The E-Butler Service, or Has the Age of Electronic Personal Decision Making Assistants Arrived?

Alex Tuzhilin
Leonard N. Stern School of Business
New York University

May 1998

Working Paper Series
Stern #IS-98-16

The E-Butler Service, or Has the Age of Electronic Personal Decision Making Assistants Arrived?

Alexander Tuzhilin

Information Systems Department
Stern School of Business
New York University

Abstract

This paper describes an *Electronic Butler* (or *e-Butler*) that provides a *customer-centric personalized* shopping services to its subscribers across a *wide* range of products. This service is provided by identifying individual customer's shopping needs from the comprehensive purchasing history of that person and providing purchasing recommendations or direct purchasing decisions for the customer. e-Butler service consists of two components -- the *Personal Shopping Assistant (PSA)* service that provides purchasing *recommendations* to the customer and the *Magic Wand (MW)* service that *directly* makes purchases it believes the customer needs without any prior consultations with the customer. In order to understand how PSA and MW services of e-Butler are related to the existing one-to-one marketing and recommender systems, a general framework classifying various personalized shopping services is presented that clearly delineates PSA and MW services from these existing systems. Moreover, the paper presents an architecture of the e-Butler service, explains what its business value is, discusses its feasibility, and describes what needs to be done to make it a successful service.

Acknowledgments. The author gratefully acknowledges several important suggestions made by John Henderson, N. Venkatraman, and P. Balasubramanian from Boston University about some concepts described in the paper and their presentation.

1. Introduction

We constantly make various purchasing decisions in our daily lives ranging from the decisions on which groceries, to which clothes, books, and appliances, to which cars and houses to buy and where. Since our daily lives are becoming increasingly busy and we are becoming increasingly pressed for time, these purchasing decisions look more and more as nuisances to some of us. Wouldn't it be wonderful to delegate at least some of these decisions to somebody else?

Well, not all of these decisions can be delegated to others. For example, buying a house, a car, or some other complex and expensive "big ticket" item cannot be delegated easily. In more general terms, it is hard to delegate *unstructured* and *semi-structured* purchasing decisions to others. Therefore, we will focus in this paper only on the *structured* purchasing decisions[1], such as buying groceries, books, CDs, etc.

The desire to delegate buying decisions to somebody else existed for many centuries. For example, in one of the Russian folk tales, the main hero learns certain magic words[2] and, when he utters these words, suddenly, delicious foods start arriving out of thin air and jumping straight into his mouth, or beautiful clothes would replace his old shabby clothes[3]. Thus, there is no need for the hero to figure out what foods to buy and where, what to cook for dinner, which clothes to buy and where – just say the magic words, and your consumer needs are instantly fulfilled.

This story projects the ultimate consumer paradise or, using mathematical terminology, the *limit* that can never be attained, and that can only exist in fairy tales. However, this limit gives us some vision of what we can strive for. After all, as one of the songs of the by-gone communist era in the former Soviet Union goes, "Mi rozhdeni chtob skazku sdelat' bilyu" (we are born to make a reality out of a fairy tale). The main question, of course, is how closely we can approach this limit. We will address this question in this paper.

Rich people in the past tried to approach this limit by using help of personal assistants, housekeepers, and butlers. A butler would know all the idiosyncratic tastes and desires of his master and would take the responsibility of making many day-to-day purchasing decisions with limited or no consultation with the master. Unfortunately, this type of service could be afforded only by the well-to-do people in the past.

[1] Although we are focusing on *purchasing* decisions in the paper, our ideas can be applied to other types of decisions that are based on transactional histories, such as medical decisions relying on patient's medical histories (e.g. which medical examinations to do and when, which treatments to prescribe, etc.), internet-based decisions (e.g., which sites to visit, which ads to show), etc.

[2] The Russian words are "Po schootchemy veleniyu, po moemu hoteniyu," which mean in English "According to pike's orders, according to my wishes..."

[3] Similar tales exist, of course, in many other nations.

In this paper we argue that Information Technology can bring this type of service to the masses through the creation of an *Electronic Butler* (or *e-Butler*) service. e-Butler provides a *personalized* shopping service to its subscribers by identifying individual customer's shopping needs and either giving him or her purchasing recommendations across a *wide range* of products or making purchases directly without any prior consultations with the customer.

The idea of an e-Butler is not entirely new. Some of its elements have already been implemented within the framework of concierge services [Peppers&Rogers98], one-to-one marketing [Peppers&Rogers93, Pine et al 95] and recommender systems [CACM97]. The concierge services, provided by such companies as World Class Concierge, Concierge@Large, Les Concierges, and Capitol Concierge, meet individual clients needs by managing an array of activities for them. Some examples of such activities include obtaining tickets to various events on a world-wide basis, locating hard-to-find items, arranging travel, making hotel and restaurant reservations, organizing itineraries for the clients, organizing meetings, and planning events. These activities are targeted to corporate as well as individual clients and save them time and trouble of organizing these activities on their own. Many companies, such as American Express, Visa, Mastercard, and General Motors, have adopted concierge services as a value-added amenity. However, concierge services are examples of *passive* services: the clients have to explicitly initiate a request for specific services they need. Moreover, concierge services are still very labor-intensive: they employ people who take care of customers requests.

In contrast to this, one-to-one and recommender services take a more proactive and more automated approach to helping their customers. They recommend them new products often without waiting for customers requests. Moreover, it is software rather than people who provides personalized recommendations. Some examples of recommender services are *First!* from Individual Inc., Instant Recommendations and BookMatcher from Amazon.com, and Catalog Navigator from Firefly. However, e-Butler is quite different from these services. Most of these one-to-one and recommender services take a *business-centric* approach to providing personal *recommendations* about a *narrow* set of products. The provided recommendations are business-centric because they usually pertain to the group of products offered by these companies, such as books by Amazon.com, customized news from Individual, etc. Moreover, as will be explained below, most of these recommendations require extensive inputs and/or feedback from the customers. Therefore, they are applicable to a narrow set of products. In contrast to this, the e-Butler service provides a *customer-centric personalized* shopping alternative across a *wide* range of products. Moreover, it not only provides purchasing recommendations, but also makes actual purchasing decisions in some cases. These differences between today's one-to-one and recommender services and e-Butler will become more apparent when we present a framework for the personalized shopping services in Section 3 and explain how existing systems and e-Butler fit into different places of this framework.

In more general terms, there is a *gap* between the existing one-to-one and recommender services on one hand and the "limit" described before in the Russian fairy tale on the other hand. In this paper, we will explore this gap and examine how the e-Butler service fills certain parts of this gap. In particular, we will explain what e-Butler service is, how it works, and will present its architecture. Moreover, we will explain what value it provides to the customers subscribing to the service. We will also examine how feasible is the e-Butler service and discuss what needs to be done to make it successful.

2. State of the Art

The idea of an e-butler system is not entirely new. Elements of it have already been implemented within the framework of one-to-one marketing [Peppers&Rogers93, Pine et al 95] and recommender systems [CACM97] that collectively fall into the category of *personalized shopping services*. In particular, several companies, such as American Express, MCI, Individual, Amazon.com, and Peapod LP,[4] have developed such services that we describe below. These services are made possible because of the recent development of *personalization technologies*, introduced by such companies as Firefly, Net Perceptions, Open Sesame, LikeMinds, and others.

American Express launched in 1997 the CustomExtras program that offers merchants, who accept their card, the ability to print their own promotional personalized messages on the monthly statements of the selected AmEx cardholders that the merchant has identified as their target market (mostly based on the past purchasing history with that merchant). To support the CustomExtras program, American Express has deployed a very large database that tracks information about customer's purchases, as well as numerous rewards and promotions offered by different merchants.

MCI has developed a decision support system (DSS) that constructs calling history profiles of individual customers. Customers' profiles are then scored in order to predict their propensity to use additional MCI services, such as international calling. The resulting recommendations of this DSS system are subsequently transferred to the MCI telemarketing system and are used for the cross-selling purposes. MCI credits this system with helping to increase average customer billings.

Peapod, Inc. is one of the leading interactive, on-line grocery shopping and delivery services based in Evanston, Illinois. It allows its customers to access Peapod's on-line groceries shopping catalogue and

[4] This is, certainly, not a complete list of companies embracing the concepts of one-to-one marketing and trying to recommend some of their products and services to customers on an individual basis. However, we believe that this is a fairly representative list of companies making impact on this area.

select products the customer wants to buy. Then the purchased products are delivered to the customer's door by Peapod delivery services. The company currently provides grocery services in eight metropolitan markets in the U.S. and serves over 60,000 households. As a part of its operations, Peapod provides a virtual supermarket for an individual customer that best suits him or her. Customers may request the lists of available products by category, by item, by brand, or what is on sale today. They can also create and save their preferred shopping lists or groups of items they typically buy together (e.g. milk products). Peapod also tries to learn about customers' shopping experiences. At the end of the session it asks "How did we do on the last order?" and uses this customer feedback to improve one-to-one relationships with individual customers. In addition, Peapod views delivery as another opportunity to interact with its customers: the delivery person asks questions about customer satisfaction with Peapod's service and for suggestions to improve it. Customers' feedback has prompted the company to make several changes to its services, including providing nutritional information, making faster deliveries (at additional price), and delivering alcoholic beverages.

Another category of companies, such as Amazon.com, BarnesandNoble.com, Yahoo, and Individual Inc., provide personalized suggestions to customers regarding which books to buy (Amazon.com and BarnesandNoble.com), which Web pages to see (Yahoo), and which news to read (Individual). For example, Amazon.com, one of the leading on-line book sellers, offers two services to its customers -- Instant Recommendations and BookMatcher. The Instant Recommendations service suggests Amazon.com customers the books that it thinks they would most likely enjoy based on the book purchasing histories of individual customers. This is a truly one-to-one service because it makes purchasing recommendations based on individual customers' profiles. An alternative BookMatcher service first asks the customer to rate a selection of books, then identifies the group of readers with similar tastes, and recommends some additional books to the customer that these readers liked. The BookMatcher is essentially a segment marketing service (and not a true one-to-one marketing service) because it tries to fit a customer into a certain segment and use purchasing patterns of similar customers as its recommendations.

As another example, the *First!* service from Individual, Inc. delivers targeted news from multiple news sources to its customers on the topics selected by them based on their interests. In particular, Individual's SMART software searches through over 400 sources containing more than 12,000 articles [Pine et al 95] for those pieces that will most likely fit the client's needs. It delivers them by whatever method the client has chosen, such as FAX or e-mail. Every week, Individual asks a new client to rate each article as "not relevant," "somewhat relevant," and "very relevant." Client's response provides a feedback loop for Individual and allows the company to learn clients' needs and preferences and therefore customize the news delivery to individual clients' needs. In the first week of service, most customers find only 40% to 60% of the articles to be somewhat or very relevant [Pine et al 95]. However, this number increases to

80% -- 90% by the fourth or fifth week [Pine et al 95].

These personalized services provided by Amazon.com, Yahoo, and Individual, rely on the personalized technologies known as *recommender* [Resnick&Varian97] or, more generally, *recommendation systems* [Stohr&Viswanathan99] that are developed by such companies as Firefly, Net Perceptions, LikeMinds, and Open Sesame. A *recommendation* system [Stohr&Viswanathan99] provides recommendations to its users about a group of products or services that they should consider purchasing or using based on the evaluation of different choices available to them. These evaluations are based on a broad range of information available from different sources, such as opinions of other people, user feedback about the recommendations, customers' profiles, etc. In contrast to this, a *recommender* system [Resnick&Varian97] is a special type of a recommendation system in which recommendations are based on the opinions expressed by other people. Recommendation systems are classified according to the source of the knowledge, or expertise, on which the system bases recommendations into utility estimation, content-based, collaborative, and expert-based systems [Stohr&Viswanathan99]. Since there are no known examples of utility estimation systems [Stohr&Viswanathan99], and expert-based systems are only distantly related to e-Butler, we will consider only *content-based* and *collaborative* recommendation systems.

Collaborative systems, such as Phoaks (http://www.phoaks.com/phoaks), GroupLens from Net Perceptions, Preference Server from LikeMinds, or Passport-based systems from Firefly specialize in the development of personalized recommendation systems for the on-line users using the *collaborative filtering* technology [Goldberg et al 92]. This technology compares tastes of an individual customer with the tastes of many like-minded customers to predict the customer's interests and make appropriate recommendations based on the interests and actions of similar customers. To illustrate how the collaborative filtering works, consider the Web site http://www.mylaunch.com that sells CDs and uses the collaborative filtering technology from Firefly. When you visit this site, you will be first asked to rate 10 CDs. Then the site will give you a list of five more CDs that it thinks that you might like and asks you to rate them. After you rate those, it will recommend you five more CDs. By examining your ratings and learning your tastes and preferences, the site generates with each new iteration more and more focused lists of recommendations that reflect your preferences and the preferences of other like-minded customers. The distinguishing feature of the collaborative filtering technology is that it does not truly provide a one-to-one approach. It is, essentially, a segmentation marketing technique because it places a customer into a group of similar customers and recommends to that customer the actions taken by similar customers in that group. The success of collaborative filtering systems depends on the availability of a critical mass of users with similar profiles and credibility of their evaluations.

Content-based recommendation systems, such as Instant Recommendations system from Amazon.com, *First*! from Individual, Syskill&Webert [Pazzani et al] and NewsWeeder [Lang95], base their recommendations on user profiles. These profiles are built using information about user preferences that is either elicited from them through questionnaires or learned from their transactional behavior over time. These profiles are constantly updated in time by obtaining *relevance feedback* from the users that specifies how much they are satisfied with recommendations. One example of a content-based recommendation system is the *First!* service from Individual, Inc. that was described above.

Content-based recommendation systems have the advantage of learning customer's needs in the one-to-one fashion better and better over time (as the Individual's case demonstrates). However, they also have the following drawbacks [Stohr&Viswanathan99]. First, reliance on the relevance feedback results in "over-specialization" [Balabanovic&Shoham97], i.e., the system works best in the restricted domains that the users have evaluated in the past. Secondly, it is often difficult to obtain proper feedback from the users about recommendations. This is especially true when the system recommends complex items that takes time to evaluate, such as some of the Web documents. In some situations it may even be impossible for the user to evaluate properly the quality of the choices presented to him or her.

Many recommender systems follow a hybrid approach in which they combine collaborative filtering with the content-based analysis. For example, Barnes and Noble Web site (BarnesandNoble.com) uses the technology developed by Firefly Networks, which combines the content-based and collaborative filtering methods. Also, Amazon.com offers two services to its customers -- Instant Recommendations and BookMatcher. Instant Recommendations uses the content-based analysis and BookMatcher the collaborative filtering approach.

Although different from each other in many respects, the personalized purchasing services described in this section have certain important features in common. First of all, the companies that offer one-to-one services, such as MCI and Peapod, and recommendation services, such as Amazon.com and Individual, take a *business-centric* approach. As a part of this service, they attempt to understand individual customers by analyzing their demographic, and transactional data pertaining to purchasing of *their* products and services. Once they understand this purchasing behavior, these companies attempt to issue recommendations of what additional products and services *produced by these companies* their customers should consider buying. This constitutes a business-centric approach to serving the customers: the personalized marketing service is just another method of pushing more and more products and services on the consumer by the companies offering this service. A typical example of this is the decision support system from MCI whose purpose was to increase customers' billing by offering them additional services that MCI felt would be relevant to the customers. An alternative *customer-centric* approach takes the standpoint of the customer and assists him or her in purchasing decisions. The e-Butler service follows this approach, and we will describe it below.

The second feature, common to several of the services described above, is that, in order to assure adequate levels of customers' satisfaction, these services require *extensive involvement of customers* in the recommendation process. For example, Individual, Barnes and Noble, Mylaunch, and Amazon.com ask customers to provide extensive feedback information on the products that these companies recommend. Although, as explained in [Stohr&Viswanathan99], this is not strictly necessary for recommendation systems, most of the existing content-based and collaborative systems require extensive user inputs. An exception to this norm is the Learn Sesame system from Open Sesame. Learn Sesame learns individual profiles of Web users in a non-intrusive manner by observing customers' habits and interests as they browse Web pages without burdening site visitors with many questions. This is achieved by building individual customer's profiles using neural network technologies.

The third feature common to the personalized services described above is that all of them offer customers a *narrow set of products*, such as books (Amazon.com), telephone services (MCI), delivery of personalized news (Individual), or grocery shopping (Peapod). Therefore, the systems supporting these services analyze only a small subset of the total set of all the customer's purchases and deploy the methods specifically suitable for that subset. Thus, they can be thought of as expert systems that were once popular in Artificial Intelligence in the 70's and the 80's. Therefore, we will call this class of services *Expert Recommendation Systems (ERS).* The tendency to focus on a narrow set of related products and services is understandable because most of the services require user inputs. It is hard enough for a customer to provide inputs on a small set of products and services, such as books and news. It would be practically infeasible to ask the customer to do this for a wide range of products and services. We would like to point out that Learn Sesame is, again, an exception here. Since it supports a non-intrusive learning of customers' profiles, this means that it can learn their buying patterns across a wide range of products and services.

In summary, the personalized purchasing services described in this section are very useful in practice and are a good starting point towards the ideals described in the introduction (such as the one from a Russian fairy tale). However, they provide only the first step towards that goal. We will describe an e-Butler service in this paper that should take us further towards that ideal. However, before describing the e-Butler service, we would like to present a broad framework for classifying various personalized shopping services and demonstrate how the services described in this section fit into this framework. Moreover, using this framework, we present other types of services broader than the ones considered in this section, including the e-Butler service, that can take us closer towards the ideals described in the introduction.

3. A Framework for Personalized Shopping Services

Personalized shopping services either *help* the user by giving purchasing recommendations for one or several categories of products and services or making some purchasing decisions *on their own*. Therefore, personalized shopping services can be classified along the following two dimensions:

- *Purchasing decisions*: does the shopping service only *recommend* purchases to its customers or does it *actually make* these purchases without any prior consultations with them? In the former case, the service acts as a *decision support* system by only recommending a purchase and leaving the actual purchasing decision to the customer. In the latter case the service *automates* purchasing decisions.

- *Class of products or services*: are purchasing decisions made about a narrow or a broad class of products or services? If purchasing decisions or recommendations are made about a narrow set of products or services, then such a service acts as an *expert system* specializing in providing services for a narrow domain of the system's expertise. Examples of such services are news selection, book and CD recommendations, groceries shopping services, and "book-of-the-month" clubs. If purchasing recommendations or decisions are made about a broad range of products, then such a service is *generic*. An example of such a service is Custom Extras from American Express because it allows different kinds of merchants to offer a very broad range of products and services to a customer.

These two dimensions give rise to the 2x2 matrix presented in Figure 1 that classifies personalized shopping services into the following four quadrants:

1. *Expert Recommendation Service (ERS)*. This service provides purchasing *recommendations* about a *narrow* range of products, such as books, news items, and groceries. Most of the developments in personal shopping services, including Internet-based shopping services provided by such companies as Amazon.com, Barnes and Noble, Mylaunch, Peapod, etc., focus on this quadrant. Examples of nonInternet-based personal services are the MCI's decision support system and Individual's *First!* service.

2. *Expert Decision-making Service (EDS)*. This service makes *automated* purchasing *decisions* about a *narrow* range of products. Examples of these services are the various "book-of-the-month" (or "record-of-the-month") clubs that automatically ship its customers a book (or a record, or a CD) each month, unless the customer notifies the club that he or she does not want the book scheduled to be sent. The problem with these types of services, as they are today, is that most of them are not personalized: most of these services do not know their customers well in a one-to-one fashion and, therefore, do not target specific books (or other products) to specific customers. Another type of EDS

are *automated replenishment systems.* They automatically reorder supplies whose stocks decrease below minimal levels. These systems deal with a single type of decision: when to reorder goods and in what quantity. Therefore, they belong to the "expert" category.

3. *General Recommendation Service (GRS).* This service provides purchasing *recommendations* about a *broad* range of products, ideally, about *all* the purchasing needs of a customer. Custom Extras service from American Express falls into this category because it allows different merchants print their own promotional personalized messages, and these messages can be about *any* product or service which can be acquired with the American Express card. Although Custom Extras falls into this category, it is still an example of a relatively simple service that barely passed the test for the GDS quadrant. This is the case because each promotional message is based on the products purchased by the customer from *that particular* merchant rather than it being based on a more comprehensive purchasing history of the customer (e.g. the set of all the AmEx transactions).

   A better example of the GRS is the *Personal Shopping Assistant (PSA)* service. The PSA service works with a comprehensive purchasing history of a customer (e.g. the purchasing history containing 80% - 90% of all the purchasing transactions performed by the customer over the past several years). It takes a set of initial inputs (e.g. I want to buy a case of wine and I am ready to spend up to $20 per bottle) and produces a set of recommendations from which the customer selects a product. For example, it can suggest ordering a case of customer's favorite Merlot of 1992 vintage from distributor XYZ (who currently runs a sale on this wine, thus making a purchase a good bargain). Moreover, the PSA service is not limited to a group of products (e.g. liquors) and issues purchasing recommendations about a broad range of products. The PSA service will be described in much greater detail in Section 4.

4. *General Decision-making Service (GDS).* Unlike EDS, that is limited to a specific category of products (such as books, wines, news services, etc.), the GDS service makes *automated* purchasing *decisions* about a *broad* range of products. Such a service has not been implemented yet, and its implementation constitutes a serious technical challenge. One specific type of a GDS is the *Magic Wand (MW)* service. As a PSA, Magic Wand service needs a comprehensive purchasing history of a customer. However, unlike the PSA, the MW service makes purchasing decisions directly, *without any prior consultations* with the customer. For example, unlike the PSA service that only informs the customer about the purchasing possibilities, the MW system can order a case of California's Merlot wine of the 1992 vintage, and it is shipped directly to the customer without any prior consultations with him or her. We call this type of service "Magic Wand" because goods arrive at a customer's door from "thin air" without any prior thought on his or her part (as if a fairy waived her magic wand). The MW service will be described in much greater detail in Section 4.

Evaluation of Quadrants in Figure 1. Which of the four types of services provide the most value to the customer? The ERS type of service requires the customer to make purchasing decisions (although it provides extensive decision support in a way described above). Moreover, it is limited to a narrow set of products that it supports and cannot give recommendations for other types of products lying outside of its area of expertise.

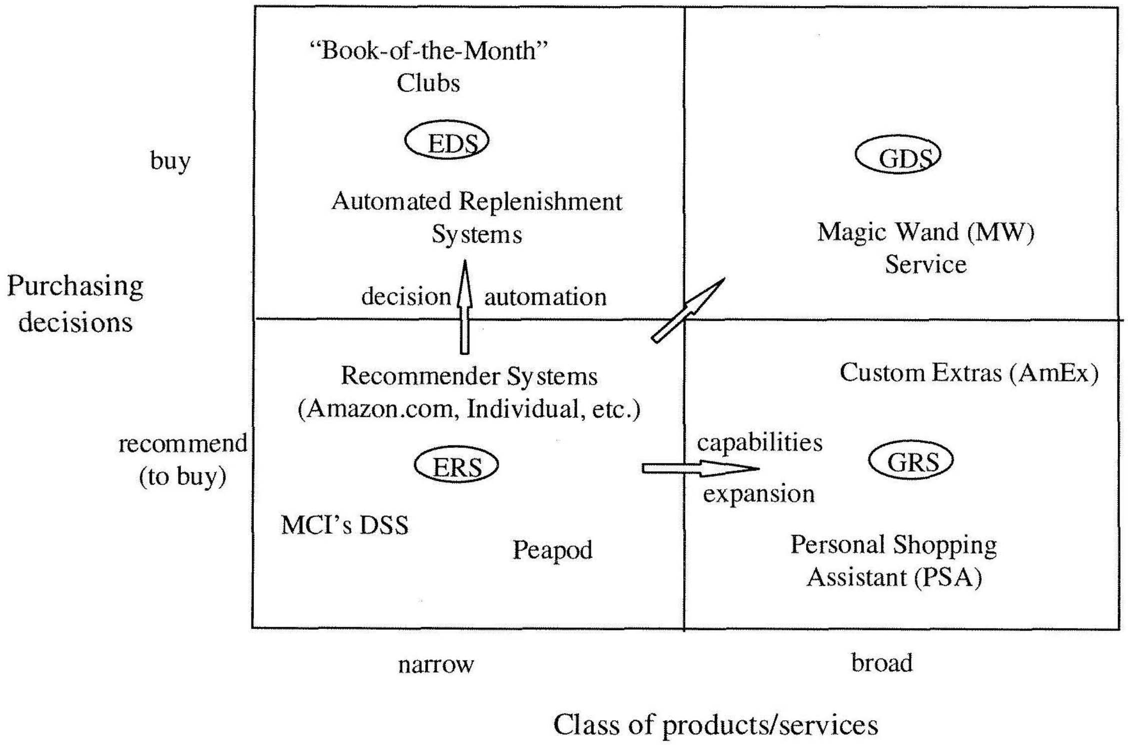

Figure 1. *Classification of Personalized Shopping Services.* Y-dimension specifies purchasing decisions: is the actual buying decision made or only buying recommendation is provided by a shopping service. X-dimension specifies the class of products and/or services about which decisions/recommendations are made: are purchasing decisions/recommendations made about a narrow or a broad group of products and/or services? Abbreviations: ERS -- Expert Recommendation Services, EDS -- Expert Decision-making Services, GRS -- General Recommendation Services, GDS -- General Decision-making Services.

The GRS type of service has a significant advantage over the ERS service in that it makes a much broader set of recommendations to the customer. By making a broader set of recommendations, the GRS service can also leverage its expertise and a more intimate understanding of the customer by providing more intelligent recommendations than the ERS service. For example, if the GRS service recommends books, CDs, concert tickets and movies to a customer, it can leverage its knowledge about purchases in other categories to provide more intelligent recommendations in a particular category. For example, such a service can provide a better recommendation of CDs if it knows not only which CDs the customer purchased in the past and how much she liked them, but also which books the customer read in the past, which movies she saw, and which concerts attended.

Similarly, the EDS service has an advantage over the ERS service in that it simplifies customer's life: the customer does not have to provide preferences and feedback inputs to the service so that it could make better recommendations. Also, the customer does not have to deal with purchasing decisions because the EDS service makes these decisions for him or her and only delivers the products. For these reasons, the EDS service has a significant advantage over the ERS service. However, the main challenge for the EDS service is to achieve accuracy rates acceptable to the customer. We will discuss this issue in Section 6.

Finally, the GDS service has advantages over the EDS and GRS services. It provides extra value to the customer by being more general than the EDS service and thus being able to make more purchasing decisions than the EDS service. It also leverages its more intimate knowledge of the customer in the same way the GRS service leverages it over the ERS service. With respect to the GRS service, the GDS service has the advantages that the customer does not have to be involved in the decision making process anymore and in that the customer does not have to provide feedback information to the service.

In summary, the GDS service provides the most value to the customer, whereas the ERS service the least. However, most of the existing personalized shopping services are of the ERS type. This situation is not surprising because of certain technological and behavioral constraints, to be explained in Section 6, that make it difficult to reach other three quadrants, especially the GDS quadrant. Nevertheless, some companies are capable of moving to other quadrants, mostly EDS and GRS, because they push the limits of the existing technologies and because of the recent technological advances (we will discuss some of these advances in Section 6). For this reason and because the GDS service provides much more value to the customer than the ERS service, we address the following important question:

How can one move from the ERS to the GDS type of service?

There are two ways to achieve this (see Figure 1):

1. ERS $\rightarrow$ EDS $\rightarrow$ GDS

2. ERS $\rightarrow$ GRS $\rightarrow$ GDS

An example of the first type of move in the book purchasing domain would be to start with an Amazon.com (or BarnesandNoble.com) type of recommendation service and move to the "book-of-the-month" type of service performed in a truly one-to-one fashion (with an extensive understanding of an individual customer and his or her reading interests and needs). After automatic book purchasing decisions are implemented successfully (the ERS $\rightarrow$ EDS move is completed), the service can be expanded to other categories besides books (e.g. CDs, movies, etc.). This corresponds to the EDS $\rightarrow$ GDS move. The main advantage of this route lies in that the ERS $\rightarrow$ EDS move is not that difficult to implement (elements of it have been implemented already by the "book-of-the-month" type of services). Similarly, the EDS $\rightarrow$ GDS move can be implemented gradually by including more and more categories of products and services at a time. The danger of this move, however, lies in that, in order to expand an EDS service to many other categories, one may need totally new processing techniques that may be quite different from the methods used in the EDS service.

The second type of move (ERS $\rightarrow$ GRS $\rightarrow$ GDS) has its own challenges. The challenge of the ERS $\rightarrow$ GRS transition lies in that most of the existing recommendation systems rely heavily on the customer feedback. Such feedback (either the initial ranking of the products or customer feedback regarding specific recommendations) can be obtained for a narrow group of products (such as books, CDs, or wines). It is much harder to obtain meaningful feedback across many product categories because it would require too much involvement from the customers, and few of them would agree to do this. However, new technologies, most notably from Open Sesame, do not require extensive customer inputs and can be used for recommending many categories of products. Moreover, the move from GRS to GDS has some additional challenges, some of which will be discussed in Sections 5 and 6.

Both the GRS and GDS systems take the *customer-centric* point of view in providing personalized shopping services: they deal with a wide range of products that are not limited to one specific manufacturer and therefore cannot support the interests of a single manufacturer. On the contrary, they help the consumer to sort through a great maze of different products in the market and simplify consumer's purchasing activities. In the next section, we present e-Butler -- a customer-centric personalized shopping service that unifies the GRS and GDS services into one common concept.

4. The e-Butler Service

In this section we present an *Electronic Butler* service (or *e*-Butler for short) that provides a *customer-centric personalized* shopping alternative across a *wide range* of products. The notion of e-Butler, as we defined it, is a broad concept. It says that e-Butler provides personal shopping services for a *wide range* of products (as opposed to a limited group of products, such as books, CDs, or news) and that it should be *customer-centric*, i.e. it should be provided by an organization that is *not* associated with manufacturing or distribution of any of the products that it recommends or buys for its customers.

Since e-Butler deals with a *wide* range of products and services, it falls into the GRS and GDS quadrants of the framework presented in Section 3 (Figure 1). According to this framework, e-Butler can provide two types of services to its customers: it can either *recommend* purchasing of certain products and services or it can actually *buy* them without any prior notification or consultation with the customer. As was stated already in Section 3, we call the former type of e-Butler service *Personal Shopping Assistant (PSA)* and the latter *Magic Wand (MW)* service. We will describe them in detail below.

The e-Butler service can derive its purchasing recommendations or decisions from a variety of sources. In particular, they can be based on the customer inputs as in recommender systems, on the customer feedback as in the case of the content-based recommendation systems, on the customer profiles, on the analysis of customer's purchasing history, or on the combination of these factors. In the rest of this paper, we will focus on the type of e-Butlers that derive their purchasing recommendations or buying decisions from the analysis of the *comprehensive* past purchasing histories of their customers. In other words, at the heart of the e-Butler service considered in this paper lies the information on the past purchasing histories of its customers that are analyzed in order to make purchasing recommendations or buying decisions.

To be effective, e-Butler should collect data on *most* of the purchases a customer makes over some period of time (e.g. 80% - 90% of the *total* number of purchases made by the customer over the past 5 years). It is certainly very difficult to collect such comprehensive data now for both technical and behavioral reasons. We will discuss the ways to address these difficulties in Section 6.1. Since our objective in this section is to explain how the e-Butler service works, we only assume at this point that such purchasing histories are gathered somehow.

The general architecture of the e-Butler service is presented in Figure 2. It consists of the following five components: the "user" component, the retail environment component, estimated purchases module, and the PSA and/or MW service modules. We will describe these components in turn now and explain how the whole system works.

The "user" component (depicted with dotted lines in Figure 2) contains the comprehensive purchasing history of the user, as explained above. It also contains a user profile that is obtained from the analysis of the past purchasing history,[5] as well as some additional external information about the user, such as his or

[5] Therefore there is an arrow from the "purchase history" module to the "user profile" module in Figure 2.

her demographic and psychographic data. Finally, the "user" component also contains the information about the current "state" of the user, such as his or her current desires (e.g. the user wants to buy lunch, wants to buy a shirt, etc.), current location (e.g. the user is in Los Angeles on a business trip), current preferences (e.g. the user prefers to have a lunch at an expensive seafood restaurant), etc.

The component "retail environment" in Figure 2 contains information about the current "state-of-the-world", i.e., which products and services for which e-Butler is responsible, are sold in which stores, at which prices, and what sales and promotions are offered for these products, when and where. In other words, this component contains information about what is "going on" in the retail environment pertaining to the products and services supported by e-Butler. For example, a typical entry for the "retail environment" component is that Macy's sells Christian Doir's perfumes and will offer 30% discount on their entire line during the week of March 23.

The third component is the "estimated purchases" module that estimates the current purchasing needs of the customer based on his or her past purchasing history, on the customer profile, on the current "state" of the user and the current "state" of the retail environment. By knowing most of the past purchases made by the customer, as well as his or her current "state," e-Butler can *estimate* the customer's future purchasing needs.[6] For example, e-Butler may estimate, based on the past purchases, that the customer may need to buy a new case of red wine, a casual shirt, and a few bottles of shampoo for oily hair within the next month. By matching predicted future purchasing needs of the customer against the information on offerings of various products supplied by the "retail environment" component, the e-Butler generates the list of *estimated purchases* that the customer should consider making. For example, given customer's need to buy some red wine, e-Butler may realize that the customer would be happy with the choice of 1994 Merlot from XYZ vineyards because the customer likes Merlot, 1994 was a very good year for Merlot, this wine falls within the price range of previous purchases, and XYZ has a very good price for their Merlot, making this purchase a really good bargain. As another alternative, the customer may consider 1995 Merlot from ABC wineries which also provides a very good wine, a bit more expensive than XYZ but certainly worth the money. Using a similar type of reasoning, e-Butler can generate recommendations about shirts and shampoos the customer needs to buy and the places where they can be bought.

[6] Therefore, insistence on having *comprehensive* purchasing history. Otherwise, it is impossible to estimate accurately future purchasing needs of the customer.

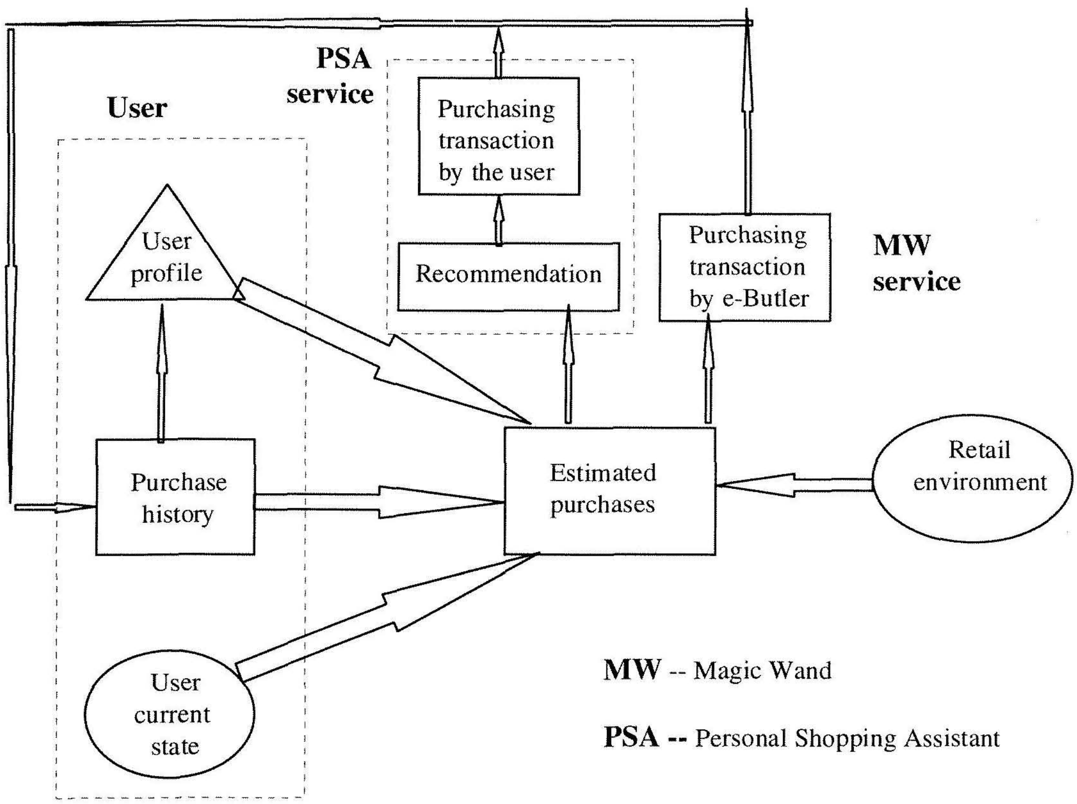

Figure 2. Architecture of the e-Butler System (see text for the details).

After e-Butler generates the list of products that the customer may need to buy, e-Butler has two choices (corresponding to the two arrows going from the "estimated purchases" module in Figure 2). As a first alternative, it can *recommend* the customer to purchase some of these products. For example, in the case of red wine, the PSA service of e-Butler can recommend the purchase of a case of the 1994 Merlot from XYZ wineries or the 1995 Merlot from ABC wineries. The "recommendation" module in Figure 2 is responsible for making the actual recommendation. After the customer receives recommendations from the PSA service, he or she makes the actual purchasing decisions (these decisions correspond to module "purchasing transaction by the user"). The second alternative is associated with making *actual purchasing decisions* based on the set of products that the customer may need to buy generated by the "estimated purchases" module of Figure 2. These purchasing decisions are produced by the module "purchasing transaction by e-Butler" in Figure 2. As explained before, this alternative corresponds to the *Magic Wand (MW)* service.

The e-Butler system works as follows (see diagram in Figure 2). First, a set of customer's purchasing needs is generated by the "estimated purchases" module based on the content of the customer's purchasing history, profile and the current-state modules. An example of a purchasing need may be the need to buy a new casual shirt. Then these estimated purchasing needs are matched against the current state of the retail environment and, as a result of this matching process, a list of estimated purchases is generated. An example of an estimated purchase is the recommendation to buy a Polo shirt from Macy's for $24.99 during the sale next week. If the PSA service of e-Butler is used, then this recommendation is presented to the customer along with others, and the decision to do an actual purchase is left to the customer. If the MW service is used instead, then this shirt is purchased by the MW service and is sent to the customer along with other purchases. In both cases, purchasing transactions are recorded as a part of customer's purchasing history, and the process continues.

To illustrate further how PSA and MW services work, consider the following examples. The first two examples deal with the PSA and the third example with the MW service.

Example 2. Based on the past purchasing history of a customer, the PSA service may discover that whenever the customer goes to Paris, she often buys wine there. The PSA service just received information that the customer purchased a ticket to Paris (this information is a part of the "user current state" module). Moreover, it knows that the duty-free shop at Charles de Gaulle airport has a very good price for the customer's favorite Chateau Margoux wine (better than anywhere else in Paris). This information is recorded in the "retail environment" module in Figure 2. Based on this information, the "estimated purchases" module comes with a recommendation to buy Chateau Margoux at the duty-free shop at the Charles de Gaulle airport, and the PSA service sends an appropriate message to the customer.

Example 3. Assume that a PSA provider company, in addition to collecting past purchasing history of its customers, supplies them with a Personal Digital Assistant (PDA) that has a wireless connection to the company's Personal Shopping Assistant (PSA) system. The PSA system estimates customer's future needs and behavior as described above. The purpose of the PDA device is to gather additional information on the current "state" of the customer (e.g. record customer's location information, preferences, desires (e.g. the customer is hungry now and wants to eat), etc.). This additional information is then transmitted from the PDA client to the PSA server over the communication lines. Using this additional information about the current "state" of the customer, the PSA system makes better estimates of the customer's needs and future behavior than without the PDA.

To illustrate how such combined service can work, assume that it is Tuesday, 11:30 am now, and the customer, Tom S., is driving in his car on I-87 in the Albany region. He has indicated to the PSA service through his PDA device that he wants to have a lunch. Based on Tom's past purchasing history, the PSA service knows that Tom, whenever traveling on business, likes to have light lunches at good quality restaurants (his company foots the bill) and that he likes sea food in general. By knowing that Tom is traveling now, the PSA service collects the following information on Tom: location – Albany region, day of the week – Tuesday, time -- 11:30am, traveling purpose -- business. This information is sent from the PDA client to the PSA server. By matching this information against Tom's purchasing history, the PSA concludes that Tom prefers a lunch at a good quality restaurant and he wants to eat light food. By matching this information with the fact that Tom likes seafood, the PSA is searching for highly rated seafood restaurants in the Albany region. If it finds any, it lets Tom know the choices by sending him an e-mail message and beeping his PDA. If the PSA service does not find any, it goes for other alternatives based on his past purchasing history (e.g. highly rated restaurants serving "light" types of food – not necessarily seafood).

Example 4. An example of the Magic Wand service would be the delivery of a case of the 1994 Merlot from XYZ wineries to the customer (without any prior notification of him/her). As another example of the MW service, Joe Smith who subscribed to this service, may discover one day that UPS delivers two shirts and a pair of jeans to his home address totally unexpectedly for Joe. Joe opens the package, examines the shirts and really likes them. Moreover, he realizes that he needs these shirts because some of his current shirts have become old and need replacement. Similarly, he needs a pair of jeans because his other pair is also becoming old and warn out. Moreover, he looks at the price tags and realizes that the prices are very competitive.

Besides buying clothes, other examples of automatic purchasing decisions made by the MW service, can include buying books, groceries, wines, and certain home accessories, renting movies, and taping certain programs on TV. Examples of purchases that are hard or even impossible to automate include purchases of big ticket items, such as houses or cars, concert tickets (or anything that requires knowledge of the person's schedule), and making travel arrangements.

It is important to note that in case of the PSA service actual purchasing decisions are made by the customer, not by the e-Butler system. Therefore, the PSA service acts as a *Decision Support System (DSS)* and falls into the GRS quadrant of Figure 1. In contrast to this, MW provides an example of a *decision automation* service and falls into the GDS quadrant of Figure 1. Both PSA and MW services deal with a broad range of products and therefore belong to the right two quadrants of Figure 1 ("broad" category quadrants).

Implementing the MW is certainly much more difficult than the PSA service for the reasons considered in Section 3 and explained further in Sections 6 and 7. In fact, to do this is a bold decision on the part of the company because the company must be certain that the customers would be happy with the products it sends automatically to them. Simply put, is the MW service realistic or is it too good to be true? The answer to this question depends on how much the customer is satisfied with the products being shipped to him or her. We will discuss the issue of achieving customer satisfaction with the MW service, with the

PSA service, and other related issues in the next section.

5. What Factors Make e-Butler a Successful Service

The success of the PSA service depends, to a very large extent, on the accuracy of predicting customer's future needs and providing the customer with *useful* recommendations. For example, if a customer finds 50% of the PSA suggestions useful for him or her (as in the case of buying wine in a duty-free shop at the Charles de Gaulle airport), the customer will, most likely, be satisfied with the service. If, on the other hand, the customer finds only 10% of suggestions to be useful, then the remaining 90% of useless suggestions will only irritate him or her, and the customer will reject this service. *Accuracy* of PSA recommendations is measured as a percent of the purchasing recommendations or transactions with which the customer is satisfied (as judged by the customer). For example, the accuracy levels in the two previous examples are 50% and 10% respectively. Another important measure is the *tolerance level* of a customer to useless recommendations. It determines the border line of accuracy rates at which the customer is indifferent to using the PSA service. If the accuracy rates are above the tolerance level, the customer will subscribe to the PSA service; if they are below the tolerance level, the customer will cancel it. Certainly, the tolerance level varies from one customer to another. For the author of this paper, the tolerance level constitutes, approximately, 25%, i.e., if less than 25% of the PSA recommendations are useful, I would refuse to use the PSA service.[7]

Similar argument also holds for the MW service. For each customer using that service, there must also be a minimal acceptable accuracy rate. For instance if Joe from Example 4 is not satisfied and returns every 10th item shipped to him by the MW service, he should probably find the MW service useful. The nuisance of returning items back is overcompensated by the advantages it provides to Joe (i.e., mainly, that he is freed from thinking what and where to buy and from actually going and buying the products). However, if Joe returns every second item back, he would certainly be unhappy with the service and, most likely, would cancel it.

Therefore, the key question for the MW service, as in the case of the PSA service, is what constitutes minimal acceptable accuracy rates for that service. In other words, at what accuracy rates Joe would still use the service and at what rates he would be irritated by its mistakes and the need to ship useless items back? As in the case of the PSA service, this borderline accuracy rate varies from one individual to another and will be called the *tolerance level* of that person. For the author of this article, the tolerance level for the MW service constitutes approximately 80%. In other words, if I have to return every 5th item

[7] Of course, this number varies from one person to another.

back because I do not like the item, I will have to think hard if I want to continue using the MW service[8].

Since the tolerance level of the MW service is much higher than the tolerance level of the PSA service, this means that the final decision whether to send a product to a customer or not should be much more stringent for the MW than the recommendation selection process for the PSA service. This means that the MW service should make a purchasing recommendation only if it estimates with very high certainty that the customer will be happy with the purchase.

Although very important, accuracy rates alone is not the only measure for determining if the customer would be satisfied with the PSA or MW services. For example, if the PSA service issues one purchasing recommendation per month or the MW service makes one purchasing decision per year, then these services would become irrelevant to the customers and hence few people would subscribe to them. Therefore, the second important factor determining the success of the PSA and the MW services is their *completeness*. Completeness is measured as a percent of the number of purchases made using the PSA or the MW services to the total number of purchases made by the customer.

Figure 3 shows under which conditions the customer is interested in the PSA or MW service. Accuracy is plotted on the x-axis and completeness on the y-axis. "Acc_min" on the x-axis of Figure 3 indicates the tolerance rates for the PSA or MW services. Of course, these numbers differ significantly between PSA and MW services. As was stated before, the tolerance rates for this author are 25% and 80% for the PSA and MW services respectively. Similarly, "acc_max" specifies the upper level for the accuracy rates that are practically attainable because of the technical, economical and behavioral considerations. On the y-axis, "comp_min" specifies the minimal completeness rate. If the completeness rate of the PSA or MW service is below this number, this means that too few recommendations or purchases are made to make this service useful to the customers. Finally, "comp_max" specifies the upper level of completeness that is practically attainable, i.e. the completeness rates above these levels are infeasible because it is very hard to give automatic recommendations on purchases of certain products (such as buying a house or a car) with current technologies. As with "acc_min," the other three constants "acc_max," "comp_min," and "comp_max" differ for the PSA and the MW services.

The four lines drawn in Figure 3 based on "acc_min," "acc_max," "comp_min," and "comp_max" divide the area into several regions. At the center is the shaded region defining the range where the PSA (or MW) services are feasible. To the left of it, separated by the acc_min line, lies the region in which the customer will be irritated and will abandon the PSA/MW service(s). Also, above the PSA (or MW) region (separated by the comp_max line) lies the area that is practically unattainable for the PSA/MW

[8] It should be noted that this is only a very crude estimation of the tolerance level. A more involved analysis depends on the return policies and conditions provided by the MW service. However, this detailed analysis is beyond the scope of this paper.

service(s). Below the comp_min line lies the area in which PSA/MW services, belonging to this region, would be useless because they would offer very few purchasing recommendations (for PSA) or purchasing decisions (for MW) to make any real impact on the customer. Finally, to the right of the acc_max line lies the area that is technologically, economically, and behaviorally infeasible because it is practically impossible for various reasons to achieve such high accuracy rates.

It is interesting to observe that the lower bounds (acc_min and comp_min) are determined by the *user*: the user will reject the e-Butler service if it produces the accuracy or completeness rates that fall below these levels. However, the upper bounds (acc_max and comp_max) are, to a certain extent, under our control. Their values are determined by several sources, some of which can be influenced by technological, behavioral, and economical factors (hence the meaning of the two arrows in Figure 3). We will address the issue of what determines acc_max and comp_max rates and how they can be increased in Section 6.

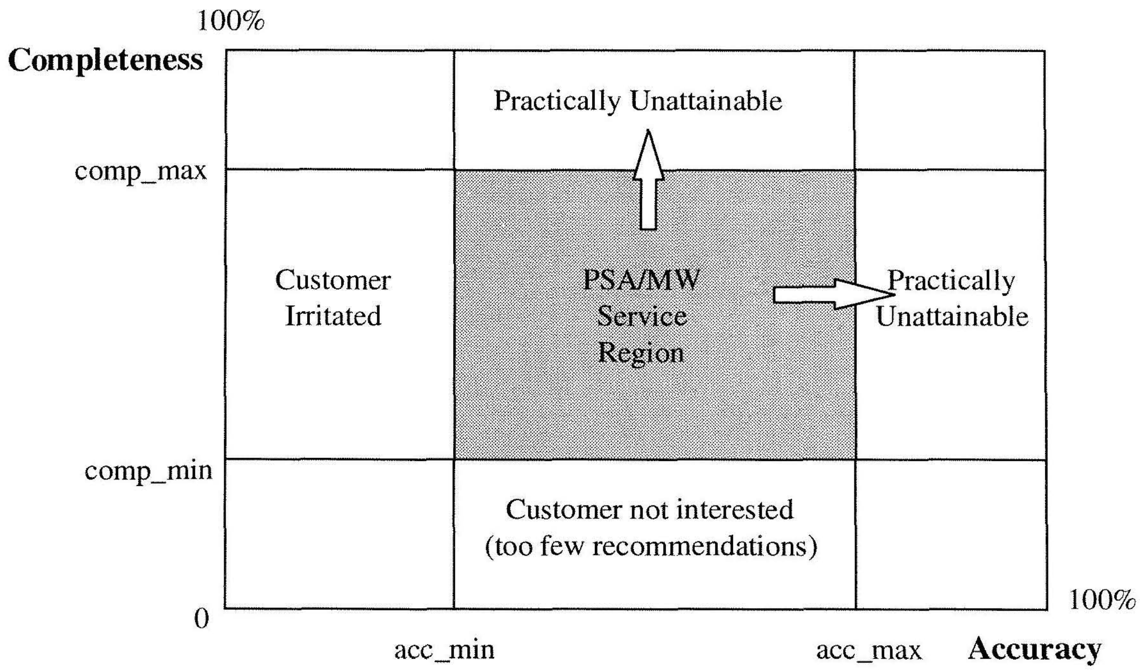

Figure 3. Customer's satisfaction with PSA and MW services. The shaded region is the area in which the customer is satisfied with the PSA/MW services.

Based on the discussions in this section, we identified the following major research questions:

1. How to achieve accuracy and completeness levels for the PSA and MW services that would land these services into the shaded region in Figure 3?
2. Once in the shaded region of Figure 3, how to increase accuracy and completeness rates that would push the PSA/MW services further to the "north-east" part of the diagram?
3. How to "push" out the upper boundaries (acc_max and comp_max) in Figure 3?

We will study these questions in the next section.

6. What Needs to Be Done to Make e-Butler a Successful Service

We posed three questions at the end of the last section. In this section we provide answers to these questions. It turns out that the answers to the first two questions are the same because the drivers that "push" PSA/MW services into the shaded region are the same that "push" these services deeper into the shaded "territory." In other words, the first two questions can be reduced to the question: How to improve accuracy and completeness rates for the PSA/MW services. We will address this question in Section 6.1. In section 6.2 we will address the second question: how to "push" out the upper boundaries "acc_max" and "comp_max."

6.1 How to improve accuracy and completeness rates for the PSA/MW services

As it follows from Figure 2, accuracy and completeness rates of the PSA and MW services depend on the accuracy and completeness rates that can be achieved by the "estimated purchases" and the "PSA (or MW) service" modules. In particular, it is important to (1) improve accuracy and provide a more complete set of estimated purchasing needs of customers; (2) given the set of estimated purchases, try to make more accurate and complete recommendations (for the PSA service) or actual purchasing decisions (for the MW service) based on this set.

The second issue is, essentially, technical and involves a trade-off between accuracy and completeness, which can be formulated as an optimization problem. In particular, if we have a list of purchases determined by the "estimated purchases" module in Figure 2, we need to decide which purchases from the list should be recommended to the customer (for the PSA service) and which should actually be purchased (for the MW service). If we recommend/buy all of the estimated purchases, the completeness value will be high, but accuracy low. If, on the other hand, we select only a few products from the list that we feel most confident about, the accuracy rate will be high but completeness rate will be low. Therefore, the key issue is how to make an "optimal" selection of products out of the list of purchases produced by the "estimated purchases" module. This problem has not been studied before to the best of our knowledge and constitutes an interesting technical problem. We will not address it in this paper, however, and will focus only on the first issue in the rest of this section.

The first issue is how to improve accuracy and completeness rates for the "estimated purchases" module in Figure 2. The accuracy and completeness of predicting customers needs (that are important both for the PSA and MW services) depend on the following factors:

1. *Comprehensiveness of customer's purchasing history.* The more complete this purchasing history is, the more accurately future purchasing needs of the customers can be estimated, and the more complete the set of these purchasing needs can be.
2. *Length of the purchasing history.* It makes a big difference if the e-Butler service has 3 months or 3 years of the purchase history of a customer. The longer the purchase history, the more accurate its predictions should be. Moreover, these predictions can be made about a larger set of products and services which should improve completeness rates as well.
3. *Inputs from the customers* regarding their current needs and preferences. This should serve as a valuable input for the e-Butler service. For example, if we know where the customer is located at the moment, as in Example 3, and we know what is the purpose of the trip (e.g. business), we can better estimate which restaurants to recommend to the customer. The more inputs we can elicit from the customer (without alienating him/her), the more accurate and complete set of predictions we can make. However, in order to minimize customer involvement in the recommendation and purchasing process and in order to avoid his or her irritation with the need to answer many distracting questions, the customer inputs should be obtained in a *non-interactive* fashion. In particular, these inputs should be obtained only once before the recommendation computations start. Otherwise, customer involvement would be too extensive, and he or she can get irritated and would reject the e-Butler service.
4. *Obtaining proper feedback* from the customer on the previously provided recommendations or purchasing decisions. The customer should provide feedback on how much he or she liked different recommendations or purchasing decisions of e-Butler. In its turn, the e-Butler service should take these recommendations/decisions into the consideration and revise its procedures in order to improve future recommendations/decisions. In other words, e-Butler should have *learning* capabilities, and these capabilities should improve accuracy rates of estimated future purchasing needs. These are the same capabilities as the ones used in the content-based recommendation systems discussed in Section 2.
5. *Good methods for analyzing past purchasing histories.* It is not sufficient just to collect past purchasing data. This data should be properly processed, extensively analyzed, and the most useful information should be extracted from it. One example of such useful information is customers

individual profiles. Such profiles would allow the e-Butler service to make better *inferences* about the future needs of the customers. The more precisely these profiles reflect the actual behavior of the customers, the more accurate and complete estimations of customer's purchasing needs can be obtained.

Therefore, in order to improve accuracy and completeness rates of the PSA and MW services, it is necessary to improve these factors. We will discuss the ways to accomplish this below.

6.1.1. Comprehensive Purchasing History.

The issue of obtaining a comprehensive purchasing history of customers' purchasing transactions is a complicated one for both technical and behavioral reasons. We will discuss these issues separately now.

Technical Issues

One way of obtaining a comprehensive purchasing history for the e-Butler service would be to issue a new (or use an existing) smart credit, debit, or e-cash card that records *individual items* purchased by the customer[9]. In addition, this card should be accepted in most of the shopping outlets. Finally, the customers should be encouraged to use this card in *most* of their transactions (we will discuss this issue below). This would allow the e-Butler service to collect most of the purchasing histories of its customers on an item-by-item basis.

Introduction of such a card is certainly a challenging project that has the following major problem. It requires installation of new Point-of-Sale systems that record information about *individual items* purchased either on the card directly (using the smart-card technology) or transfer this information to the e-Butler system through communication lines. For this feature to be useful, it should be installed at most of the shopping outlets, which makes it a very large project.

The industry has already taken some steps in this direction. One example of a more "intelligent" Point-of-Sale (POS) system is the Pinnacle suit of products from Hypercom, Inc. The Pinnacle Transaction Environment allows electronic financial transaction processors to provide value-added features to their clients' payment systems. In particular, a financial transaction consists of the core transaction, containing the merchant number, personal account number and the total transaction amount, and the value-added data, containing data such as the product code, Card Verification Value, tip amount and the POS

---

[9] In comparison, most of the existing credit or debit card transactions record only the total amount of purchases in a transaction without listing individual items purchased. In some cases, one can obtain individual item information; however, it is not easy or even impossible to do this in general.

condition code. Pinnacle extracts this value-added data and stores it on the Pinnacle Network Terminal Server. One implication of this is that this server contains information on the *individual products* purchased by customers that arrive to the server from the connected terminals. If individual Pinnacle Network Terminal Servers can be connected into one system that shares the data from individual servers, then this data can contain a comprehensive collection of purchases of some of the customers.

In summary, it is not easy *today* from a purely technical point of view to collect comprehensive histories of customer purchases. However, new technologies are "around the corner," and we believe that customer comprehensive purchasing histories can be collected using these new technologies within the next few years. However, we also believe that the "bottleneck" is not technology but customers' behavior. We will discuss this issue in the next subsection.

Behavioral Issues

Even if it becomes technologically feasible to collect comprehensive purchasing histories, it is not clear if the customers want to do that. In fact, as some evidence indicates [Venkatraman98], many customers resist the idea of putting all of their financial information, such as checking and saving accounts, CDs, credit card, investment, and insurance information into one place. Certainly, purchasing information is quite different from the personal financial information, and many people might feel different about putting most of their purchasing information into one source. For example, some people link their credit cards with a frequent flier program and use the same credit card for most of their purchases to earn extra free miles. Nevertheless, it is still not clear how many people would follow this model rather than the financial information model. Therefore, we can encounter resistance from many customers to placing their comprehensive purchasing information into one source.

However, this problem can be addressed as follows:

- *Educate the customers* on the advantages of the e-Butler system and demonstrate what value it can bring to them. As was demonstrated in Section 4, e-Butler provides much value to its subscribers. The PSA service frees busy customers from the chores of making purchasing decisions and shopping around for the products they needs. In addition to this, the MW service automates purchasing decisions and frees the customers from the shopping process itself. Using various marketing methods, these advantages should be communicated very clearly to the right segments of the population. The primary segment initially targeted by the e-Butler system should be the busy professionals who have neither time nor desire to shop in a conventional way. After this initial segment "buys" into the e-Butler concept, and e-Butler demonstrates its value in practice, e-Butler can then be targeted to other segments of the population.

- *Address the privacy issues* and assure the customers that their purchasing information will be kept in *strictest* confidence. One of the main reasons why many people do not want to put all of their financial information into one source is the concern over the *privacy* of this information. This is certainly a very serious problem. Even though such companies as Firefly, Open Sesame and others address this issue by making strong commitments to consumer privacy by proposing and adhering to strict privacy standards, such as Open Profiling Standard, the concern for privacy still remains [Markoff98]. However, we believe that the concern over privacy can be properly addressed in the context of the e-Butler service because it will be provided by some company that will have *vital interests* in keeping customer's data in the strictest form of privacy. This is the case because the survival of the company itself is at stake if customer privacy is breached. Therefore, this company should keep customer information in such confidentiality as banks keep customers financial information, or Coca-Cola keeps the Coke secret. Still there is a task of convincing the customers that this is the case, and this task is not an easy one. However, we believe that with the right degree of persistence, this task can be accomplished.

- *Provide additional incentives for enforcing behavioral patterns* that allow to collect comprehensive purchasing histories, such as accumulation of frequent flier miles, giving bonus points, etc. In fact, some people try to use the same credit card in *most* of their purchasing transactions in order to earn more frequent miles using that credit card. This is a good indication that these incentives can work for e-Butler as well.

Another problem pertaining to collecting comprehensive histories of customers purchases is that not all of these purchases are made for the individual consumption. Some of the purchases are made as gifts. It is difficult to differentiate between personal and gift purchasing behavior of the customers, and this makes it more difficult to estimate their future purchasing needs. For example, assume that a customer purchased two shirts as a gift. The e-Butler service would think that the customer has enough shirts and would not recommend any shirt purchases for some time, while the customer may need new shirts. However, one can use various methods to differentiate between gifts and personal purchases. For example, the e-Butler service can encourage the user to record gift purchases. Also it can use various heuristics to distinguish between the two. For example, if the purchased shirt is of the size that is different from the sizes of the household members, this means, most likely, that it is a gift. Finally, the gift biases would lead only to the overestimation mistakes (e.g. the e-Butler service thinks that the customer has enough shirts, whereas he needs to buy some more). These mistakes would *not* result in wrong (and annoying) recommendations that are the most harmful for the e-Butler service.

In summary, it is harder to resolve behavioral issues hampering the collection of comprehensive purchasing histories for e-Butler customers. However, this is not an impossible task, and we believe that, eventually, it could be successfully resolved.

6.1.2. Length of Purchasing Histories.

Clearly, the longer the purchasing history of a customer is, the more accurate estimation of future purchasing needs e-Butler can make. This means that the most serious problem is with the new users subscribing for the e-Butler service because they don't have any purchasing histories recorded by the service. Therefore, the e-Butler service should first accumulate the purchasing histories for some time before it can start giving purchasing recommendations or making buying decisions for the new customers. This means that a system of incentives should be developed for the first few months after a person subscribes for that service and before purchasing recommendations or buying decisions start being issued. Alternatively, other methods for providing recommendations and/or buying decisions can be used initially. For example, any of the recommender systems [CACM97] can be used for that purpose.

In addition, the first purchasing recommendations will be based on a short purchasing history, and therefore, the e-Butler service will be able to make initially only few reliable recommendations. However, both completeness and accuracy of the e-Butler service should increase over time, when longer purchasing histories are generated (and the customers provide feedback on previous recommendations as discussed in Section 6.1.4).

6.1.3 Inputs from the Customers.

These inputs are very valuable and certainly should be encouraged by the e-Butler service. One incentive for the customers to specify their current and anticipated needs lies in the increased accuracy and better service provided by the e-Butler system. Nevertheless, it is also useful to develop an additional system of incentives to facilitate even more inputs from the user.

However, excessive requests for inputs can irritate customers and make them abandon the service. For example, if a customer is asked for inputs by e-Butler 100 times a day, he or she would certainly be irritated and would quit the service. Therefore, requests for customer inputs should be made very judiciously by e-Butler.

6.1.4 Customer Feedback.

Once the e-Butler service starts issuing recommendations or purchasing decisions, it is important to obtain the feedback from the customers on how much they liked or disliked different recommendations or decisions. This valuable information should be fed back to e-Butler and the system should *learn* how to revise its recommendation or decision making strategies based on the customers' feedback. There are numerous learning methods developed in the Machine Learning branch of Artificial Intelligence field that can be applied to this problem [Mitchell97]. As we discussed in Section 2, content-based recommendation systems use some of these techniques, and e-Butler can utilize these approaches.

6.1.5 Methods for Analyzing Purchasing Histories.

Customer's purchasing needs generated by the "estimated purchases" module in Figure 2 should be determined based on the patterns and trends detected in the purchasing histories. For example, it may be discovered that a customer tends to buy Bordeaux wine when she goes to Paris. This pattern is certainly more useful for recommendation and purchasing capabilities of e-Butler than a simple collection of facts enumerating all the purchases that the customer made in Paris.

There are several approaches proposed on how to analyze purchasing histories of customers. They include collaborative filtering, content-based recommendations, non-intrusive learning, and user-interaction methods. Collaborative filtering was described in Section 2 within the context of collaborative systems, such as GroupLens from Net Perceptions, Preference Server from LikeMinds, and Passport-based systems from Firefly. Similarly, content-based recommendation systems, such as Instant Recommendations from Amazon.com, *First!* from Individual, and Syskill&Webert were also described in Section 2. All of these systems require extensive user participation in the analysis of purchasing histories and recommendations: collaborative filtering systems require extensive user inputs and content-based recommendation systems require extensive customer feedback. Therefore, there is a danger of overburdening the customer with excessive requests for information.

In contrast to this, the Learn Sesame system from Open Sesame constructs *individual* customer profiles from the Web click-stream (Web logfile) data in a *non-intrusive* manner without requiring extensive customer inputs. These profiles are build by analyzing histories of customer interactions with a Web site. Learn Sesame uses neural network technology for the analysis of the click-stream data and for the construction of individual profiles.

Learn Sesame discovers statistically strong patterns in the customer's click-stream data. However, it is our firm belief that it is, ultimately, a human who has to examine and validate the patterns for their usefulness and strength. Therefore, we are currently developing new methods for validating individual customer patterns obtained from purchasing histories through the involvement of the user in this process.

In summary, we considered various methods for improving accuracy and completeness rates of recommendations and purchasing decisions made by e-Butler. Although some of these methods a relatively easy to implement, others require modification of customer's behavior, which is very hard to achieve. However, we believe that none of the issues are insurmountable, and all of them can be resolved in due time. However, much work is needed for that purpose.

6.2 How to "push out" the upper boundaries "acc_max" and "comp_max" in Figure 3

We will first discuss the issue of "pushing out" the value of comp_max (this corresponds to the upper arrow in Figure 3) and then the value of acc_max (right arrow in Figure 3).

Completeness (comp_max). Consider such products as books and CDs, and other products, such as houses and cars. It is certainly much easier to recommend new books and CDs rather than new homes or cars. This is the case because we usually buy books and CDs much more often than houses or cars and because purchasing a book or a CD is a much less involved decision than purchasing a house or a car (purchasing a new house or a new car is such an unstructured decision that it is very hard to design machine-generated recommendations for these products). Therefore, books and CDs are on one extreme and "big ticket" items, such as houses and cars, are on the other extreme of the ease-of-recommendation spectrum.

In general, there are products for which it is easy to provide purchasing recommendations and other products for which it is very hard (or even impossible). Between these two extremes lies the whole spectrum of products for which it is "somewhat" hard to provide a good recommendation. An example of the "middle-of-the-road" product is a vacation. Given several years of the past purchasing history of a customer, including a few vacations taken in the past, it is possible to come up with a recommendation for the next vacation the customer may like, although it is not easy to do so.

One way to increase completeness rates for the e-Butler service is to identify such "middle-of-the-road" products and services and to develop *special subsystems* that specialize in the recommendations for these types of products[10]. For example, a vacation-planning recommendation system would deal exclusively with vacations, and a car-buying recommendation system with cars. One advantage of this approach lies in that these expert recommendation systems can utilize domain-specific knowledge that helps produce better recommendations than the generic recommendation system. Therefore, they could help to increase the upper bound of completeness rate (comp_max). An obvious disadvantage of this approach lies in that this is a step back to the Expert Recommendation Systems and the drawbacks related to that class of

[10] We would like to point out that these special subsystems for the "middle-of-the-road" products would *complement* a generic recommendation system for the easy-to-recommend products, such as books, wines, and CDs. In other words, the two types of systems would operate side-by-side rather than one replacing the other.

systems. In particular, this is certainly a very labor-intensive process requiring building an ERS system for each product category (vacations, cars, etc.).

Accuracy (acc_max). In order to "push" the value of acc-max up, we need to identify the factors that affect the accuracy rates. In Section 6.1 we identified five factors that affect these rates. Therefore, we need to determine the components of these factors that are very hard (or even impossible) to influence (and thus the accuracy rates will not go above the limits set by these factors). In order to answer this question, we examine each of them one by one.

Comprehensiveness of purchasing history. We identified technology-based and behavioral factors that affect the completeness of a purchasing history. We also argued that the technology-based factors can be affected easier than the behavioral factors. Therefore, behavioral factors, such as unwillingness to provide complete purchasing history and unwillingness to purchase all of the products with a smart card identify one type of limit for the accuracy rates.

Length of purchasing history. This is, certainly, not a factor.

Customer inputs. This is another factor limiting accuracy that is very hard or even impossible to influence because it is very hard to convince a customer to provide inputs if he or she does not want to do so. Therefore, lack of customer inputs is still another factor limiting recommendation accuracy.

Customer feedback. This is still another limiting factor that is hard or even impossible to influence. As in the case of customer inputs, it is very hard to convince the customer to provide feedback if he or she does not want to do this.

Methods for analyzing purchasing histories. This is not a limiting factor because these methods can be perfected as our understanding of them advances.

In summary, all of the factors imposing "hard-core" constraints on the value of "acc_max" are *behavioral* and constitute unwillingness on the part of the customer to 1) provide a complete purchasing history to the e-Butler service, 2) purchase most of the products with a smart card, 3) provide systematic inputs about his or her preferences, and 4) provide feedback about the levels of satisfaction with the purchased products and services. We would also like to point out that all of these behavioral factors vary significantly from one customer to another.

In this section we explained what needs to be done to make e-Butler a successful service, i.e., what needs to be done to move it into the shaded region in Figure 3 and further up into the north-eastern part of that region. In particular, we identified the following critical issues:

1. how to collect a comprehensive purchasing histories of e-Butler customers
2. how to start e-Butler service (without any prior history)
3. how to move the upper boundaries (acc_max and comp_max) in Figure 3 (the low boundaries acc_min and comp_min are determined by the user and are fixed and not "movable")
4. how to do the analysis of the past purchasing histories well in order to make better estimations of future purchasing needs
5. how to make actual purchasing recommendations (for PSA service) and purchasing decisions (for the MW service) based on the estimated purchases analysis done by e-Butler.

Although we explored some of these issues in the paper, certainly, more work is required in order to understand them well and to resolve all of them successfully.

7. Feasibility of the e-Butler Service

e-Butler has not been implemented yet and, therefore, we don't have any first-hand evidence that it will work in practice. However, we believe that the technical issues, discussed in Section 6, pertaining to the successful implementation of the *PSA* service will be resolved soon. Therefore, the bottleneck lies in the behavioral issues identified in the previous section, such as unwillingness of the customers to provide complete purchasing history, unwillingness to purchase most of the products with a smart card, unwillingness to provide customer inputs and customer feedback. Moreover we argued in Section 6 that, although these bottlenecks are serious, they are not unsolvable, and could be overcome. We believe that if we manage to overcome these behavioral bottlenecks and when the technical solutions are in place, the PSA service should be able to achieve the accuracy and completeness rates acceptable to the public (in other words, it should land in the shaded area in Figure 3).

The next question is whether or not the Magic Wand service is feasible. As was explained in Section 5, the accuracy rates for the Magic Wand service should be much higher than for the PSA service (e.g. above 80%). Is it feasible to achieve such rates even if we manage to resolve all the technical and behavioral problems described in the previous section?

We believe that it should take a very intimate understanding of the customer and his or her needs, including a very thorough analysis of customer's purchasing history over a long period of time to achieve such high accuracy rates for *selected* products. Examples of such selected products are books and wines. If we study purchasing histories of people interested in purchasing books or wines over a long period of time, we can know their preferences and habits well enough to start unsolicited sending of these products to them and expect acceptable accuracy rates. Examples of items that should never be used in the MW service are concert tickets or any other products, the consumption of which requires knowledge of customer's schedule. Moreover, in no case should the MW service be offered to a new or a recently joined customer. However, if we start offering MW service for *selected* products, then we fall back to the EDS type of service (from Figure 1), such as the book-of-the-month or similar clubs, and the service becomes no longer generic.

Therefore, we believe that it is infeasible to start the MW service from "scratch." What could be done instead is the initial introduction of the PSA service and gradual development of a one-to-one relationship with a customer by accumulating his or her purchasing history and constantly studying the customer's feedback. Once the customer is well-known to the PSA service, *gradually* and very *cautiously* we can start introducing the MW service on selected products as an *augmentation* to the PSA service, carefully studying the customer's feedback. If the customer feedback is positive, we can gradually expand the MW service to a larger set of products until we reach some stable point of satisfactory completeness and accuracy rates. However, if the accuracy rates are not satisfactory for a customer, we can scale down the MW service for that customer, leaving him or her only with the PSA service.

This approach corresponds to the ERS $\rightarrow$ GRS $\rightarrow$ GDS path in Figure 1, where the GRS $\rightarrow$ GDS transition should be made very carefully and cautiously. However, we believe that the alternative path ERS $\rightarrow$ EDS $\rightarrow$ GDS is also promising. In particular, it avoids the bottlenecks of the ERS $\rightarrow$ GRS transition. This approach corresponds to the development of expert decision-making services that are similar to the book-of-the-month clubs and then gradually and cautiously expanding them to other types of products and services, such as a wine-of-the-month, a video-of-the-moment, a shirt-of-the-season type of clubs. This is, certainly, a very labor-intensive process because each of the product types has a corresponding expert system designed for that product. But once we get enough experience with designing such expert decision-making systems, we can start building *generic* decision-making systems. Hopefully, the PSA service will mature at that point, and we will be able to merge the ERS $\rightarrow$ GRS $\rightarrow$ GDS and ERS $\rightarrow$ EDS $\rightarrow$ GDS paths into one integral approach. Such approach corresponds to the arrow from the ERS to the GDS quadrant in Figure 1.

In summary, we believe that it is feasible to provide the MW service for a *selected* group of products and services, but it should be done very carefully and very cautiously in a manner described in this section.

8. Conclusions

In this paper, we described an *Electronic Butler* (or *e-Butler)* that provides a personalized shopping service to its subscribers by identifying individual customer's shopping needs and providing purchasing recommendations across a wide range of products or making purchases directly without any prior consultations with the customer. We also identified the two components of e-Butler -- the Personal Shopping Assistant (PSA) that provides purchasing recommendations and the Magic Wand service that makes direct purchases.

We also addressed the issues of feasibility of the PSA and MW services and identified technological and behavioral constraints for the successful deployment of the e-Butler service. Since e-Butler has not been implemented yet, it is impossible to provide any hard-core evidence of its feasibility. However, we argued that the technologies for solving all the technical constraints associated with the implementation of e-Butler are either in place now or are "around the corner." It is more difficult to address the behavioral constraints. However, we maintain that they are not insurmountable and will be solved in due time.

In this paper we considered an e-Butler service that makes *personalized purchasing* recommendations or buying decisions. Clearly, it can be extended to other types of personalized recommendation and decision making services, such as personalized health care and personalized Web-browsing services. For example, e-Butler can be extended to include personalized recommendations about when to visit a doctor, which medicines to take and when, which exercises to do and when, and which types of foods to eat based on the health considerations. Although we have not considered them in the paper, these services should not be difficult to incorporate into e-Butler because they use the same types of technologies and because they have similar types of behavioral constraints.

9. Bibliography

[Balabanovic&Shoham97] "Fab: Content-Based, Collaborative Recommendation," C*ommunications of the ACM*, March 1997, vol. 40, no. 3, pp. 66-72.

[CACM 97] Special Issue on Recommender Systems," *Communications of the ACM,* vol. 40, no 3, March 1997.

[Goldberg et al 92] Goldberg, D, Nichols, D., Oki, B. M. and Terry, D. "Using collaborative filtering to weave and information tapestry." *Communications of the ACM*, 35(12), December 1992, pp. 61 – 70.

[Lang95] Lang, K., "Newsweeder: Learning to filter netnews," *Proceedings of the 12th International Conf. on Machine Learning,* (Tahoe City, CA), 1995.

[Markoff98] Markoff, J. "Firefly sale sparks debate on privacy," *International Herald Tribune,* April 11, 1998.

[Mitchell 97] Mitchell, T. M. *Machine Learning.* McGraw-Hill, 1997.

[Pazzani et al] Pazzani, M., Muramatsu, J. and Billsus, D., "Syskill & Webert: Identifying Interesting Web Sites," http://www.ics.uci.edu/~pazzani/

[Peppers&Rogers 98] Peppers, D. and Rogers, M. "Concierge Services and Customer Loyalty," *INSIDE 1to1*, On-Line Newsletter published by Marketing 1to1/ Peppers and Rogers Group, May 28, 1998.

[Peppers&Rogers 93] Peppers, D. and Rogers, M. *The One-to-One Future.* Doubleday, 1993.

[Pine et al.95] Pine, B. J. II, Peppers, D. and Rogers, M. "Do you want to keep your customers forever?" *Harvard Business Review,* March-April 1995.

[Resnick&Varian97] Resnik, P. and Varian, H. R. "Recommender Systems," *Communications of the ACM,* vol. 40, no 3, March 1997.

[Stohr&Viswanathan99] Stohr, E. A. and Viswanathan, S. "Recommendation Systems." In *Emerging Information Technologies,* K. E. Kendall (ed.), Sage Publication, Inc., 1999.

[Venkatraman98] N. Venkatraman, Personal communications, Boston, MA, April 1998.