Using Demonstration to Promote Information Products

Metin Çakanyıldırım* and Tevfik Dalgıç

School of Management, University of Texas at Dallas, Richardson, Texas 75083.

May 21, 2002

Abstract

Value of information products, such as software, cannot be determined at a glance but it is critical for purchase decisions. Customers should test and/or learn these products and compare the product capabilities against their needs. Vendors allow customers to evaluate the products via demonstration. Either they make products available to customers for free for a limited amount of time (demonstration phase) or they hand out a limited (demonstration) version of the product without any time limitation. We coin the term "demo phase (version) strategy" for the former (latter) demonstration strategy, and the "combined strategy" for the combination of the two. Customers value a product by summing values of satisfactory features discovered during the demonstration. We study vendor profits as a function of strategy parameters, the justification value, the number of features in a demonstration version and the length of demonstration phase. We argue that version and phase strategies can be made equivalent in terms of product value by choosing demonstration parameters appropriately. We establish unimodality properties of profit function and provide closed form solution for a special case. We show that the combined strategy is dominated by the phase strategy. We also provide numerical experiments and draw managerial insights.

*metin@utdallas.edu

1 Introduction

In this paper we will study how customers use the demonstration mechanism ("demo" from now on) to evaluate information products, such as software, electronic magazines, archives. During demonstration, customers are given free access to products (or some versions of them) to experiment with product features and to acquire product specific knowledge. The knowledge acquired with experimentation accumulates over time and is never forgotten. We can envision the knowledge about a product starting from a certain level and increasing to another level in time as suggested in [21]. We will use this learning process to analyze the evaluation of the products. Evaluation will correspond to mapping the product specific knowledge to product value. We will present customers' buy-or-not decision in conjunction with product value and price. In this context, demo is a valuable mechanism to convince customers of the high value of products as confirmed in [9] by an AOL manager who reports that: "the trial offers are worthwhile because nearly three-quarters stay on as full-paying members." We will introduce two principal demo strategies and study the trade offs involved in finding optimal strategy parameters.

In practice, the popularity of demo is rising because of three fundamental factors. First, information products are becoming complex, this trend is especially pronounced in the software industry. This makes it harder for vendors to communicate the features of the products to potential customers. Traditional promotion channels, such as advertisements and brochures, are not very effective for complex products. Realizing this, vendors are allowing customers to have first hand experience of products during demo. Second, technology has developed enough to make demo a viable alternative for promotion. Doubtlessly the biggest contributor in this respect is the Internet. It makes access to products convenient and hence is extensively used by vendors. Using cookies, it is technologically possible to disable a product at the end of demo and prohibit multiple demo accesses/downloads of the same product onto the same computer. Finally, two conditions must be satisfied roughly before a purchase: first product awareness and then sufficiently high perceived (by customers) product value. The technology is making the product search so easy that focus has been shifting from increasing product awareness to convincing the customer for a purchase. Although a primary application area of our models would be the promotion of small-scale software on the Internet, large-scale software vendors also flirt with the demo strategies. Some enterprise planning and supply chain management software vendors are willing to hand out their software to potential customers for demo purposes, if they believe that these customers are likely to make a purchase.

In addition to above mentioned three factors justifying the study of demo strategies, focus on information products supports the validity of our modelling assumptions. We do not study production costs because additional cost of producing one more information product is negligible, once the first product is designed and developed. This is supported by the observation that customer feedbacks are not implemented immediately but are considered for the future releases of software, so additional customers / feedbacks do not increase the production cost of the current release. This allows us to use "profits" and "revenues" synonymously. We assume that evaluation process is the same for individuals making purchase decisions for themselves or for an organization, so observations of this paper apply to both individuals and organizations who would be called customers. We assume that product testing causes negligible costs to customers.

The Internet allows small and large international companies market products to a large customer base scattered around the globe by eliminating obstacles created by geography, time zones, and location [22]. This is more so for small companies which cannot afford to promote their products through more expensive marketing channels, promotion and publicity campaigns and industrial fairs. The Internet has decreased marketing expenses for many medium and small companies to go international beyond their home country markets [11]. Especially software companies put up their products on the Internet and illustrate the capabilities of the software to potential customers before the purchase. This has created opportunities for global software producers to distribute their products [1], many companies, including GE, Bank of America, Target and American Express, are going offshore to develop and maintain their software. Because of not having an established brand name, small and global software producers have hard time in selling their products. On the other hand, customers find it equally hard to find and evaluate software packages to satisfy their needs. This hardship is more pronounced for small companies which cannot afford to dedicate large resources to software evaluation [12]. Demo strategies solve not only vendors' promotion problem but also help customers to evaluate software.

Product evaluation relates to new information gathering, updating existing information and perception formation tendencies of customers as they use or test products. [16] suggests that beliefs about an attribute of a product is a weighted combination of initial beliefs and observed beliefs. [4] argues that, while choosing a auto repair service center, a consumer's experience with a center is more important than externally obtained information on that center. [2] studies how customers' experience with a product determine their satisfaction. [7] study consumers' brand choice process while product diffusion is happening. [3] argue that customers may construct preferences over time rather than having a stationary set of preferences. This can happen when there are contingencies among preferences where the value of each preference depends on the value of other preferences. On the other hand, [18] investigated the issue of persistent preferences for product attributes which are always favored over other preferences. It also considers product trial as a necessary tool to develop an opinion about a brand as result of key attributes of that brand.

Some product attributes (price, user friendliness) can be more easily evaluated than others (reliability, compatibility). Those attributes that take time to be evaluated cause product information to be incomplete initially, say at the first contact with the product. In a recent review, [15] mentions two approaches about treating incomplete information. Incomplete information generally causes customers to discount the value of a product and/or ignore the attributes with incomplete information. [23] show that when consumers have uncertain beliefs about products, they offset the expected utility of products by a certain multiplier of the variance of beliefs about the utility. Beliefs about an attribute can also be obtained by updating prior beliefs with observed beliefs in a Bayesian fashion [23].

Marketing literature have investigated product evaluation from different points of view. [10] categorizes products according to how attributes are learned: play products, functional products, time products, undemonstrable products. It places information products into time product category, saying that discovery of their attributes take substantial time. [14] and [25] distinguish attributes as experimental (called search in [25]) attributes which are observed after some use, and nonexperimental (called learning in [25]) attributes. According to [10], attributes of learning (search) products can be evaluated before (only after) the purchase. [14] first argues that product trials have powerful influences on brand evaluation and considers only experimental attributes during trial.

Software products are not physical goods so they do not wear down. Once they satisfy customer needs, they may be used for a long time unless the needs change. Although each producer has its own development plans, upgrades take place frequently. Consequently, customers often buy different versions of a product. Since different versions have different functionality, we treat every version as a different product. As a result, all software customers are first-time buyers. Demo strategies are logical approaches to attract first time buyers. Unfortunately, most of the previous studies address buy-or-not decisions for repeat buyers. In the context of first time buyers who evaluate a product and decide during the product demo, literature lacks a study of product valuation, pricing and associated profit. Therein lies the contribution of this paper where an attempt is made for using demo as a promotion method and its interaction with pricing.

The rest of this paper is organized as follows: In the next section, we introduce version and phase strategies, and obtain product values as a sum of values of evaluated features. In Section 3, we express vendor profits and study their unimodality in decision variables. This is ensued by numerical examples and associated managerial insights. Finally, we provide a brief conclusion in Section 5. To maintain the flow of our arguments, we postpone most of the proofs to the Appendix.

2 A Model of Product Evaluation

Customers want information products that can reliably perform certain functions and meet the current needs. When vendors release products, they claim that products perform some functions. However, customers cannot be certain about the performance of a new product because of potential problems such as low reliability, bugs, lacking compatibility with existing products. Thus, customers prefer to evaluate product performance preferably at their own site, before paying for them.

We envision a product as a bunch of features, where we use the term “feature” to mean a single function of given product. During a demo, customers evaluate the features one by one to see if they are functioning properly. At any time during the demo, customers have a set of features already evaluated and the remaining set of features to be evaluated. The value of the evaluated and properly functioning features can be assessed right away. However, the value of features to be evaluated is uncertain because customers have incomplete information about them. Product evaluation can be done in several ways. In the first approach, customers learn the features of a product one by one while evaluating it. Customers may tolerate a couple of bugs or unexpected outcomes during learning, although these problems may lower the perceived value of the product. Too many problems will naturally make customers reject the product. In the second approach, customers test the features of the product with test cases prepared in advance [17]. Naturally, the outcome of these tests must be verifiable independently, without using the product being tested. Some vendors use test cases to test products before the release. [19] and the references therein provide some literature on economics of product testing by vendors.

Two demo strategies are currently popular among vendors: demo phase and demo version. In the demo phase strategy, customers have access to the entire product for a limited time, i.e. during demo phase. In the demo version strategy, customers are given a limited (demo) version of the product for forever. The demo version includes only a subset of the features of the full version. Since learning a product requires more time than testing it, customers who evaluate a product by testing are less time sensitive. Since the demo version strategy does not account for time, it is more appropriate for testing. Both demo phase and version strategies can be relevant for learners, although the demo phase explicitly studies learning speed of customers and hence provides a richer model.

At the beginning of a demo, customers may know all features (promised to work properly) if vendors mention these in product manuals. In practice, the total number of features is much larger than the number of features learned during a demo. Software tend to have many features, some of which are never used even years after the purchase. Along these lines, [20] says that: "…our experience tells us that the typical UNIX users only use about 40% of the tools and utilities available to them." Let $\mathcal{U}$ be the set of features of a product and set $U = |\mathcal{U}|$, these observations motivate taking U as infinity when necessary. Let S be the set of features that function properly according to customers' specifications/needs. We call features in S as successful (good) features and the others as defective (bad) features. Since S usually relates to customer specifications, a vendor cannot determine S. Even when S is independent of customer specifications, the vendor may not know the set of the good features [19]. Neither demo phase nor demo version strategy allows customers to discover entire S because customers do not have enough time or all features, respectively, in these strategies. Consequently, demo allows customers only to estimate S.

We define the value of product features using a value function V. We require two natural axioms on V:
Positivity: $V(\emptyset) = 0$ and $V(C_1) \leq V(C_2)$ if $C_1 \subseteq C_2 \subseteq \mathcal{U}$. Additivity: $V(C) = \sum_{k \in C} V(\{k\})$ for all $C \subseteq \mathcal{U}$.
Positivity merely states that the value of extra features cannot be negative. Additivity can be motivated by assuming that there is neither synergy nor duplicity among product features. For an example of synergy consider Netscape, Internet Explorer web browsers and Outlook e-mailing software. The former browser comes with built-in e-mail reading/sending capability unlike the latter browser. Some customers may value Netscape more than Explorer and Outlook together. The driving force behind synergy is the usefulness of integrated features. However, accounting for information products without additivity can be hard to manage, so additivity is often built-in to cost models. For example, McGraw Hill sells books chapter by chapter and prices each chapter independent of other chapters.

We use value function to assign a monetary value to product features. The product value is $V(\mathcal{U} \cap S) = V(S)$ but S will appear as a random set to a customer and ranges from $\emptyset$ to $\mathcal{U}$. Let p be the success probability for each feature in U. We assume that this probability is the same for all features. Then, treating value functions as random variables we obtain that

$$V(\{i\} \cap S) = \bar{Y}_i = \left\{ \begin{array}{lll} Y_i & \text{with probability } p & \text{i.e. feature } i \text{ is good} \\ 0 & \text{with probability } 1-p & \text{i.e. feature } i \text{ is bad} \end{array} \right\}$$

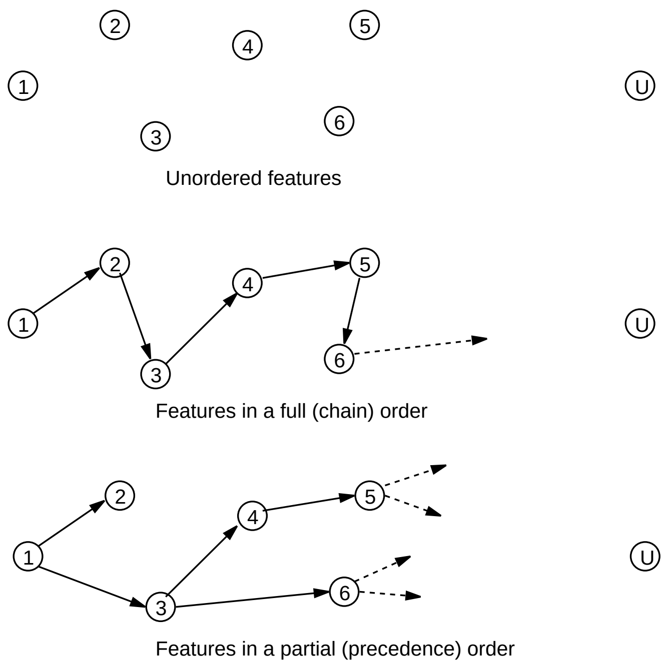

Figure 1: Unordered, fully ordered and partially ordered product features.

where $\bar{Y}_i$ (Y_i) is the value of ith feature (given that it is good).

Customers evaluate features in a sequence, possibly going from a simple feature towards a more complicated one. There can be precedence relations among features forcing the customer to evaluate one feature before another. In the ultimate case, these relations will imply a complete order. We depicted three cases of product features in Figure 1. The evaluation sequence of features is irrelevant for our analysis of version strategy. For the phase strategy, we assume there is a sequence determined a priori to our analysis, and number the features according to this sequence.

Customers evaluate features one by one to build C_t, the set of features already evaluated by time t, so $C_0 = \emptyset$. Since we assume that the acquired knowledge is not forgotten, C_t never contracts, i.e. $C_s \subseteq C_t$ for $s \leq t$. $C_t \cap S$ is the set of good features that are evaluated by t. At time t, the (observed) value V_t of a product is:

$$V(t) = V(S \cap C_t) \text{ and } V_0 = 0. \tag{1}$$

At time t, C_t is observed so $C_t \cap S$ is not a random set anymore, but $\mathcal{U} - C_t$ is not observed and $((\mathcal{U} - C_t) \cap S)$ is a random set. $V(S)$ can be significantly or slightly larger than $V(t)$ depending on the success rate of the features in $\mathcal{U} - C_t$. Using (1) and the Positivity axiom,

$$V(s) = V(S \cap C_s) \leq V(S \cap C_t) = V(t) \text{ for } s \leq t \text{ and } V(t) \overset{t\to\infty}{\longrightarrow} V(S) \tag{2}$$

The process $V(t)$ is nondecreasing in a sample path sense. For the case of version strategy with N features, the product value can be obtained by setting $t \leftarrow \infty$ and $S \leftarrow S \cap \{1, \ldots, N\}$. From a customer's point of view, the perceived value of the product is nondecreasing in time or the extent of the demo version so this explains the vendor's motivation to extend the demo length and expand the demo version.

2.1 Demonstration version strategy

In the demo version strategy, vendors hand out a demo product with limited features to customers who base buy-or-not decision of the full product on the evaluation of the demo product. For example, Amazon makes several chapters of books available for promotion. Newspapers use this strategy by giving free access to their web sites where only certain news articles are featured to promote the paper copy. It is possible to listen to the first 30 sec. of songs on the Internet without buying them. In these examples there is not much to learn about features, rather customers glance at demo versions to decide if they like it. Thus, the evaluation resembles testing rather than learning.

Remember that N is the number of features available in a demo version. Unless evaluating features cost substantially with respect to the price of the software, customers will evaluate all the available features before committing to buy a product. Then the set of features evaluated just before buy-or-not decision is $\{1, \ldots, N\}$. The value of these features is V_N

$$V_N = \{\sum_{i \in C} Y_i \text{ with probability } X_N(p) = |C| \text{ for all } C \subseteq \{1, \ldots, N\}\} \tag{3}$$

where $X_N(p)$ denotes a binomial random variable with N trials and p probability of success.

In general, a vendor supplies many customers. Depending on potential usage, current needs, different customers may attach slightly different values to the same feature. Although we may argue for deterministic feature values for a single customer, those values become stochastic for a randomly chosen customer from a population. From vendor's point of view, feature values are then stochastic variables. If customers divide product features into similar groups in terms of hierarchy/importance, features are likely to have similar values (in a stochastic sense). For example, a customer can give edit and view features of Microsoft Word equal value and treat these as features. It is also possible to go down one step below in command hierarchy and treat all subcommands of edit and view as features with equal value. If feature values are not similar, features can be divided into smaller pieces or united together to have more or less identical feature values. Motivated by these observations, we use identically and independently distributed (iid) random variables to represent feature values. This allows us to express the product value without studying subsets of $\{1, \ldots, N\}$ as in (3) and yields

$$V_N = \sum_{i=1}^{N} \bar{Y}_i = \sum_{i=1}^{X_N(p)} Y_i \tag{4}$$

We scale the feature values by $E(Y_i)$ so that new feature values have expected value of 1 and variance of σ^2.

We can specialize V_N for two practical cases

Case 1. Many features are tested : When many features are included in the demo version, i.e., N is large.

Case 2. Deterministic values: If vendors know the value customers attach to each feature with certainty, then $Y_i = 1$ for all i after scaling values by their expected value.

Let V_N^1 and V_N^2 denote the value of the product under Cases 1 and 2:

$$V_N^1 = \mathcal{N}(Np, N(p\ \sigma^2 + p - p^2)) \quad \text{and} \quad V_N^2 = X_N(p) \tag{5}$$

where $\mathcal{N}$ denotes a Normal random variable. Although V_N^1 converges to the Normal distribution, we write it as equal for notational simplicity. V_N^1 is obtained by applying the central limit theorem to the random sum of (4). Since knowing feature values reduces uncertainty, V_N^2 is less variable than V_N^1. We will use the valuations of (5) to compare version and phase strategies.

2.2 Demonstration phase strategy

Some vendors choose to promote their products by allowing customers to evaluate full versions of the products during a demo phase. Demo phase lengths are 14 days for Encyclopedia Britannica, 30 days for EditPlus editor, 6-12 months for the computer software that comes with college textbooks and 1000 hours for AOL membership. These products are more complicated than those of demo version and they must be learned. Actually this strategy seems to be more commonplace than the version strategy in practice.

Different customers tend to have have different learning times for the same feature, because of their (educational / professional) backgrounds and previous exposure to similar features. As we argued for stochastic feature values before, we can argue for stochastic learning times. The traditional way of finding out learning times and feature values is via marketing surveys or experiments. In addition to these, cookies can be used to record and send learning times back to vendors for software products. As long as customers are informed about these cookies at the beginning of demo phase and they agree to proceed, there will be no privacy issues. Thus, learning times for software can be found out very accurately. We suppose that vendors collect learning time and feature value information from customers and make statistical analysis to obtain necessary distributions.

Since U is very large with respect to the number of features that can be learned during a demo and it is unlikely that a customer learns more than one feature in a small time interval, we assume that the number of features learned by time t is Poisson with rate λt, denoted by $Po(\lambda t)$. Then each customer learns a feature in time exponentially distributed with rate λ and the time at which learning of nth feature is completed becomes Gamma random variable with parameters n and λ, denoted by $\Gamma(n, \lambda)$. From (4), the value of the product at time t,

$$V(t) = V_{Po(\lambda t)} = \sum_{i=1}^{Po(\lambda t)} \bar{Y}_i = \sum_{i=1}^{X_{Po(\lambda t)}(p)} Y_i = \sum_{i=1}^{Po(p\lambda t)} Y_i, \tag{6}$$

where $X_{Po(\lambda t)}(p) = Po(p\lambda t)$ because of the thinning of a Poisson process with success probability p. We further assume that Y_i are independent of learning times. Then the product value representation in (6) has compound Poisson distribution.

Let τ_c be the time the product value crosses c (≥ 0), it can be defined by

$$V(t) \geq c \Leftrightarrow \tau_c \leq t.$$

We coin the term justification time (of value c) for τ_c. If feature values are deterministic then $\tau_c = \Gamma(c, \lambda p)$.

We assume that all features are learned with the same rate λ. Once features are divided and united to have identical feature values (as in the last subsection), no more degrees of freedom is left to achieve identical learning times. However, if values are proportional to learning times (the value of a feature learned quickly worths to customer less than than other features and vice versa), then identical learning times assumption will be satisfied automatically. Without this assumption, learning time of each feature must be specified separately. Learning time data requirements for such a model will be enormous. We believe that identical learning times assumption allows us to understand fundamentals of the phase strategy without leading to an overly complex model.

We can now examine $V(t)$ under two special cases introduced earlier.

Case 1: Many features are learned: $Po(p\lambda T) \to \infty$. In this case, the number of features learned in $[0, T]$ is large so we can use a diffusion approximation [24]:

$$V^1(T) = \lambda T E(\bar{Y}_i) + \sqrt{\lambda T E(\bar{Y}_i^2)} \cdot z = \lambda T p + \sqrt{\lambda T p(\sigma^2 + 1)} \cdot z$$

where z is the standard Normal variable $\mathcal{N}(0, 1)$. Note that $E(V^1(T)) = p\lambda T$ and $Var(V^1(T)) = \lambda T p(\sigma^2 + 1)$.

Case 2: Deterministic values: $Y_i = 1$ for all features. Since we normalize all values by $E(Y_i)$, $V^2(t) = Po(p\lambda t)$.

We borrow the definitions below from reliability theory, these definitions will later be used to establish structural properties of vendor's profit. A random variable with cdf F will be called increasing (decreasing) failure rate if $1 - F$ is logconcave (logconvex). It will be called decreasing inverse failure rate if F is logconcave. We use acronyms IFR and DFR for increasing (decreasing) failure rate and DIFR for decreasing inverse failure rate. IFR property for Normal is established in Lemma 3 using logconcavity of Normal density. In the case of $V^2(t)$ we have Poisson density which is IFR by Lemma 4, thus we obtain the following corollary.

Corollary 1 *Both $V^1(t)$ and $V^2(t)$ are IFR.*

We now compare the product value under demo version and phase strategies. It is natural to investigate if the affect of increasing the demo time T is equivalent to that of increasing the number of demo features N; can we choose parameters of demo version and phase strategies so that product values are the same at the end of demo? Consider a given product under Case 2 and note its value with two strategies: $V_N^2 = X_N(p)$ and $V^2(t) = Po(p\lambda T)$. When p is small and N is large, we can approximate a Binomial distribution by a Poisson distribution:

$$V^2(T) = Po(p\lambda T) \longrightarrow X_N(p) = V_N^2 \quad \text{if} \quad N = \lambda T \;\text{ and }\; N \longrightarrow \infty \tag{7}$$

In other words, whether demo version or phase strategy is used, the customer will obtain (stochastically) the same value for the product as long as $N = \lambda T$. This result can be generalized for the case of identically and independently distributed feature values, proof is in the Appendix.

Theorem 1 *Marketing strategies of demo phase of* $[0, T]$ *and demo version of* N *features yield approximately the same random product value at the end of the demo if* $N = \lambda T$ *and* N *is sufficiently large.*

Theorem 1 applies to product values only, it has no relation to profits. We will later revisit these strategies.

Now consider customer and vendor decisions together when the vendor offers two strategies simultaneously and forces the customer to choose one (perhaps by installing cookies on customer's computer to prevent multiple trial downloads of the same product). A customer with learning rate λ_c would like to evaluate as many features as possible during the demo which can be achieved by choosing the demo version (phase) option if $\lambda_c < N/T$ ($\lambda_c > N/T$). Actually λ is the average of all λ_c over all customers, call customers with $\lambda_c < \lambda$ as slow learners and the rest as fast learners. Since all the fast learners would choose the demo phase option, they on the average will learn more than N features in $[0, T]$. Then for equivalence of simultaneously offered strategies we need $T < N/\lambda$. In any case, offering two demo strategies simultaneously, vendors can achieve self-customization (by customers) of demo according to their learning speed. Currently, no vendor does this to the best of our knowledge.

Another model of product promotion is combining version and phase strategies. In this case, vendors provide a limited version of the product with N features for a limited time T. The demo ends at time T or at the completion of learning Nth feature, whichever is earlier. For example, Game Commander, which sells voice controlled interfaces for computer games, offers limited (in terms of actions per command, commands per game and concurrent operation with chat) versions for 30 days. Emerald Publishing of UK gives free access to only 3 of its publication databases for a month. The value of the product $V_N(T)$ under the combined strategy can be written as :

$$V_N(T) = \sum_{i=1}^{N \wedge Po(\lambda T)} \bar{Y}_i.$$

where $\wedge$ denotes minimum. Having developed the product value, we will relate the value to buying and will study vendor's profit in the next section.

3 Vendor's Profit

In this section we will investigate how demo time and demo features affect vendors' profits. Since software production costs are negligible, we will take revenue as a measure of profit. Instead of the total revenue over all customers, we will study the revenue per average customer. Effectively this is nothing but scaling the revenue by the total number of customers. Revenue per customer relates to the sales price and the probability of sales at the end of the demo.

To understand the probability of sales, the thought process of customers leading to buying must be examined. Assuming that customers are behaving objectively, they would buy a product if its value is more than its price. Customers actually use demo phase or version to estimate the product value. If the product value at the end of demo is larger than a threshold value c (called purchase justification value from now on), purchase happens. This process is also discussed qualitatively in [10] which says: "The [product evaluation] ... process concludes with the decision whether to buy the product after the demo." Clearly c and the product price are positively related but they do not have to be equal. Actually, customers do not usually expect to justify the whole price of the product during the demo so c is often smaller than the price.

Although customers form a lot of perceptions during the demo, they still have incomplete information about the product because they cannot finish evaluating all product features. Since some of these unevaluated features will be useful to customers, we are motivated by [15] to assume that customers discount the price of products to get c and compare c to the value of evaluated features. A hardly-pleased (easily-pleased) customer wants most (a small portion) of the price be justified during the demo, that customer will set c at a large (small) value with respect to the price. Then, the revenue made from each customer can be written as (discounting constant $\times$ c) $\times$ P(Sale happens). If all customers have the same discounting constant, that constant is irrelevant for optimizing the total profit from the entire customer population because total profits are proportional to $c \times P$(Sale happens). To keep our models of reasonable size, we assume that all the customers use the same constant. If this fails, some of customers are pleased hardly and some easily, so demo must be customized by optimizing demo parameters N, T and c separately for hardly and easily pleased customers. Implementation of different demo parameters for different types of customers is a challenge, however the Internet can overcome this and allow for extensive customization of demo. Once such customization is done, our assumption (of same discounting constant for all customers) is satisfied automatically. Since Y_is are iid, it will be convenient to consider c as multiples of $E(Y_i)$, i.e. $c \leftarrow c/E(Y_i)$, then c will be unitless. In our models, customers will consider buying the product if $V_N \geq c$ and $V(T) \geq c$.

3.1 Profits under version strategy

Customers consider buying the product if the value of the demo version is greater than the purchase justification value c. However, if customers can satisfy their daily needs with the demo version, they will not buy the product. Consequently, $V_N \geq c$ is not a sufficient condition for buying. Thus, the number of features in the demo version must be large enough to raise the appetite of customers but it must be small enough so that customers really need the full version. Let the random variable R be the maximum value of the demo product that does not meet average customer needs, for convenience assume that R is scaled, i.e. $R \leftarrow R/E(Y_i)$. We study the profit under Cases 1 and 2:

$$\Pi^1(c, N) = cP(c \leq V_N^1 \leq R) = c \int_{x=c}^{\infty} P(R \geq x) f_V(x) dx,$$

$$\Pi^2(c, N) = cP(c \leq V_N^2 \leq R) = c \sum_{i=c}^{N} P(R \geq i) P(X_N(p) = i), \tag{8}$$

where f_V is the probability density function for $\mathcal{N}(Np, N(p\ \sigma^2 + p - p^2))$. When the customer need can be estimated very well and does not vary much among customers, it can be taken as a deterministic number. Then we write $R = r$

(with probability 1). We can specialize the revenue for deterministic r:

$$\Pi^1(c,N) = c\int_{x=c}^{r} f_V(x)dx \quad \text{and} \quad \Pi^2(c,N) = c\sum_{i=c}^{N\wedge r} P(X_N(p)=i). \tag{9}$$

We are interested in maximizing profits by choosing strategy parameters c and N. The next lemma ensures that profit functions behave well for optimization purposes.

Lemma 1 *a) Profit $\Pi^1(c,N)$ is unimodal in justification value c if customer needs R is DFR or deterministic.*
b) $\Pi^1(c,N)$ is unimodal in the number of features N if R is deterministic and coefficient of variation of feature values is less than or equal to $\sqrt{1+pN}$.
c) $\Pi^2(c,N)$ is unimodal in c if R is DFR or deterministic.
d) $\Pi^2(c,N)$ is unimodal in N if R is deterministic.

Proof: a) The proof below is only for random R, similar steps yield the proof for deterministic $R=r$. First consider the derivative:

$$\begin{aligned}\frac{d\Pi^1(c,N)}{dc} &= \int_{x=c}^{\infty} P(R\ge x)f_V(x)dx - cP(R\ge c)f_V(c) = \left(\int_{x=c}^{\infty}\frac{P(R\ge x)f_V(x)}{P(R\ge c)f_V(c)}dx - c\right)P(R\ge c)f_V(c)\\ &= \left(\int_{x=0}^{\infty}\frac{P(R\ge x+c)}{P(R\ge c)}\frac{f_V(x+c)}{f_V(c)}dx - c\right)P(R\ge c)f(c)\end{aligned} \tag{10}$$

It suffices to show that the integral inside the first parenthesis decreases with c. Note that

$$\frac{f_V(x+c)}{f_V(c)} = \exp\left(\frac{-x^2-2x(c-Np)}{2N(p\sigma^2+p-p^2)}\right)\downarrow \text{ in } c. \tag{11}$$

Let f_R and F_R be pdf and cdf of R, then for an arbitrarily fixed $x\ge 0$

$$\frac{d}{dc}\frac{P(R\ge x+c)}{P(R\ge c)} = \frac{-f_R(x+c)(1-F_R(c))+f_R(c)(1-F_R(x+c))}{(1-F_R(c))^2} = \frac{1-F_R(x+c)}{1-F_R(c)}\left(-\frac{f_R(x+c)}{1-F_R(x+c)}+\frac{f_R(c)}{1-F_R(c)}\right)\le 0. \tag{12}$$

where the inequality follows from the DFR property for R. Thus, $P(R\ge i+c)/P(R\ge c)$ is decreasing in c. Combining this with (11), we show that the integral inside the parenthesis in (10) decreases in c. Therefore the derivative of Π^1 can have only one sign change which implies unimodality.

b) Lemma 6 of the Appendix says that if $X(N)\sim\mathcal{N}(N,N\bar{\sigma}^2)$ for some $\bar{\sigma}$ independent of N then $P(c\le X(N)\le r)$ decreases in N as long as $N>(c+r)/2$. Since $P(c\le V_N^1\le r)=P(c/p\le X(N)\le r/p)$ for $X(N)\sim\mathcal{N}(N,N(\sigma^2/p+1/p-1))$, $P(c\le V_N^1\le r)$ is decreasing in N if $N>(c+r)/(2p)$. Thus, we are not concerned with $N>(r+c)/(2p)$. Using the standard normal variable z,

$$\Pi^1(c,N) = \int_{\frac{c-Np}{\sqrt{N}\tilde{\sigma}}}^{\frac{r-Np}{\sqrt{N}\tilde{\sigma}}} f_z(z)dz$$

where $\tilde{\sigma}^2 = p\ \sigma^2+p-p^2$. Upon differentiation this yields

$$\frac{d\Pi^1(c,N)}{dN} = \frac{1}{2\tilde{\sigma}N}\left(-(r/\sqrt{N}+p\sqrt{N})f_z(\frac{r-Np}{\sqrt{N}\tilde{\sigma}})+(c/\sqrt{N}+p\sqrt{N})f_z(\frac{c-Np}{\sqrt{N}\tilde{\sigma}})\right).$$

Setting this equal to zero, we obtain a necessary condition for N:

$$1+\frac{r-c}{c+pN}=\exp\left(\frac{r^2-c^2}{2N\tilde{\sigma}^2}+\frac{p(c-r)}{\tilde{\sigma}^2}\right)$$

To conclude that there can be at most one N satisfying the necessary condition, we will show that the left hand side derivative is larger than the right hand side derivative for all $N \le (r+c)/(2p)$:

$$-\frac{p(r-c)}{(c+pN)^2}\ge -\frac{r^2-c^2}{2N^2\tilde{\sigma}^2}\exp\left(\frac{r^2-c^2}{2N\tilde{\sigma}^2}+\frac{p(c-r)}{\tilde{\sigma}^2}\right)$$

Since $N \le (r+c)/(2p)$, the term in the exponent is nonnegative. Then it is sufficient to show that:

$$\frac{1}{(c+pN)^2}\le\frac{1}{N\tilde{\sigma}^2}$$

This holds if $N\tilde{\sigma}^2 = Np(\sigma^2+1-p) \le c^2+2pN+p^2N^2$ or if $\sigma^2+1-p \le 2+pN$. Therefore $\sigma^2 \le 1+pN$ is a sufficient condition for unimodality. Proof of c) is similar to b), proof of c) and d) are deferred to the Appendix. □

The condition on coefficient of variation in Lemma 1.b is not strong. For example, when modelling feature values with a normal distribution, coefficient of variation must be taken much smaller than 1 to ensure nonnegativity of feature values. Moreover, Np will be a large number for any practical values of N and p.

Lemma 1 can be used in a reliability context. Consider a product with N parallel units such that after the failure of a unit, an unused unit starts operation. The product operates until the last units fails. Suppose that each unit has independent lifetime, Normally distributed with sufficiently small coefficient of variation. Lemma 1.b implies that there is a single number N maximizing the probability of product lasting at most r and at least c. On the other hand, now consider designing a product with N parallel units so that we maximize the probability of having at least c and at most r operational units, where each unit may not function independently with probability p. Lemma 1.d implies that there is a single number N maximizing this probability. In both cases, we are manipulating the parameter N of the product value distribution to trap it into a given interval that triggers a purchase. Thus choosing N is equivalent to giving the customers impression that the product worths its price and it is useful for the needs. This is the use of demonstration to promote products.

To motivate our understanding of optimal justification price c and the optimal number of features in the demo version, we examine necessary conditions on c and N for maximizing $\Pi^2(c,N)$. These conditions are obtained in the proof of Lemma 1.c and d. First the necessary condition on c:

$$cP(X_N(p)=c)P(R\ge c)=\sum_{i=c}^{N}P(X_N(p)=i)P(R\ge i)$$

Since c is discrete, equality may not be achieved exactly, so our necessary conditions are approximate. These conditions can be made exact by defining first order differences as c or N increase and decrease, but this complicates exposition without improving understanding which is our main goal. When justification price c increases by 1 purchase probability and profit decrease by $P(X_N(p)=c)P(R\ge c)$ and $cP(X_N(p)=c)P(R\ge c)$, this is the marginal cost of increasing c. However, 1 unit more revenue is made from customers who still buy the product, then marginal revenue is $\sum_{i=c}^{N} P(X_N(p) = i)P(R \geq i)$. As a result, optimal c sets marginal revenue equal to marginal cost.

The necessary condition on N can also be motivated in the special case of deterministic customer need r, we have

$$P(X_N(p) = c - 1) = P(X_N(p) = r).$$

Suppose that we multiply both sides with the probability p of $N+1$st feature being good. The left hand side becomes the probability that the product could not be justified with N features but can be justified with $N+1$ features. Then, the right hand side is the probability that the product does not meet customer needs with N features but will meet them with $N+1$ features. In addition to p, if we multiply left and right hand sides with justification value c, the left (right) hand side becomes the marginal revenue (cost) of increasing N. Thus, necessary condition on N also set marginal costs equal to marginal revenues.

When $N = |U|$, demo and full versions are the same. Then, no product can be sold and revenue is zero, so we must have $N < |U|$. In view of (8) and $N < |U|$, the total number of product features has no effect on the optimal number of features included in the demo version as long as the justification price is fixed. The justification price may go up and down with the price of the product rather than the total number of features in the product because customers do not use many features of the product anyway. Some vendors may prepare a demo version by including roughly a certain percentage (say 20-25%) of all features, this practice can result in suboptimal profits.

When the sequence for evaluating features is not known but N is and features are of different value, vendors must choose to include valuable features in the demo version. When there is a strict order for checking features (chain structure) vendors must include the first N features in the demo version. When feature values satisfy additivity axiom and features can be checked independently, the greedy approach works; order features from high to low values and include first N features in the demo version. However, features in general are neither totally orderable nor totally independent. There is a precedence relationship among features, feature i will precede feature j if i can be checked only after j. In that case, we would have a precedence network to represent the precedence structure. For deterministic feature values, the formulation of this problem is a special case of knapsack problem on a noncyclical network. When there are cross value terms like $v_{i,j}$ (violating additivity axiom) and there are no precedence structure, choice of N features reduces to a quadratic knapsack problem [5].

3.2 Profits under phase strategy

Let the net present value of the expected profit from an average customer be $\Pi(c, T)$ and for now let α be the discounting rate compounded continuously. We express profits as functions of price c and demo time T.

$$\Pi(c, T) = c \exp(-\alpha T) P(V(T) \geq c) = c \exp(-\alpha T) P(\tau_c \leq T). \tag{13}$$

These profit expressions are written under the assumption that customers wait until the end of the demo phase to decide on buy-or-not. Customers may want to delay this decision as much as possible to check more features and to delay payments. If customers decide prematurely to buy ad pay as soon as $V(t)$ touches c, then

$$\Pi(c, T) = c \int_0^T \exp(-\alpha t) P(V(t) \geq c, V(s) < c \text{ for } s \in (0, t)) dt.$$

This case is not interesting because the integrand is always positive so the revenue is strictly increasing in demo length. Then vendors must let demo continue until the arrival of the next version. Since this is not the case in practice, we believe that most of the software buyers decide on buy-or-not at the end of the demo phase so we will use (13) as the profit expression.

There is another interpretation of (13) when customers evaluate several other products simultaneously along with the product we are studying. A customer buys the product if $V_T \geq c$ and the value of all the other products remain below their prices from 0 to T. In other words, the time each one of other products' value surpasses its justification value is beyond T. If this time is exponentially distributed with α_i for product i, the earliest purchase justification time for other products is exponential with rate $\sum_i \alpha_i$. Thus, the probability of not switching to another product in $[0, T]$ is equivalent to the earliest time being larger than T, i.e., $exp(-\sum_i \alpha_i)$. Then we can set $\alpha = \sum_i \alpha_i$ and add the discount rate.

The discount rate of (13) can be far greater than the interest rate. This can be justified with short software product life cycles. A particular version of a software loses a big portion of its value once a new version is introduced. Actually older versions are often given away for free. Since new versions come about every one or two years, the discount rate must be steep enough to substantially bring down the product value over two years.

Our aim is to find optimal $c, T \geq 0$ pairs to maximize the revenue $\Pi(c, T)$. Demo phase for a product can last only until the introduction of the next version of the same product. One can place an upper bound on T without altering our analysis and conclusions.

Lemma 2 *a) If the value $V(T)$ of product at the end of demo phase is IFR then $\Pi(c, T)$ is unimodal in c.*
b) If the justification time τ_c is DIFR then $\Pi(c, T)$ is unimodal in T.

Proof: a) Fix time T. Suppose that $V(T)$ is a continuous random variable with pdf f_T, cdf F_T and $\bar{F}_T = 1 - F_T$. Note $d\Pi/dc = \exp(-\alpha T)\{\bar{F}_T(c) - cf_T(c))\} = \exp(-\alpha T)c\bar{F}_T(c)\{1/c - f_T(c)/\bar{F}_T(c)\}$ so the critical c is found from the necessary equation $1/c - f_T(c)/\bar{F}_T(c) = 0$. For $c = 0$, $1/c - f_T(c)/\bar{F}_T(c) = \infty$. But it decreases as c increases because $V(T)$ is IFR, i.e., $-f_T(c)/\bar{F}_T(c)$ decreases in c. Then there is a unique c solving the necessary condition, equivalently profit is a unimodal function. The derivative is positive before the unique c and negative afterwards so the unique c maximizes the profit. When $V(T)$ is a discrete random variable, above steps can be traced by replacing $f_T(c)$ with $P(V(T) = c)$ to conclude that the profit is still unimodal. However, there could be two prices c and $c + 1$ maximizing the profit.
b) Fix price c. Let F_c and f_c be cdf and pdf of the justification time τ_c. Consider the necessary condition: $d\Pi/dT = -c\alpha \exp(-\alpha T)F_c(T) + c\exp(-\alpha T)f_c(t)$. T uniquely solves $\alpha = f_c(T)/F_c(T)$ because τ_c is DIFR. Furthermore, before unique T the derivative is positive and afterwards it is negative, so the unique T maximizes the profit which is a unimodal function in T. $\square$

IFR and DIFR conditions of Lemma 2 can be combined into a single condition: $P(V(t) \geq c)$ is logconcave in t and c. Note that the condition is only along t and c axes so we do not require joint logconcavity.

Lemma 2 implies that the profit is maximized by the unique demo time T and price c which solve the necessary conditions

$$(1/c) = \frac{dP(V(T) \leq c)}{P(V(T) \geq c)} \qquad \alpha = \frac{dP(\tau_c \leq T)}{P(\tau_c \leq T)} \tag{14}$$

When $V(T)$ is a continuous random variable $dP(V(T) \leq c)$ must be interpreted as the density function of $V(T)$ at c. Otherwise, we take $dP(V(T) \leq c) = P(V(T) = c)$ for discrete random variables. Necessary conditions indicate that the optimal justification value depends heavily on feature values whereas optimal phase length is very sensitive to competition and discount rate. We will later revisit this observation with numerical examples.

Let us write the first condition as

$$P(V(T) \geq c) = c\, dP(V(T) \leq c).$$

This condition simply asserts the equality of marginal revenue and the marginal cost at the optimal price; The left hand side of the condition is the extra revenue generated by increasing price from c by to $c+1$, if the customer is willing to buy at the price of $c+1$. The right hand side of the condition is the revenue lost from customers who would buy at price c but not at price $c+1$. Now multiply both sides of second condition by c:

$$c\, dP(\tau_c \leq T) = c\alpha P(\tau_c \leq T).$$

The left hand side is the extra revenue made from customers who justify a purchase not by T but immediately after T. The right hand side is the amount of revenue discounted by expanding demo length. Note that $\exp(-\alpha T) = 1 - \alpha T$ for small αT so that revenue is discounted approximately at a rate of α. Once more necessary conditions state the equivalence of marginal revenue and marginal cost.

With Corollary 1, we have already discussed that IFR property holds for $V^1(T)$ and $V^2(T)$. When feature values are deterministic, justification times become Gamma random variables and they are DIFR by Lemma 5. When we use the Normal approximation, justification times are DIFR by Lemma 3. As a result, assumptions of Lemma 2 are not very restrictive. The next corollary follows from applying these observations to Lemma 2. A full characterization of IFR for total feature values and DIFR for justification times is beyond the scope of this paper, however related results can be found in [13]. DIFR property is common among nonnegative random variables (e.g. Gamma and Poisson see Lemmas 4 and 5), also see [6].

Theorem 2 *When feature values are deterministic or many features are learned during the demo phase (allowing Normal approximation), profits are unimodal in justification value and demo length.*

As a result of Theorem 2, either for a given justification price or a given demo length we can obtain unique demo length or price, respectively, maximizing the profit using necessary conditions of (14). However, the corollary does not assert the existence of a unique price and duration pair maximizing the profit together. Since the profit is not concave even in price or duration separately, the hope to obtain uniqueness in price and duration together is limited.

Specializing necessary conditions for deterministic feature values, we can obtain close form solutions for demo length:

$$c = \frac{P(Po(p\lambda T) \geq c)}{P(Po(p\lambda T) = c)} = \frac{F_{\Gamma(c,p\lambda)}(T)}{f_{\Gamma(c,p\lambda)}(T)} \frac{f_{\Gamma(c,p\lambda)}(T)}{P(Po(p\lambda T) = c)} = \frac{1}{\alpha} \frac{f_{\Gamma(c,p\lambda)}(T)}{P(Po(p\lambda T) = c)}$$

The last term above is the distribution of $\Gamma(c, p\lambda)|Po(p\lambda T) = c$, it is well known that this random variable is distributed as the maximum of c uniform random variables over $(0, T)$. Note that this distribution is independent of the rate of the Poisson. Then,

$$\frac{f_{\Gamma(c,p\lambda)}(T)}{P(Po(p\lambda T) = c)} = \frac{c}{T}$$

which implies $T = 1/\alpha$. This yields the next result:

Corollary 2 $\Pi^2(c, T)$ *is maximized at* $T = 1/\alpha$ *thus finding the optimal profit reduces to maximization only along the justification price.*

Interestingly, optimal demo length does not depend on the learning rate λ, it is just a function of α representing the discount factor and competition. Then the optimal demo length can be chosen without knowing whether customers learn fast or slow (different customers can learn at different rates), or the learning rate, as long as customers learn according to a Poisson process. On the other hand, price increases with learning rate λ and decreases with α: fast learners can justify higher prices and higher competition draws prices down. We make similar observations also for $\Pi^1(c, T)$ with our numerical examples.

Profits of version and phase strategies should not be compared because they are applicable under different situations. Version (phase) strategy focuses on customer needs R (discounting and competition rate α). To understand this distinction consider two vendors that sell editor software and academic paper archives. There is more competition among editor vendors than academic paper archive providers. This is because academic papers are not substitutable for each other. If customer needs are overestimated with version strategy, few customers will make a purchase because demo version will be sufficient for daily needs. Then an archive provider should promote its product by focusing on customer needs rather than competition and discount rate. The source of strong competition in editor market can be another editor vendor or a newer version of the same editor scheduled to be released in the future. That is why editor vendors are more time sensitive in promoting their products. Consequently, we suggest demo (phase) strategy for archive (editor) vendors.

Version and phase strategy can be combined by giving a demo version to customers for a limited time. This combined strategy can be compared with the phase strategy because the product will expire in both strategies. The profit $\Pi(c, N, T)$ associated with a combined strategy can be expressed as

$$\Pi(c, N, T) = c \exp(-\alpha T) P \left(\sum_{i=1}^{N \wedge Po(\lambda T)} \bar{Y}_i \geq c \right).$$

When writing this expression we assume that customers decide on buy-or-not at T. This would be the case for customers who are looking for a product to improve on what is currently used, such relaxed customers can manage their operations with the current product for a while. If $\Gamma(N, \lambda) \geq T$, $\Pi(c, N, T) = \Pi(c, T)$, otherwise $\Pi(c, N, T) \leq \Pi(c, T)$, that is in terms of profits made from relaxed customers phase strategy dominates the combined strategy.

Now consider urgent customers who need the product immediately. These customers decide to buy as soon as the purchase is justified. For $\Gamma(N, \lambda) \geq T$ $\Pi(c, N, T) = \Pi(c, T)$, thus consider only $\Gamma(N, \lambda) < T$. Depending on where the justification time falls we have three possible situations:

1. $\tau_c \leq \Gamma(N, \lambda) \leq T$: With the phase and combined strategy customers buy at τ_c.
2. $\Gamma(N, \lambda) \leq \tau_c \leq T$: With the phase strategy customers buy at τ_c. With the combined strategy, they decide not to buy at $\Gamma(N, \lambda)$.
3. $\Gamma(N, \lambda) \leq T \leq \tau_c$: With the phase and combined strategy customer decide not to buy at T and $\Gamma(N, \lambda)$ respectively.

In Situation 1, customers buy with both phase and combined strategy at the same time. In Situation 2, customers buy under the phase strategy but not under the combined strategy. As a result the phase strategy provides higher profits than the combined strategy in the case of urgent customers. We suggest that the phase strategy is used always instead of the combined strategy.

4 Numerical Experiments

In this section, we consider the version and phase strategy where many features are learned during the demo so we study $\Pi^1(c, N)$ and $\Pi^1(c, T)$. We concentrate on Case 1 because Case 2 has a partial closed form solution by Corollary 2. In addition to profit, purchase probabilities are also of interest. We choose our parameters so that the resulting purchase probability is about the AOL membership purchase probability of three quarters [9]. We do not have exact estimates of the parameters p, σ, distribution of R, α or λ so we vary these parameters over reasonable ranges as in Figures 2 and 3. Note that our parameter ranges always yield purchase probabilities between 60% and 85% so they can be argued to be approximating AOL's case.

For the version and phase strategies we set $p = 0.3$ and $\sigma = 0.2$ for our base case. Recall that p is the probability of a feature successfully matching a customer need, this probability is substantially smaller than the probability of a feature operating successfully so we take $p = 0.3$. In the version case, we define customer needs as a discrete random variable with mass at only two points. For the base case, we set $E(R) = 18$ and $var(R) = 4$. In the case of phase strategy, we take a month as a measure of 1 time unit and set $\alpha = 0.3$ and $\lambda = 10$. Since $exp(-0.3) = 0.75$, the expected profit from a sale reduces by 25% every month the sale is postponed. This drop is large because of two factors: new versions and competition. New versions of software are introduced every 1-2 years, this reduces the value of current versions drastically as time passes. Competition cuts down the probability of purchase as the demo process is prolonged, hence pulling expected profits further down. We justify seemingly small learning rate of $\lambda = 10$ per month by noting that a

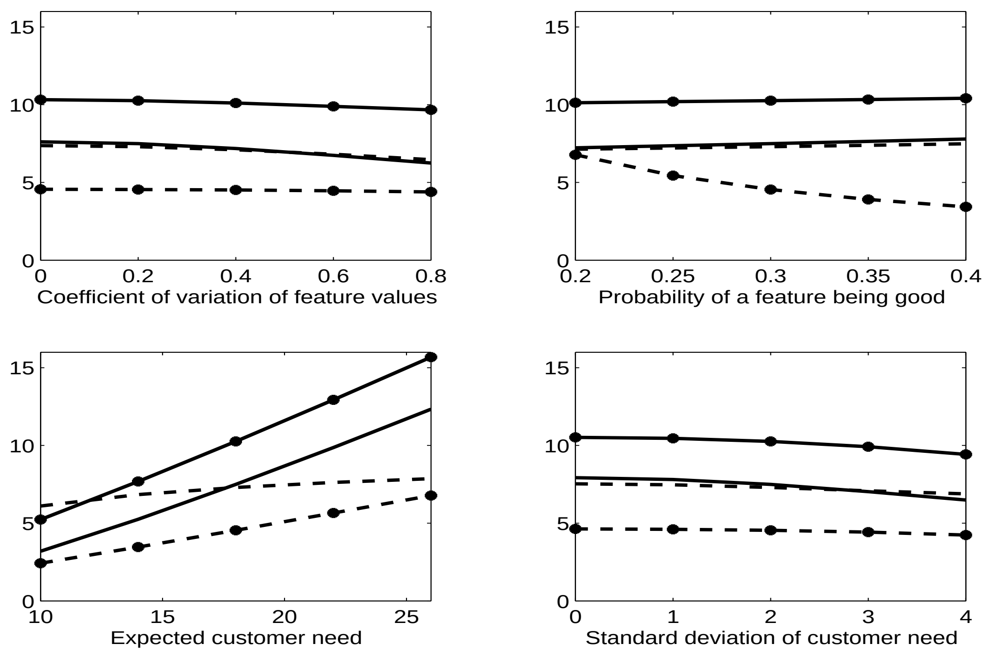

Figure 2: Version strategy performance measures and optimal phase parameters as inputs parameters vary. Profit (—), 10*Purchase Probability (--), Optimal Justification Value (•–•) and Optimal Number of Features/10 (•--•).

customer is likely to spend only a portion of working hours for learning the product. In the version and phase strategy experiments, we fix 3 parameters out of 4 $(p, \sigma, E(R), var(R))$ and $(p, \sigma, \lambda, \alpha)$ respectively and vary the fourth to obtain the graphs in Figures 2 and 3, i.e. every graph illustrates the effect of varying a single parameter in the base case.

We provide managerial insights for version strategy using Figure 2. Our insights for the coefficient of variation of feature values is common to both version and phase strategies, compare Figure 2 and 3.

Uncertainty of feature values: In neither version nor phase strategy, uncertainty in the feature values has a big effect although a close examination reveals that it brings profits and purchase probabilities down. Since we assume that feature values are independent, extreme values cancel each other when they are summed. Thus, the uncertainty of a feature value has a little effect on the overall profit figures. The practical implication of this observation is that vendors should not bother to accurately estimate the variance of feature values. The observation also supports the deterministic feature value assumption made for Case 2.

Probability of feature being good: At first thought, one would expect the profit to grow with software that meets more of customer needs, i.e., higher p. However, we must remember that if most of the customer needs are satisfied with the demo version, customers will not buy the full version. That is why the number of demo features reduce with p. This reaction keeps other three performance measures stable. As a result, vendors would undercut their own sales if they meet

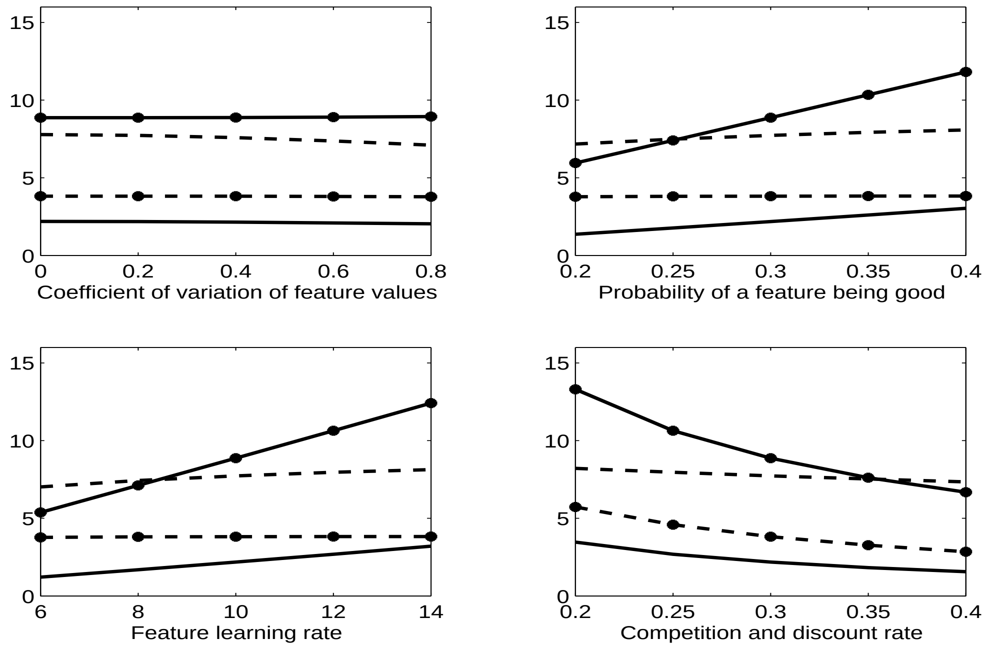

Figure 3: Phase strategy performance measures and optimal parameters as inputs parameters vary. Profit (—), 10*Purchase Probability (--), Optimal Justification Value (●–●) and Optimal Phase Length (●--●).

most of customer needs with demo products.

Expected value and standard deviation of customer needs: All the performance measures increase (decrease) with the expected value (standard deviation) of customer needs. These results are intuitive: more customer needs let vendors sell expensive products which are promoted with extensive demo versions. Consequently, vendors should come up with creative use of their products to push customer demand up. On the other hand, variability in customer needs makes demonstration harder and pulls down the profits. Customizing the demonstration process (different parameters for different set of customer classes) according to customer needs can reduce the standard deviations of the needs and boost vendor profits.

We provide managerial insights for phase strategy using Figure 3.

Probability of feature being good and feature learning rate: We treat these two cases together because we can write all performance measures in terms of λp by (6). This time, profits, prices, phase length (only slightly) and purchase probability go up with λp. A higher probability of good features indicates a product that is customer-oriented and almost bug-free. However, such a superior product would be hard and costly to develop so our observations are not practical for implementation. On the other hand, learning rate can be increased relatively easily by providing product manuals, one-step set up procedures, help, demo files, illustrative examples, product web sites, product chat and phone lines. Finally, between two uncertain model parameters feature values and learning rates, learning rates have a bigger impact on the performance measures. It worths to spend effort on data gathering and analysis to find out learning rates.

Competition and discount rate increases: In this case a vendor would pull down its phase length substantially and price somewhat (so the justification value c drops). Although this keeps purchase probability stable, profits drop almost about the same rate as prices. Then the best way to react to increased competition is by cutting down the window of vulnerability (demo phase) while damping the effect of shorter demo phase on sales by reducing prices.

Optimal justification values vary drastically in 4 experiments out of a total of 8 in Figures 2 and 3. Since the justification value is proportional to the product price, our experiments often advise reacting to parameter changes by modifying the product price. Unlike traditional product vendors, information product vendors have a lot of flexibility to adjust their prices. For example, information product vendors can reduce prices drastically and still stay profitable because their marginal production cost is almost zero. That is why our observations must be used with great care outside the domain of information products.

5 Conclusion

In this paper, we study demonstration strategies to promote information products. We coin terms "version", "phase" and "combined" for demonstration strategies that currently exist in practice exactly in the ways we have defined. Our contribution is modelling and analyzing these strategies and this is the first such effort as far as we know. We first study the product testing and learning mechanisms to estimate product values at the end of demos. Then we use product values to trade off key performance measures and strategy parameters: the product demand (represented by purchase probability), the product price (represented by justification value), the number of demo features and the length of demo phase. These trade offs lead to necessary optimality conditions which can also be motivated by setting marginal costs equal to marginal revenues. We also present several results guaranteeing the unimodality of our profit functions and hence the uniqueness of solutions to optimality equations.

We establish that product values at the end of version and phase strategies converge asymptotically if strategy parameters are chosen appropriately. However, we point out this is not to say that profits under these strategies will be equal for they apply in different contexts. A sale can be considered in two steps: first customer is convinced of the need for the product then the customer chooses the vendor we are studying. Our formulation of version strategy emphasizes the first step so it is appropriate when competition is limited. That is why special attention is then given to the customer needs. When there is competition, the vendor is under pressure to instigate the sale immediately so time sensitive phase strategy becomes more appropriate. We also mention that evaluation of products can be thought as testing and/or learning activity. When it is purely testing, phase strategy becomes irrelevant because testing often takes much shorter time than learning. Some vendors combine version and phase strategy, we argue that this strategy is beaten by the phase strategy irrespective of whether customers are "relaxed" or "urgent". Thus, we do not study the combined strategy in detail. We discuss the concept of self-customization (by customers) of demo strategy when version and phase options are offered simultaneously; fast (slow) learners are likely to choose phase (version) strategy. We also talk about customizing demonstration according to customer needs to reduce standard deviation of customer needs and to increase profits. We note that our performance measures are more sensitive to learning rates than feature values and advise that vendors concentrate their efforts on estimating learning rates. Finally, we point out that our models are more applicable when prices can be adjusted freely, i.e., for information products.

In order to obtain above results we have made several assumptions. First, we assume that product values are zero initially and are nondecreasing (2). If customers initially have positive expectations about product values, then the expected product values may decrease after each bad feature. This corresponds to discovering a shortcoming of the product and is discussed in [10]. All of our models can be generalized by allowing product value to drop. Second, we assume that feature values and learning times are identical and point out that either one of these assumptions can be made without loss of generality. Relaxing these assumptions can lead to a richer setting. Third, one can introduce the concept of "panicking customer" who would buy at the end of demo phase when the current product value is slightly below the justification value. A panicking customer does this to avoid search and evaluation costs for another product, if the current one is not purchased. Finally, our model of competition from other vendors is crude and can certainly be refined further. These ideas yield various venues for future research.

Acknowledgements: This research was supported by a summer research grant from the School of Management, University of Texas at Dallas. We thank to Shun-Chen Niu for his helpful comments and contributions to the proof of Lemma 1.d.

Appendix: Proofs

Lemma 3 *Normal is IFR and τ_c associated with $V^1(t)$ is DIFR.*
Proof: Suppose that $X \sim N(\mu, \sigma^2)$. Then for any c, we show the logconcavity of

$$P(X \geq c) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) I_{[x \geq \frac{c-\mu}{\sigma}]} dx.$$

where $I_{[x \geq \frac{c-\mu}{\sigma}]}$ is an indicator function. It suffices to show the logconcavity of the integrand in (c, x) by Theorem 5.1 of [8]. Integrand is logconcave if the Normal density and the indicator function are so. Clearly Normal density is logconcave in x. The indicator function is logconcave in (c, x), because for $0 \leq \alpha \leq 1$

$$I^{\alpha}_{[x_1 \geq \frac{c_1-\mu}{\sigma}]} I^{1-\alpha}_{[x_2 \geq \frac{c_2-\mu}{\sigma}]} = I_{[x_1 \geq \frac{c_1-\mu}{\sigma}]} I_{[x_2 \geq \frac{c_2-\mu}{\sigma}]} \leq I_{[\alpha x_1 + (1-\alpha) x_2 \geq \frac{\alpha c_1 + (1-\alpha) c_2 - \mu}{\sigma}]}.$$

Noting that multiplication of functions preserve logconcavity, we establish the first assertion.

Let $\tilde{c}(T) = (c - p\lambda T)/\sqrt{p\lambda T(\sigma^2+1)}$. We use algebra to show that $\tilde{c}(T)$ is a convex function in T:

$$\frac{d\tilde{c}(T)}{dT} = -\frac{p\lambda^2(\sigma^2+1)}{2\sqrt{t}(p\lambda(\sigma^2+1))^{1.5}} - \frac{cp\lambda(\sigma^2+1)}{2(p\lambda t(\sigma^2+1))^{1.5}} \quad \uparrow \text{ in } T$$

Since $\log P(z \geq c)$ is decreasing and concave in c (from part (a)) and $\tilde{c}(T)$ is convex, $\log P(z \geq \tilde{c}(T))$ is concave in T. Then $\log P(\tau_c \leq T) = \log P(z \geq \tilde{c}(T))$ implies that $\log P(\tau_c \leq T)$ is concave in T, i.e., τ_c is DIFR. □

Lemma 4 *Poisson is IFR and DIFR.*
Proof: First consider the reciprocal of the failure rate:

$$\frac{P(Po(\lambda) \geq c)}{P(Po(\lambda pT) = c)} = \frac{c!}{(\lambda)^c e^{-\lambda}} \sum_{i \geq c} \frac{(\lambda)^i e^{-\lambda}}{i!} = \sum_{j \geq 0} (\lambda)^j \frac{c!}{(j+1+c)!}$$

The last term decreases in c so Poisson is IFR. Now consider the reciprocal of the inverse failure rate:

$$\frac{P(Po(\lambda) \leq c)}{P(Po(\lambda pT) = c)} = \frac{c!}{(\lambda)^c e^{-\lambda}} \sum_{i=0}^{c} \frac{(\lambda)^i e^{-\lambda}}{i!} = \sum_{j=0}^{c} (\lambda)^{-j} \frac{c!}{(c-j)!}$$

The last term increases in c so Poisson is DIFR. □

Lemma 5 *Gamma is DIFR.*
Proof: Using L'Hopital's rule, note that $\lim_{T\to 0} G(T)/g(T) = 0$ and $\lim_{T\to\infty} G(T)/g(T) = \infty$. We now show that $G(T)/g(T)$ is increasing in T:

$$\frac{G(T)}{g(T)} = \frac{\int_0^T \lambda p \exp(-\lambda pt)(\lambda pt)^{c-1}/(c-1)! dt}{\lambda p \exp(-\lambda pT)(\lambda pT)^{c-1}/(c-1)!} = \int_0^T \exp(-\lambda p(t-T))(t/T)^{c-1} dt \tag{15}$$

Make a change of the variable $u = t - T$:

$$\frac{G(T)}{g(T)} = \int_{-T}^{0} e^{-\lambda pu}(1 + u/T)^{c-1} du$$

Taking the derivative

$$\frac{d}{dT}\frac{G(T)}{g(T)} = \int_{-T}^{0} e^{-\lambda pu}(1+u/T)^{c-2}(1-u/T^2)du - (-1)e^{-\lambda p(-T)}(1-T/T)^{c-1} = \int_{-T}^{0} e^{-\lambda pu}(1+u/T)^{c-2}(1-u/T^2)du$$

Going back to the original variable $t = u + T$,

$$\frac{d}{dT}\frac{G(T)}{g(T)} = \int_{0}^{T} e^{-\lambda p(t-T)}(1+\frac{t-T}{T})^{c-2}(1-\frac{t-T}{T^2})dt \geq 0$$

$G(T)/g(T)$ is 0 at $T = 0$ and increases to ∞ with T.

Proof of Theorem 1

It is sufficient to argue that V_N converges in distribution to V_T if $N = \lambda T$. Let the probability space of Y_i be $(\Omega_i, \mathcal{F}_i, \mu_i)$. Then for every fixed $a \geq 0$,

$$P(V_N \leq a) = \ldots \int_{\Omega_2}\int_{\Omega_1} P(\sum_{i=1}^{X_N(p)} Y_i \leq a, Y_1 = y_1, Y_2 = y_2, \ldots)d\mu_1 d\mu_2 \ldots = \ldots \int_{\Omega_2}\int_{\Omega_1} P(X_N(p) \leq h(a, y_1, y_2, \ldots))d\mu_1 d\mu_2 \ldots.$$

The existence of function $h(a, y_1, y_2, \ldots)$ follows from $Y_i \geq 0$ for all i. We can obtain similar expressions for the demo phase case:

$$P(V(T) \leq a) = \ldots \int_{\Omega_2}\int_{\Omega_1} P(Po(p\lambda T) \leq h(a, y_1, y_2, \ldots))d\mu_1 d\mu_2 \ldots.$$

By the Poisson approximation to Binomial distribution, $P(X_N(p) \leq h(a, y_1, y_2, \ldots)) \to P(Po(p\lambda T) \leq h(a, y_1, y_2, \ldots))$ as N grows. Noting that $0 \leq P(X_N(p) \leq h(a, y_1, y_2, \ldots)) \leq 1$ and invoking dominated convergence theorem successively theorem:

$$\ldots \int_{\Omega_2}\int_{\Omega_1} P(X_N(p) \leq h(a, y_1, y_2, \ldots))d\mu_1 d\mu_2 \ldots \to \ldots \int_{\Omega_2}\int_{\Omega_1} P(Po(p\lambda T) \leq h(a, y_1, y_2, \ldots))d\mu_1 d\mu_2 \ldots$$

This implies $P(V_N \leq a) \to P(V(T) \leq a)$ for all a and completes the proof. □

Lemma 6 *Let be $X(N) \sim N(N, N\bar{\sigma}^2)$ then $P(c \leq X(N) \leq r)$ decreases in N as long as $N > (c+r)/2$.*
Proof: We study two cases: $N \geq r$ and $r > N \geq (c+r)/2$. Note that

$$P(c \leq X \leq r) = P\left(\frac{c-N}{\sqrt{N}\bar{\sigma}} \leq z \leq \frac{r-N}{\sqrt{N}\bar{\sigma}}\right)$$

where z is the standard Normal variant. For $N \geq r$, $c - N < r - N \leq 0$. As we decrease N, the interval $[(c-N)/(\sqrt{N}\bar{\sigma}), (r-N)/(\sqrt{N}\bar{\sigma})]$ becomes wider and moves from negative towards zero. Since z is more likely to be around zero than be negative, the probability of z falling into the interval increases as N decrease.

For $r > N > (c+r)/2$, $c - N < 0 < r - N < |c - N|$. Then for a small $\epsilon > 0$,

$$P\left(\frac{c-N}{\sqrt{N}\bar{\sigma}} \leq z \leq \frac{r-N}{\sqrt{N}\bar{\sigma}}\right) \geq P\left(\frac{c-N-\epsilon}{\sqrt{N}\bar{\sigma}} \leq z \leq \frac{r-N-\epsilon}{\sqrt{N}\bar{\sigma}}\right) \geq P\left(\frac{c-N-\epsilon}{\sqrt{N+\epsilon}\bar{\sigma}} \leq z \leq \frac{r-N-\epsilon}{\sqrt{N+\epsilon}\bar{\sigma}}\right)$$

where the first inequality follows from $f_z((c-N)/(\sqrt{N}\bar{\sigma})) < f_z((r-N)/(\sqrt{N}\bar{\sigma}))$ for $0 < r - N < |c - N|$ and the second from the shrinking of the interval as N increases. Reading the inequality above in terms of $X(N)$ and $X(N + \epsilon)$, we establish the result. □

Proof of Lemma 1.c and d

c) We provide the proof for random R. We show that $\Pi^2(c+1,N)-\Pi^2(c,N)$ can have at most one sign change.

$$\begin{aligned}\Pi^2(c+1,N)-\Pi^2(c,N) &= -cP(X_N(p)=c)P(R\geq c)+\sum_{i=c}^{N}P(X_N(p)=i)P(R\geq i)\\ &= P(X_N(p)=c)P(R\geq c)\left(\sum_{i=c}^{N}\frac{P(X_N(p)=i)}{P(X_N(p)=c)}\frac{P(R\geq i)}{P(R\geq c)}-c\right)\end{aligned}$$

After inserting binomial probabilities and some algebra, we obtain

$$\begin{aligned}\Pi^2(c+1,N)-\Pi^2(c,N) &= P(X_N(p)=c)P(R\geq c)\left(\sum_{i=c}^{N}\frac{c!(N-c)!}{i!(N-i)!}p^{i-c}(1-p)^{-(i-c)}\frac{P(R\geq i)}{P(R\geq c)}-c\right)\\ &= P(X_N(p)=c)P(R\geq c)\left(\sum_{i=0}^{N-c}\binom{N-c}{i}\frac{1}{\binom{i+c}{i}}(p/(1-p))^{i}\frac{P(R\geq i+c)}{P(R\geq c)}-c\right).\end{aligned}$$

Using (12) with $x=i$, we know that $P(R\geq i+c)/P(R\geq c)$ is decreasing in c for every i. Applying this observation to the term inside the parenthesis, we conclude that the term decreases in c as well and complete the proof.

d) Clearly $\Pi^2(c,N)$ of (9) increases in N for $N<r=R$. We will examine only $N\geq r$ in which case,

$$\Pi^2(c,N)=c\sum_{i=c}^{r}P(X_N=i).$$

Conditioning $X_{N+1}(p)$ on the $i+1$st trial:

$$P(X_{N+1}(p)=i)=pP(X_N(p)=i-1)+(1-p)P(X_N(p)=i), \tag{16}$$

after reorganizing terms

$$P(X_{N+1}(p)=i)-P(X_N(p)=i)=p(P(X_N(p)=i-1)-P(X_N(p)=i)).$$

Then the first difference of profit becomes

$$\begin{aligned}\Pi^2(c,N+1)-\Pi^2(c,N) &= c(\textstyle\sum_{i=c}^{r}P(X_{N+1}(p)=i)-\sum_{i=c}^{r}P(X_N(p)=i))\\ &= cp(P(X_N(p)=c-1)-P(X_N(p)=r))\end{aligned}$$

To finish the proof, it is sufficient to note that $P(X_N(p)=c-1)-P(X_N(p)=r)$ is positive for $c-1\leq N<r$ and to claim that it has at most one sign change as N increases.

Claim: If $P(X_N(p)=c-1)-P(X_N(p)=r)\leq 0$ then $P(X_{N+1}(p)=c-1)-P(X_{N+1}(p)=r)\leq 0$.

Proof of the Claim: Using (16) for $i=c-1$ and $i=r$, $P(X_{N+1}(p)=c-1)-P(X_{N+1}(p)=r)\leq 0$ if $P(X_N(p)=c-2)-P(X_N(p)=r-1)\leq 0$, in addition to the condition in the Claim. Now consider the (unique) mode of $X_N(p)$ ($mode(X_N(p))$), $c-1$ and r relative to each other. Three cases follow:

1. $mode(X_N(p))<c-1\leq r$: This case implies $P(X_N(p)=c-1)-P(X_N(p)=r)>0$, so it is not possible.

2. $c-1 \leq mode(X_N(p)) < r$: Under this case, $P(X_N(p)=c-2) \leq P(X_N(p)=c-1) \leq P(X_N(p)=r) \leq P(X_N(p)=r-1)$.

3. $c-1 \leq r \leq mode(X_N(p))$: This directly implies $P(X_N(p)=c-2) \leq P(X_N(p)=r-1)$.

Putting all the cases together claim is proved. Then using the claim, we conclude that $\Pi^2(c,N+1)-\Pi^2(c,N)$ changes its sign at most once. □

References

[1] I. Amoribieta, K. Bhaumik, K.Kanakamedela and A.D. Parkhe (2001). *Programmers abroad: a primer on offshore software development.* The McKinsey Quarterly, No.2: 128-139.

[2] W.O. Bearden and J.E. Teel (1983). *Selected determinants of consumer satisfaction and complaint reports.* Journal of Marketing Research, Vol.20: 21-28.

[3] J.R. Bettman, M.F. Luce and J.W. Pain (1998). *Constructive consumer choice processes.* Journal of Consumer Research, Vol.25 December: 187-217.

[4] G.J. Biehal (1983). *Consumers' prior experiences and perceptions in auto repair choice.* Journal of Marketing, Vol.47: 82-91.

[5] A. Billionnet and F. Calmels (1996) . *Linear programming for the 0-1 quadratic knapsack problem.* European Journal of Operational Research, Vol.92 No.2: 310-325.

[6] H.W. Block, T.H. Savits and H. Singh (1998). *The reversed hazard rate function.* Probability in Engineering and Informational Science, Vol.12 No.1: 69-90.

[7] R. Chatterjee and J. Eliashberg (1990). *The innovation diffusion process in a heterogeneous population: A micromodelling approach.* Management Science, Vol.36 No.9: 1057-1074.

[8] M.L. Eaton (1082). *A review of selected topics in multivariate probability inequalities.* The Annals of Statistics, Vol.10 No.1: 11-43.

[9] F. Gibney and D. Eisenberg (2002). *The engine stalls at AOL.* Time, Vol.159 No.6: 39-42.

[10] A. Heiman and E. Muller (1996). *Using demonstration to increase new product acceptance: controlling demonstration time.* Journal of Marketing Research, Vol.33 No.4: 422-430.

[11] T. Dalgıc and I. Hezar (2002). *Principles of global niche marketing.* Working paper, OSIM, University of Texas at Dallas.

[12] M. Janson and A. Subramanian (1996). *Packaged software: Selection and implementation policies.* INFOR, Vol.34 No.2: 133-151.

[13] M. Kijima (1998). *Hazard rate and reverse hazard rate monotonocities in continuous time Markov chains.* Journal of Applied Probability, Vol.35: 545-556.

[14] D.S. Kempf and R.E. Smith (1998). *Consumer processing of product trial and the influence of prior advertising: as structural modelling approach.* Journal of Marketing Research, Vol.35 No.3: 325-338.

[15] R. Kivetz and I. Simonson (2000). *The effects of incomplete information on consumer choice* Journal of Marketing Research, Vol.37 No.4: 427-448.

[16] M.H. Meyer and A.Sathi (1985). *A multiattribute model of consumer choice during product learning.* Marketing Science, Vol.4 No.1: 41-61.

[17] J.D. Musa (1993). *Operational profiles in software reliability engineering.* IEEE Software, Vol.10 No.2: 14-32.

[18] A.V. Muthukrishan and F.R. Kardes (2001). *Persistent preferences for product attributes: the effects of the initial choice context and uninformative experience.* Journal of Consumer Research, Vol.28 June: 89-104.

[19] S. Ozekici, I.K. Altinel and S. Ozcelikyurek (2000). *Testing of Software with an operational profile.* Naval Research Logistics, Vol.47 No.8: 620-634.

[20] W.A. Parrette (1991). *UNIX: For application developers.* McGraw-Hill Inc, New York.

[21] J.M. Pennings and F. Harianto (1992). *Technological networking and innovation implementation.* Organization Science, Vol.3 No.3: 356-382.

[22] J.A. Quelch and L. R. Klein (1996). *The Internet and international marketing.* Sloan Management Review, Vol.37 No.3 : 60-75.

[23] J. Roberts and G.L. Urban (1988). *Modelling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice.* Management Science, Vol.34 No.2: 167-185.

[24] M. I. Taksar (2000). *Optimal risk and dividend distribution control models for an insurance company.* Mathematical Methods of Operations Research, Vol.52: 1-42.

[25] V.A. Zeithaml (1988). *Consumer perceptions of price, quality and value: a means-end model and synthesis of evidence.* Journal of Marketing, Vol.52 July: 2-22