors in the horizontal dimension. Typically, transactions occur in the vertical dimension, with personalized products and services flowing from firm to customers (in some cases, through channels or intermediaries). Customer information, the critical input for personalization, flows in the reverse direction. In understanding strategic issues related to a firm’s personalization strategy, it is important to consider the role and influence of competitors, complementors, suppliers, and channels. Several papers have dealt with the effects of personalization that include product differentiation, price discrimination, first-mover advantage, and bundling. These topics have been explored under both monopoly and duopoly situations.

Figure 1   The Enhanced Value Net
We view the personalization process itself as consisting of three main stages: learning, matching, and evaluation. In the learning stage, a firm collects data on its customers and uses the data to learn about the customers’ preferences and tastes. In the matching stage, the firm uses the knowledge of customer preferences to develop offerings that best satisfy their preferences and to target these to the appropriate market segment. The last stage consists of evaluating the effectiveness of learning and matching efforts in providing meaningful personalization to the firm’s customers.

The learning process starts with the collection of data from customers. The data may be explicitly provided by the customer or inferred from the customer’s interactions at a firm’s site. We examine methods for the explicit collection of data from online customers, and their effectiveness, and also the methods of inferring a customer’s preferences from a variety of customer interactions. A number of well-established techniques in the management sciences have been employed in traditional (offline) environments for modeling customer needs. We discuss how these existing techniques could be adapted for use in the online environment, and we detail the need for new techniques.

In the next stage of the personalization process, customers are matched with desired products, targeted communications, and even personalized prices based on their preferences. Such matching is performed by recommendation systems that usually employ rule-based, collaborative filtering, and content filtering techniques. The final stage, evaluation, requires the development of appropriate metrics for assessing the effectiveness of a personalization program. Although important, the last two stages of personalization (matching and evaluation) are outside the scope of this survey.

3. Personalization and Firm Strategy
In the following discussion, we examine the role of personalization for each of the interactions represented in the Value Net framework. First, we look at the interaction between a firm and the consumer under both monopoly and competition. Next, we assess the impact of personalization on a firm’s interaction with its suppliers, complementors, and intermediaries.

3.1. Firm, Customer, and Competitors
We discuss emerging research and highlight the need for models that consider strategic responses of customers, privacy and trust, differentiation through personalization, price discrimination, and bundling. Game theory is the dominant paradigm in most of the papers.

Strategic Customer Behavior. The most important strategic consideration between a firm and its customers is the bargaining power of the customer. The Internet reduces information asymmetry between the consumer and the firm, leading to an increase in buyer power relative to the firm (Afuah and Tucci 2001, p. 30). Effective personalization strategies can help shift the power back in favor of the firm. In this interaction, a customer may behave strategically to provide only certain kinds of information, and may even selectively distort information to benefit himself. This interplay provides interesting research issues. For instance, is it profitable for a monopolist to provide personalization when customers act strategically? Under competition, when will personalization be profitable to a firm?

Shaffer and Zhang (2000) show that personalized pricing can benefit firms under competition. In their model, customers are heterogeneous in their preferences and do not act strategically. In contrast, Villas-Boas (2001) shows that in the presence of strategic customers, a monopolist who gives incentives to new buyers could be worse off for offering promotions selectively. This is because strategic customers could forego purchases in the first period to avoid being recognized as an “existing” customer in the second period. Chen and Zhang (2001) extend the above analysis to competition and show that targeted pricing is profitable even in the presence of strategic customers. This research demonstrates the need for modeling strategic customer behavior in understanding personalization effects.
Privacy and Trust. A firm would like to obtain as much information as possible about a customer before engaging in a transaction with her. The customer, on the other hand, would like to obtain personalized service by providing minimal information. In particular, the customer would not like to provide information that would reveal her reservation price for the product. The firm should, therefore, provide incentives (monetary or nonmonetary) to the customer to share this information. Alternately, it could impose sharing information as a mandatory requirement for receiving recommendations (Resnick and Varian 1997). Further, as earlier stated, users may deliberately manipulate personal data to benefit themselves. The effectiveness of alternate mechanisms for eliciting personal information needs examination. In this context, it is relevant to study the role of trust and methods of establishing trust (Resnick and Varian 1997).

Currently, the market for personal information is based on the notion that the institution that has gathered the information also owns the information (Laudon 1996). Laudon suggests the possibility of creating a “national information market” in which information about individuals is bought and sold at a market-clearing price. In such a market, an individual would have the ability to grant to institutions the right to use their personal information for a predetermined period of time and specified nature of use. On the other hand, Rust et al. (2002) posit that, with declining costs of collecting personal information, privacy will be harder to protect and a market for privacy will emerge where customers can purchase their ideal level of privacy. Research is needed on the mechanism design of such markets and the pricing of customer information. How will competition in a given industry affect pricing in such markets? What is the role of the government in monitoring such information markets?

Product Differentiation. Personalization of product attributes can enable firms to better differentiate their products and earn higher profits in equilibrium (Shaffer and Zhang 2000). In a competitive scenario, Shaked and Sutton (1982) have shown that product differentiation leads to reduced price competition in equilibrium due to better extraction of consumer surplus. However, with personalization, Ulph and Vulkan (2000) show that there exists an additional “enhanced competition effect” because firms are competing for smaller and smaller segments of consumers, and this effect offsets the benefits of differentiation through surplus extraction. They show that when consumers are relatively homogeneous in taste, the competition effect dominates the surplus effect, thereby making firms worse off with personalized pricing. The model assumes that both firms are symmetric in that they both have full information about consumers. It is important to examine the strategic effects when one of the firms is superior to its competitors in terms of personalization technology. This allows one to separate differentiation in “products” space and “information” space. Syam and Dellaert (2002) model this interaction and find that, in equilibrium, firms will pay for customer information. Further, the payment can act as a market signal of quality. They also show that such payment will make it unattractive for firms to customize in all situations. In these models, the firm’s decision to engage in personalization is treated as exogenous. Future research could endogenize the firm’s decision to personalize product attributes.

The literature on differentiation (Tirole 1988) discusses two types: vertical (based on quality valuations), and horizontal (based on taste variations). For example, personalized pricing represents vertical differentiation, while personalization of product attributes represents horizontal differentiation. A study of the strategic impact of both types of differentiation within a durable goods oligopoly is given in Desai (2001). He examines a firm’s product-line strategies when customers differ in both quality valuations and tastes. He identifies optimal price and quality levels in equilibrium that reduce the cannibalization effect between high- and low-quality products. This highlights the need to consider both types of differentiation effects in personalization.

Firms have an incentive to develop multiple variants of a product to satisfy the market’s need for variety. The low cost of personalization enhances the ability of firms to offer more variants, especially for digital products. We need to understand the conditions that lead to proliferation of product variants. Alternately, when will firms offer only one version?
Bhargava and Choudhary (2001) present conditions when firms will offer a single version for information goods even with zero marginal costs.

The strategic timing of adoption of personalization by competing firms—whether sequential or simultaneous—has implications for differentiation. Dewan et al. (2000a) show that when firms in a duopoly simultaneously adopt customization, there is reduced differentiation and, hence, greater price competition. However, customization allows firms to charge higher prices, which offsets the effect of price competition. In their model, Dewan et al. (2000a) assume that firms incur an additional cost to customize their products. They show that when one firm adopts a customization strategy, it is able to improve its market share and profits at the expense of other firms. However, it then becomes optimal for other firms to also adopt customization, which, in turn, leads to excessive investments in customization, leading to lower profits for all the firms. When firms adopt personalization sequentially, there is an advantage to the early adopter (Dewan et al. 2000b). Further, they show that by heavily investing in personalization, a firm can deter entry of potential rivals.

When competing firms provide free products, the effect of differentiation on price competition is not relevant. What, then, is the effect of personalization in this context? Further, differentiation can be affected through different types of personalization (e.g., product features, content, or prices). The type of personalization that leads to maximal differentiation in different situations needs to be explored.

**Price Discrimination.** Personalization allows firms to inexpensively estimate their customers’ valuations and, hence, enables finer price discrimination. A taxonomy commonly used for price discrimination considers three types (Pigou 1932). When a firm is able to charge different prices to different customers (e.g., auctions), it is termed first-degree price discrimination. A firm engages in second-degree price discrimination when it makes available a set of related offerings with fixed prices associated with each, and customers choose the product that best fits their needs. Applications of second-degree price discrimination include product-line pricing and versioning (Varian 2001b). In third-degree price discrimination, firms charge different prices to different segments.

In traditional markets, first-degree price discrimination is not a practical approach, because it is quite expensive or sometimes impossible for a firm to gauge an individual consumer’s willingness to pay. With access to customer data and the tools to analyze data in real time, Internet firms are able to estimate individual customer valuations accurately and cheaply. This enables first-degree price discrimination. Brynjolfsson and Smith (2000) found that Internet retailers change their prices more frequently and in smaller increments than traditional retailers, thus, demonstrating the potential of the Internet for finer price discrimination.

The following papers explore the link between price discrimination and the decision to provide personalization. Ulph and Vulkan (2001) study the interplay between a firm’s ability to provide mass customization and the choice of first-degree price discrimination in a duopoly using Hotelling’s framework. A firm that can provide mass customization can employ either first- or second-degree price discrimination. The authors show that a firm is always better off using first-degree price discrimination if it also mass-customizes and vice versa. Dewan et al. (1999) explore the incentives for a monopolist engaging in second-degree price discrimination to develop more customized products vis-à-vis standardized products. They show that a decrease in personalization costs leads to greater customization at the expense of standardized products. Research in second-degree price discrimination has established that a firm should offer multiple versions when consumers have linear valuations and marginal costs are convex in quality (Salant 1989, Spence 1980, Stokey 1979). Bhargava and Choudhary (2001) generalize these results to nonlinear valuation functions and nonconvex marginal cost functions. Kannan and Jain (2003) present a model for the pricing of multiple versions of digital content.

Promotions can be viewed as price discrimination devices (Narasimhan 1984). Personalized promotions are seen as being more profitable to firms because the discount is selectively given only to price-sensitive customers. Some researchers have challenged this
assumption and demonstrate that one-to-one promotions by symmetric competing firms can lead to lower profits to all firms (Shaffer and Zhang 1995, Fudenburg and Tirole 2000). Shaffer and Zhang (2002) examine a scenario where firms are asymmetric with respect to size. They find that while there is increased price competition, the firm with a larger market share and a more loyal customer base can be better off when both firms offer targeted promotions, as compared to when neither does so. In the context of Internet coupons, Cheng and Dogan (2001) show that coupons should not be provided to all customers or the firm will lose the advantages of price discrimination. Firms should either impose a cost to acquire coupons online (through registration or printing coupons) or engage in targeted coupons.

When customers find it costly to provide information about their preferences to firms, they may be unwilling to register with multiple firms. Models of personalization should consider such costs to the customer, which could lead to lock-in effects. How will switching costs affect market equilibrium? The results will depend on whether switching costs are known to the firm or not. What factors impact the ability of a firm to use personalization as a lock-in strategy? Liebowitz and Margolis (1990) suggest that one is likely to observe greater personalization in relatively slower-growth markets.

Haggling is another mechanism for price discrimination. Desai and Purohit (2002) find that when segments differ in their costs of haggling and competing firms choose the same price format, haggling is more profitable than a fixed-price policy. When the proportion of nonhagglers in the market is low, firms will choose a fixed policy in equilibrium. Online haggling as a mechanism for personalized pricing is a novel idea that needs further research.

Similarly, models should consider the firm’s costs of personalization so as to endogenize the decision to personalize. Chen and Iyer (2002) model a firm’s decision to engage in customized pricing in a competitive scenario. They show that, in equilibrium, both firms will invest in personalization if customers are sufficiently heterogeneous (i.e., market differentiation is high) and the cost of personalization is high. However, when market differentiation or costs are low, a firm may choose not to invest in personalization. A firm’s decision to personalize could also depend on the size of its customer base, the depth of information available on each customer, and customer loyalty.

**Bundling.** One form of personalization is the bundling of products (especially content) based on consumer preferences. The Internet allows bundling of content at low marginal costs and could lead to large-scale bundling. Bakos and Brynjolfsson (1999) show that large-scale bundling can create “economies of aggregation,” which favor large aggregators even if network externalities and economies of scale are ignored. This raises interesting possibilities for aggregation of traditional products with information goods, leading to multimarket multiproduct bundling.

When providing a bundle, is it better for a firm to design the bundle or for customers to create their own bundle? Chuang and Sirbu (1999) show that allowing customers to self-select a bundle can often improve a firm’s outcome. Hitt and Chen (2001) extend this stream of work to show that for a monopolistic setting, such a mechanism outperforms individual selling and pure bundling when marginal costs of providing the goods are greater than zero and customers have heterogeneous preferences. The above model can be extended from a monopoly to a competitive situation.

**Other Issues.** (1) Relative importance of personalization: Firms can differentiate by using other strategies, such as developing a strong brand or partnering with highly-visible companies. When does personalization become a significant source of differentiation relative to other alternatives? What kinds of interaction effects exist between the multiple sources of differentiation? For instance, does personalization enhance or diminish the effect of branding?

(2) The role of shopbots: Shopbots are software agents that perform information-gathering tasks for either the buyer or the seller. How will the strategic effects of personalization change in the presence of shopbots?

(3) Sharing of customer information between competitors: When will firms share information to improve targetability? Chen et al. (2001) explore conditions under which sharing will occur.
(4) Dynamic collection of preference data: Personalization can change over time depending on changes in preference data. While most models do not account for dynamic personalization, Raghu et al. (2001) develop a model for dynamic profiling of customer preferences.

3.2. Firm, Suppliers, Complementors, and Channels
The impact of personalization on a firm’s strategy could, in turn, affect its partners. For instance, personalization leads to proliferation of product variants, which has implications for supplier strategy. How does product proliferation affect a firm’s sourcing strategy—single sourcing versus multiple sourcing? How does it affect supplier’s inventory? At present, there is little exploration of such issues.

An important set of issues revolves around a firm’s willingness to share customer information with its partners (termed vertical information sharing). If the intermediary provides personalization to its customers, how willing will it be to share information about its profitable customers with the parent firm? How can firms affect their partner’s willingness to share information? Sharing private information about downstream demand or costs affects the relative bargaining power between the players and would, in general, discourage information sharing. Li (2002) examines an additional leakage effect of vertical information sharing, which occurs when competing firms infer information about each other based on their actions with a common supplier. He shows that the leakage effect can discourage the sharing of demand information.

When a product has two complementary components (e.g., hardware and software), each being produced by a different firm, there are several questions of interest. How does personalization for one product affect demand for the other product? Each player would like to promote its own product while retaining the ability to recommend products from multiple complementors. What are the mechanisms for aligning a firm’s, and its complementor’s, ability to personalize? How would each player personalize to strengthen bargaining power relative to a complementor? How should the additional consumer surplus extracted by the complements be optimally shared? If versioning is more costly for one firm than the other, what impact should this have on revenue sharing? How do customization capabilities drive the choice of a complementor? When would a firm want exclusive rights over its complementor’s products for effective personalization?

When firms sell their products through their intermediaries, both can engage in personalization. When is it advantageous to the firm to have a direct contact with the customer and when should its intermediary be allowed to do the personalization? Does the nature of the product (e.g., products versus services) affect this decision? The intermediary possesses greater power through first-hand knowledge of customers in traditional retail environments. On the Internet, since it is inexpensive for a firm to maintain direct customer contact, such retailer power is expected to diminish. For effective personalization, when is vertical integration a desirable strategy?

Another research issue relates to channels as a coordinating mechanism. For example, an electronic mall can track visitors’ movements across all storefronts, and make that information available to participating stores. When would it be worthwhile for the marketplace to engage in provisioning this type of personalization services? How should the personalization provider charge for their services? What are the implications to firms that do not participate? (Stated differently, what kinds of firms would prefer to not participate?) What impact does it have on the customer, and what are the privacy implications in this context?

In summary, there are several open research questions regarding the role of personalization in a firm’s strategic behavior. Different models in the personalization context could consider different combinations of factors such as monopoly or competition, symmetric or asymmetric firms, horizontal or vertical differentiation or both, and whether the choice of personalization is exogenous or endogenous. In addition, models could consider the costs of personalization to a customer, and whether a customer behaves strategically. Further, we found no research on empirical validation of the assumptions or the results of analytical models providing substantial opportunities for researchers. Table 1 summarizes key research issues.
Table 1  Impact of Personalization on Firm Strategy

<table>
<thead>
<tr>
<th>Interactions</th>
<th>Subtopics</th>
<th>Research questions</th>
<th>Representative articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm, customer, and competition</td>
<td>Strategic customer behavior</td>
<td>What are the benefits of personalization in the presence of a strategic customer?</td>
<td>Chen and Zhang 2001, Villas-Boas 2001</td>
</tr>
<tr>
<td></td>
<td>Privacy and trust</td>
<td>What incentives are necessary to (a) get information from customers, and (b) reduce distortion of information? Is a “national information market” viable?</td>
<td>Laudon 1996, Resnick and Varian 1997, Rust et al. 2002</td>
</tr>
<tr>
<td></td>
<td>Product differentiation</td>
<td>Under what conditions would a firm offer multiple versions for better personalization?</td>
<td>Dewan et al. 1999; Ulph and Vulcan 2000; Dewan et al. 2000a, b; Varian 2001b; Desai 2001; Shaffer and Zhang 2002</td>
</tr>
<tr>
<td></td>
<td>Bundling</td>
<td>When are dynamic pricing strategies viable?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other issues</td>
<td>When is personalization a significant source of differentiation?</td>
<td>Farrell and Shapiro 1989, Chen et al. 2001</td>
</tr>
<tr>
<td>Suppliers, complementors, and intermediaries</td>
<td>Product proliferation</td>
<td>How does product proliferation affect a firm’s sourcing strategy?</td>
<td>Varian 2001b</td>
</tr>
<tr>
<td></td>
<td>Vertical information sharing</td>
<td>What are the incentives for sharing customer information with its partners?</td>
<td>Li 2002</td>
</tr>
<tr>
<td></td>
<td>Complementors</td>
<td>What are the mechanisms for aligning a firm’s, and its complementor’s, ability to personalize?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vertical integration</td>
<td>For effective personalization, when is vertical integration desirable?</td>
<td></td>
</tr>
</tbody>
</table>

4. The Personalization Process—Learning

The process of personalization can be broken down into three stages: learning, matching, and evaluation. Learning involves data collection and inference from data. In the next subsection, we explore the research issues pertaining to data collection. Then, we examine the process of inference from the data about customer preferences. While published research on learning in the personalization context is scant, many established models in marketing are useful. We provide a quick review of such models and discuss issues that need to be addressed.

4.1. Data Collection

Data can be collected by either asking the customer using online surveys and registration forms directly or tracking the customer’s interactions with the website/firm. The firm’s data can be supplemented with data from external sources such as credit card companies and credit rating agencies. Customer relationship management software facilitates the collection of data from multiple touch points between a firm and its customers, and provides analysis tools to create a unified view of the customer.

4.1.1. Directly Asking the Customer. Ideally, firms need data at an individual level to understand a
customer’s preferences and provide personalization. Website registration forms generally collect minimal information about a customer: name, address, telephone number, and e-mail address. This data by itself permits relatively little personalization (e.g., some targeting is possible based on residential location). To more effectively personalize, firms need information on a customer’s demographics and preferences and often use online surveys. For instance, to get an e-mail account at Hotmail.com, a customer encounters a survey regarding free online magazine subscriptions and is asked to indicate interest in receiving promotional information from various categories.

A central issue is that consumers are unwilling to provide much information unless they can see a clear benefit (Schwartz 1997, p. 72). Hence, firms offer incentives such as free products or participation in lotteries in return for more information. Some firms employ creative methods (e.g., interactive menus and games), while others collect data over multiple interactions. Research is needed to develop better data collection mechanisms and understand the effectiveness of alternate schemes for gathering information in different situations. Even with incentives, customers may not provide as much detailed information as desired, so firms use sample surveys, conducted either online or offline, to infer segment-level preferences.

Online surveys provide advantages of low turn-around time, lower costs, and the ability to reach “harder-to-reach” respondents such as busy executives and salespeople. Smith and Leigh (1997) and Miller And Dickson (2001) provide good reviews of online market research techniques and associated research issues. Important issues are whether online samples are representative of the target population, the effectiveness of alternate methods of administration, and the falsification of information by consumers. Table 2 summarizes the research issues related to data collection.

4.1.2. Tracking the Customer. The important types of data available from tracking a customer’s online interactions are:

1. Transaction data/point of sale data: This includes information on purchased items, their prices, time of purchase, and other conditions at the time of transaction. A visitor is identified as an existing customer through user registration or with the help of cookies.

### Table 2: Issues in Data Collection

<table>
<thead>
<tr>
<th>Issues</th>
<th>Research questions</th>
<th>Representative articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incentive mechanisms</td>
<td>How effective are monetary and nonmonetary incentives?</td>
<td>Geng et al. 2001</td>
</tr>
<tr>
<td></td>
<td>How does obtaining permissions improve the response rate and data quality?</td>
<td></td>
</tr>
<tr>
<td>Potential biases</td>
<td>In which situations is bias aggravated? How can online information gathering be improved?</td>
<td>Saris 1991, Kalfs 1993, GVU 1998</td>
</tr>
<tr>
<td>Adaptive surveys</td>
<td>What is the effect of adaptive surveys on data quality and response rates?</td>
<td></td>
</tr>
<tr>
<td>Privacy and trust</td>
<td>How can a firm credibly signal its good intentions in online surveys? How do privacy and trust affect data quality?</td>
<td>Miller and Dickson 2001</td>
</tr>
<tr>
<td>Sampling issues</td>
<td>How can representative samples be drawn? When that is not possible, how should adjustments be made?</td>
<td>Miller 1999, Miller and Gupta 2001</td>
</tr>
<tr>
<td>Reliability and validity of measures</td>
<td>Are results obtained from online surveys consistent with those obtained by traditional methods? If not, what are the causes for these differences? What are the benefits and limitations of alternate forms of delivery of surveys (e.g., e-mail/Web-based)?</td>
<td>Wilke 2000, Schafer and Wydra 2000</td>
</tr>
<tr>
<td>Role of technology</td>
<td>How do differential download times affect response rates? How does this distort the manner in which people view the same survey?</td>
<td>Miller and Dickson 2001</td>
</tr>
<tr>
<td>Tracking data</td>
<td>What are the advantages and limitations of site centric data and user-centric data? What models are appropriate for sparse data? How should one merge individual- and aggregate-level data?</td>
<td>Resnick and Varian 1997, Padmanabhan et al. 2001</td>
</tr>
</tbody>
</table>
(2) Web and application server logs: Web log files capture data such as browser host IP address, date and time of the interaction, the URL for the page requested, the referrer field, and a cookie field. Application server log files can include additional information such as the data queried from back-end databases, the thread-ID for the applet that was executed, and special events that may have occurred during the interaction.

(3) Cookies: Cookies are small text files placed on the hard disk of a browser host machine that helps identify a user both within a session and across sessions. They help understand browsing behavior within a session and track repeat user visits.

Electronic data collection avoids some of the pitfalls of direct asking. It is fast, accurate, and unobtrusive to the customer. However, there are a number of issues that arise when analyzing the data. Server logs collect data at the firm’s end, while panel data (as collected by Nielsen NetRatings or Media Metrix) capture a panel member’s Web log from the user’s computer. Server logs capture detailed information about the customer interaction with the firm’s website (i.e., provide site-centric data) but do not capture before-and-after activity of the customer on other websites. They also do not record usage of cached data. Panel data capture all Web activity at the user’s machine (user-centric data) but typically does not capture it in detail. Both data sources are useful in different domains and sometimes need to be jointly used in inference. Site-centric data present a subset of activities from user-centric data and cannot be used to understand competitive effects, or adequately explain site choice or purchases. However, the data are useful to model within site navigation, and optimize content and design. Research is needed to understand the bias in using one type of data over the other for various applications. Padmanabhan et al. (2001) show the superiority of user-centric data over site-centric data for predicting purchase in a given session or at a future session. Future research needs to provide methods to match insights from site-centric data with those from user-centric data during personalization. Russell and Kamakura (1994) present an approach to combine individual-level scanner data and store-level data to refine estimates. It is important to develop models for combining data from heterogeneous sources.

Resnick and Varian (1997) point out that Internet data is typically of high dimensionality (i.e., a large number of attributes are needed to describe the product space), but data on a given individual are sparse. This necessitates pooling of data across different customers, different sources, and even different categories, which presents many research opportunities. Click-stream data also poses interpretation problems when many people share a single computer. To whom should the firm personalize? How can individual users be identified? We discuss these issues in the next section.

4.2. Inference About Customer Preferences.
Firms employ a number of methods to infer customer preferences, purchase behavior, browsing behavior, and to deliver personalization. Broadly, the inference tasks can be grouped as prediction (purchases, Web visits, and so on), clustering and classification, and understanding preferences. Table 3 provides an overview of the three categories of models with commonly used techniques and representative applications.

4.2.1. Prediction Models. Traditional methods for prediction and modeling preferences include regression analysis, discrete choice models, neural networks, Bayesian networks, and other AI techniques. Neural networks and other AI techniques offer good predictions but are not as useful for understanding customer behavior. The choice of the technique depends on scalability and trade-offs between speed and sophistication. There are research opportunities for comparing the effectiveness of alternate techniques on multiple dimensions and in developing hybrid models.

Discrete choice models are useful in predicting purchases, website choice, Web page selection within a website, and in understanding customers’ responses to stimuli on a Web page. In these models, the dominant paradigm is customer utility maximization. Utility functions are defined and estimated as a function of customer and product characteristics. Generally, a
Table 3  Model Categories and Representative Applications

<table>
<thead>
<tr>
<th>Model categories</th>
<th>Commonly used techniques</th>
<th>Representative applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction models</td>
<td>Regression analysis, Logit and Probit models, Neural networks, Bayesian networks, Hazard rate models</td>
<td>How can firms predict customer response to targeted offers? What is the effect of prices and promotions on the probability of purchase? How likely is a customer to click on a given advertisement? How can firms predict duration of stay on a Web page?</td>
</tr>
<tr>
<td>Clustering and classification models</td>
<td>Automatic Interaction Detection (AID) / Chi-Squared AID (CHAID), Clustering, Latent class segmentation, Classification Algorithms and Regression Trees (CART)</td>
<td>To which segment does a customer belong? What characteristics define a segment? How can firms classify customers?</td>
</tr>
<tr>
<td>Preference models</td>
<td>Expectancy value model, Conjoint analysis, Ideal point model, Collaborative filtering, Hierarchical Bayes (HB) conjoint</td>
<td>How much does a customer value different attributes? What is the utility of alternate combinations of attributes to a customer? What product should be recommended to a customer?</td>
</tr>
</tbody>
</table>

Linear specification is used and utility for an alternative \( j \) is specified as

\[
U(j) = \beta X + \gamma Z + \varepsilon_j,
\]

where \( X \) is the vector of product attributes, \( Z \) is vector of consumer characteristics, and \( \beta \) and \( \gamma \) are parameter vectors. The choice of the distribution of the error term \( \varepsilon_j \), leads to the multinomial logit and probit models. These models have been extensively used in scanner data analysis. The models are estimated at an aggregate level using maximum likelihood methods (see Guadagni and Little 1983 for logit) or simulation methods (e.g., method of simulated moments, McFadden 1989 for probit). The logit model is easy to estimate, while the probit model, with an additional computational cost, allows for a flexible correlation structure to model the dependencies between alternatives and between the explanatory variables.

Recent work with click-stream data employs discrete choice models. Bucklin and Sismeiro (2001) model the decision to stay or exit a website using a binary probit model and find that the propensity to stay is a function of the depth of the site visit and number of repeat visits. In another paper, Sismeiro and Bucklin (2003) model the predictive power of user tasks (e.g., filling a cart, configuring a car, and so on) on the probability of a purchase. Montgomery (2001) proposes a dynamic ordered probit model to study the effect of covariates on the probability of exit, continuing to browse, or making a purchase. Website choice has also been modeled as a logit function of site loyalty, use of links, e-mail, and prior visits (Goldfarb 2001).

The challenges for researchers arise due to the nature of click-stream data. First, the attribute space is large and data on individual customers are sparse. Therefore, models are estimated using pooled data across customers. This leads to issues of aggregation bias and the need for estimation of individual-level parameters. Second, the Internet permits collection of data of multiple types that can be used for learning customer preferences. Proper integration of these data poses modeling challenges. Third, on the Internet, the customer actively creates the choice context and the firm can intervene at different stages in the purchase process (Bucklin et al. 2002). This could lead to endogeneity of marketing mix variables. Fourth, when a customer visits a website, her intention (whether to browse, search, or purchase) is not clear, which makes the prediction of timing of an event difficult. These issues are discussed below.

Aggregation Bias. Discrete choice models can be estimated for each individual if a sufficient number of transactions have been made in a given category. In most cases, data are not adequate for individual-level estimation. Therefore, researchers pool data across customers, which causes problems of aggregation bias. When pooling across customers, the estimates of the choice model will be biased if differences between individuals (such as differences in their preferences or in their responses to marketing variables)
are ignored (Guadagni and Little 1983). To mitigate the heterogeneity issue, initial attempts modeled the parameters as a function of customer characteristics. However, demographic variables explain a small percentage of the variance in response parameters (Rossi et al. 1996). Therefore, models were developed to control for unobserved heterogeneity. These include mixture models (Kamakura and Russell 1989), random intercept models (Chintagunta et al. 1991), random coefficient models (Gonul and Srinivasan 1993), or Bayesian hierarchical methods in multinomial probit models (Rossi and Allenby 1993). These methods are computation intensive and research is needed to compare the alternative methods of controlling for heterogeneity in the personalization context, in which both the attribute space and the choice space become larger.

Estimation of Individual-Level Parameters. Firms would ideally like to estimate individual-level parameters from aggregate data. Research employing Markov Chain Monte Carlo (MCMC) Bayesian hierarchical methods (Allenby and Lenk 1994) has demonstrated that it is feasible to make inferences about individual household parameters even for high-dimensional problems. Such parameters could be the basis for cross-selling, customizing promotions, or sending targeted communications. In the same manner, one can obtain website-specific parameters, or even page-specific parameters.

Integration of Multiple Types of Data. When combining individual- and aggregate-level data, proper integration is necessary. For example, how can segment-level preferences for some attributes be combined with individual preferences on other attributes to create a complete preference function? Recent studies demonstrate the use of Bayesian hierarchical methods in addressing this issue. Ansari et al. (2000) develop a movie recommendation system using a Bayesian preference regression model that permits even page-specific parameters. They incorporate preference heterogeneity across customers and unobserved product heterogeneity. In their model, a movie rating $r_{ij}$ for a customer $i$ and movie $j$ is a linear function of $w_j$, a vector of movie attributes, $z_i$, a vector of customer characteristics, and $x_{ij}$, a vector of movie and customer variables. The parameters are random (to capture unobserved heterogeneity) and are assumed to be distributed normal:

$$r_{ij} = x_{ij}^\prime \mu + z_i^\prime \gamma_i + w_j^\prime \lambda_j + e_{ij},$$

$$e_{ij} \sim N(0, \sigma^2), \quad \lambda_j \sim N(0, \Lambda), \quad \gamma_i \sim N(0, \Gamma).$$

$\mu$, $\gamma_i$, and $\lambda_j$ are parameters to be estimated. $\Lambda$ and $\Gamma$ are variance covariance matrices. The model is estimated using MCMC methods (Gelfand and Smith 1990) and is shown to perform better than collaborative filtering methods. Ansari and Mela (2000) develop a similar model in the context of targeted e-mails.

Endogeneity of Marketing Mix Variables. Another important issue in using the above models is the appropriate modeling of endogeneity of marketing mix variables. In traditional response models, the explanatory variables are assumed to be exogenous. However, in the context of personalization, prices and promotions are tailored to an individual customer and, hence, it is important to consider these variables as endogenous (Leeflang and Wittink 2000). This necessitates the specification of models for the endogenous variables and estimation of a system of equations. For instance, researchers need to model the process by which the price charged to a customer is determined by the customer’s actions, and simultaneously estimate the customer’s choice model and price determination model (Berry 1994, Chintagunta et al. 2001). Another alternative is to use instrumental variables for the endogenous variables as in Nevo (1997), who uses prices and advertising of ready-to-eat cereals in a similar geographical region as instruments to estimate the demand function in a given region. One may use lagged values of included variables as instruments as in Villas-Boas and Winer (1999), who develop a model to account for endogeneity of marketing mix variables in a multinomial logit model for analysis of household-level scanner data. In the personalization context, the endogeneity problem is worsened because the product, the context of the offer, and the offer are all determined as a function of the customer demographics and past actions.
Timing-Related Issues. It is useful to have the ability to predict the time when a customer is likely to visit again, make a purchase, or exit a website in deciding when to offer a promotion or an incentive. Models can also help to determine the number of visits or purchases in a given time period. Popular timing models in marketing employ hazard functions (i.e., the probability of purchase during a certain time interval \( t + \Delta t \), given that the customer has not purchased until time \( t \)). In Jain and Vilcassim (1991), the hazard function is specified as

\[ h(t \mid X, \theta) = h_s(t)\varphi(X)\phi(\theta), \]

where \( h_s(t) \) is the baseline hazard function, \( \varphi(X) \) is a function of explanatory variables, and \( \phi(\theta) \) is the specification of the unobserved heterogeneity. The authors use the Box-Cox formulation of Flinn and Heckman (1982) to specify the baseline hazard function and estimate the model using maximum likelihood estimation. Hazard models can accommodate censoring of data (Heckman and Singer 1984, Cox 1972, Lancaster 1979). Using click-stream data, John-son et al. (2000) model the duration of visits at a website as a function of the number of repeat visits, and find that repeat visits are shorter. Bucklin and Sis-meiro (2001) use a proportional hazards model to predict the duration of visits on websites. Timing models can be used to study when a customer is likely to switch to a competing website and factors that influence it. These models also help firms in optimal allocation of resources to customers over time.

4.2.2. Clustering/Classification Models. As stated earlier, personalization is often conducted at a segment level, based on the group’s preference function. Therefore, it is important to identify the number of distinct segments and their needs. Once segment-level behavior is understood, new customers need to be classified as belonging to a segment based on their profile to receive their personalized services. Over the years, a large number of techniques have been developed to group customers. Some commonly used techniques are Automatic Interaction Detection (AID)/Chi-Squared AID (CHAID), cluster analysis, Classification Algorithms and Regression Trees (CART) (Breiman et al. 1984), and latent class segmentation. Wedel and Kamakura (1998) provide a comprehensive evaluation of segmentation techniques.

Bhatnagar and Papatla (2001) present a logit model to segment customers based on their search behavior to deliver personalized advertisements. Based on patterns in Web-browsing behaviors, Moe (2001) clusters website visitors into buyers, browsers, searchers, and knowledge developers. In a related effort, Montgomery (2000) presents a Bayesian model using user-centric data to infer the characteristics of a Web surfer. If firms know the demographic characteristics of different websites, it is possible to compute the conditional probability of a visitor having a specified characteristic (e.g., Prob (visitor is male \( \mid \text{websites visited} \)).

Research issues in this area are: How often should segmentation be performed? What are the trigger events that suggest the need for a fresh segmentation? Can customers be classified into different segments over time? Customers could evolve from being a prospect, to a customer, to a supporter, and finally to an advocate (Brown 1999). While a few dynamic segmentation models have been developed (Wedel and Kamakura 1998), little is known about the stability of these methods. Sometimes different contexts suggest different segment memberships for the same individual. For example, a customer using Yahoo’s services could be a novice in high-technology products, while being an expert in stock trading. Future research should help us understand the customer characteristics, behaviors, and contextual variables that are most useful for personalization.

4.2.3. Preference Models. The Web provides an excellent medium for conducting experiments. Based on careful experimental designs, different samples of customers can be provided different products or offers. The responses of the customers can be analyzed to learn about the utility and the importance weight associated with each attribute of the product. In conjoint analysis (Green et al. 1972, Srinivasan and Shocker 1973), an established technique in marketing, the utility for a product is decomposed into the utility of its discrete attributes and is specified as

\[ U(X) = \sum_{i=1}^{m} \sum_{j=1}^{r_i} u_{ij} * x_{ij}, \]
where $U(X)$ is the overall utility of an alternative, $u_{ij}$ is the utility associated with the $j$th level of the $i$th attribute, $x_{ij}$ is one if the $i$th attribute is present and zero otherwise, $m$ is the number of attributes, and $r_i$ is the number of levels of attribute $i$. Customers may reveal their preferences by ranking multiple products or by choosing among a selected set of products. The estimated parameters can be used to simulate market shares of new products, find different segments of customers, and design optimal products.

A common problem in applying conjoint analysis is respondent fatigue. For example, the number of all possible combinations of three levels each for five attributes is $3^5$ or 243. Solutions to reduce fatigue are adaptive conjoint analysis (Johnson 1987, 1991; Green and Srinivasan 1990) and hybrid conjoint models (Green and Krieger 1996). Choice-based conjoint analysis (CBA) uses customer choice as the dependent variable to provide estimates at an aggregate level (Louviere and Woodworth 1983, Mahajan et al. 1982, Batsell and Louviere 1991). Analogous to discrete choice models, individual-level estimates have been obtained by employing the HB methods (Lenk et al. 1996). In online transactions, customers click on items and each click represents a choice among a set of alternatives. By carefully varying the choices, consumer preferences can be estimated using CBA. The choice of appropriate experimental design for different types of personalization needs to be studied. HB models will permit integration of information across multiple pages and websites to provide individual-specific estimates.

Other Issues. (1) Joint preference functions: In group decision making (e.g., family vacations), how can a firm develop joint preference functions? What models are relevant for personalization to groups? (2) Identification of individual user: When a family shares the same computer, it is important for a firm to identify the family member who is online at any given moment, which is not easy at this time. (3) Preferences for bundles: In many situations (e.g., groceries, music, movies, or news items on a website), a customer is interested in a group of items. McAlister (1979) applies linear programming techniques to find optimal bundles based on preference judgments and perceived similarities between products. (4) Consideration sets: How many items should be offered for a customer’s consideration and which items should be offered? (5) Scalability of models: The models discussed in this section are practical for data sets that have about a few thousand observations. However, when the data set is in millions of records, some of these models may not scale up. Hybrid methods that incorporate AI techniques and econometric approaches could be developed. (6) Changing preferences: Firms should understand the difference in preferences due to purchase objective (e.g., whether a purchase is for consumption or to give as a gift). Similarly, it is important to understand the difference between transient and permanent preferences. Preference may change over time for some attributes, while remain unchanged for others. Table 4 summarizes these issues.

4.3. A Word About Matching
A number of methods such as collaborative filtering and rule-based systems are employed in making personalized recommendations to customers. The underlying model recommends the product that maximizes a consumer’s utility function $u(x, z)$ defined over product attributes $x$ and consumer characteristics $z$. Utility is measured by a rating provided by the customer for each product. In collaborative filtering, a score $r_{xz}$ is computed that is a function of the ratings of a prespecified number of most similar users for the same product. The function can be a simple average or a weighted sum using a measure of similarity between customers as a weight. Similarity between customers could be modeled using a cosine function, or a correlation coefficient over the set of items rated by the customers. These functions are optimized using neural networks, clustering techniques, and Bayesian networks (Adomavicius and Tuzhilin 2002).

Research is needed to meld the theoretical rigor of management science with the speed of AI search algorithms. In discrete choice models, matching can be viewed as a prediction problem in which the model
predicts choice of an item based on consumer and product characteristics. The estimation of such models is time consuming, but predictions can be made in real time. Researchers can develop new models that consider maximization of a firm’s objectives (profits, revenues, or customer satisfaction) or a customer’s objectives (search, browse, or purchase). Personalization alternatives can also be chosen to maximize the probability of a purchase, the probability of a return visit, the number of items in a basket, or customer lifetime value. Alternately, constrained optimization models could be developed in which the firm’s objective function could be maximized subject to achieving a certain level of a customer’s objective function. Geoffrion and Krishnan (2001) discuss applications and prospects for operations research in the context of personalization. While a detailed discussion of this topic is outside the scope of our study, we provide a few examples below.

Advertising networks schedule banner advertisements for their clients, keeping in mind site requirements and customer preferences. Optimal banner placement strategies present a rich opportunity for the employment of traditional operations research models. Adler et al. (2002) and Kumar et al. (2000) have employed scheduling techniques to maximize the advertising revenue for a site subject to capacity constraints across several Web pages. Recommendation systems help reduce the choice set of a customer from a large number of alternatives to a select set of relevant ones. Such “k out of n” choice scenarios could be modeled as the knapsack problem and solved using integer programming methods. Matching of products to customers and customers to products (e.g., disposal of stock-out items) could be modeled using the generalized assignment problem. When allocating scarce resources (e.g., allocating customers to service representatives in call centers or controlling congestion at back-end servers), firms can use priority queuing approaches based on customer potential (Tan et al. 2003).

5. Conclusion

In this paper, we focus on research opportunities in the management sciences in the context of personalization. We have approached research issues at two different levels. First, we look at the strategic role of personalization for a firm. We examine the role of personalization in the interactions between a firm and other key players in the firm’s value system, survey current research, and suggest avenues for future
research. Next, we focus on one of the key activities a firm must undertake to effectively provide personalization, namely learning customer preferences. We discuss existing approaches to customer modeling and suggest how these and newer models could be used for personalization applications.

Our focus has been primarily on online environments. However, future developments in wearable computers could reduce the distinction in online and face-to-face interactions. There are three important differences in interactions across these two environments. They are the ability for nearly instantaneous customer identification (e.g., through IP address or a cookie), greater ability to capture more information about the customer (e.g., through a Web log), and greater ability to recall more information about a customer once identified (through real-time database access). An intriguing future possibility is the use of wireless heads-up displays in ordinary-looking eye-glasses that a sales person could use in a traditional brick-and-mortar store. This could enable the salesperson to overcome the three main differences between online and face-to-face interactions. If customer acceptable, a radio-frequency identification (RFID) tag embedded in the discount card that many stores issue could enable registered customer identification as soon as they pass through the door, and noninvasive biometric technology might identify some of the rest who do not have a tag. Wireless connectivity to profiling databases could help the salesperson in making recommendations based on prior interactions with the customer, and voice recognition systems could capture the new interaction for future use. We should mention here that several issues in human-computer interaction would need to be resolved before these devices could be successfully deployed.

While we have attempted to provide a reasonably comprehensive survey of the involved issues and current research, we make no claims that the survey is exhaustive. Our hope is to increase awareness of the importance of personalization in a firm’s strategic and operational considerations, and to illustrate some of the important problems and opportunities for researchers in that context. Given the interdisciplinary nature of these issues, researchers must be able to view problems from the different perspectives and be able to bring to bear tools and techniques from different disciplines to make significant contributions. The difficulty in doing this well is compounded by the pace of technology changes, which continually create new ways in which firms can personalize products and services.

Acknowledgments
The authors are indebted to the guest editors, the associate editor, and the reviewers for providing valuable suggestions that have substantially improved the organization of this paper. They particularly appreciate their quick turnaround time for each round of review.

References


Cheng, H. S., K. Dogan. 2001. One-to-one marketing with Internet coupons: Spatial models of price discrimination. Working paper, University of Texas at Dallas, Richardson, TX.


——, A. Gupta. 2001. Studies of information, research, and consulting activities (SIRCS): Fall 2000 survey of organization. A. C. Nielsen Center for Marketing Research, University of Wisconsin-Madison, Madison, WI.

Moe, W. 2001. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. Working paper, University of Texas, Austin, TX.


MURTHI AND SARKAR
The Role of the Management Sciences in Research

varian, h. r. 2001. economics of information technology. working paper, university of california, berkeley, ca.
villas-boas, j. m. 2001. competing with experience goods. working paper, university of california, berkeley, ca.

wedel, m., w. a. kamakura. 1998. market segmentation: conceptual and methodological foundations. kluwer academic publishers, boston, ma.
wilke, j. 2000. a landmark study of the differences between mall-intercept and internet interviewing from bases. eleventh annual advanced research techniques forum, monterey, ca.

1362
management science/vol. 49, no. 10, october 2003