UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

SAMUELSON LAW, TECHNOLOGY & PUBLIC POLICY CLINIC
CENTER FOR CLINICAL EDUCATION

ACTING CLINICAL PROFESSOR AND DIRECTOR
Deirdre Mulligan
(510) 642-0499
dmulligan@law.berkeley.edu

SCHOOL OF LAW (BOALT HALL)
BERKELEY, CALIFORNIA 94720-7200
TELEPHONE (510) 643-4800 $
FAX (510) 643-4625
http://www.law.berkeley.edu/academics/samuelson/

February 26, 2007

Federal Trade Commission/Office of the Secretary
Room H-135 (Annex E)
600 Pennsylvania Avenue, NW
Washington, DC 20580.

Re: Negative Option Workshop - Comment P064202

Dear Commissioners,

Thank you for giving us the opportunity to discuss our research at the January 25, 2007 workshop on negative options marketing. The Samuelson Law, Technology & Public Policy Clinic submits the following comments to supplement the presentations of Jens Grossklags and Nathan Good. These materials are relevant to several key issues in negative option marketing: consumers' reading behavior; installation behavior; presentation of price; intertemporal decision making; and transaction costs.

Reading Behavior

In a recent workshop paper, *Empirical studies on software notices to inform policy makers and usability designers*,[1] Jens Grossklags and Nathan Good perform a preliminary assessment of issues around presentation of notices. They report on average length, average readability and provide views on reading behavior during the experiment:

> We have presented results on readability and presentation of EULAs from 50 popular free or free-to-try programs available for download on a distribution page. We suggest

[1] Available online at http://usablesecurity.org/papers/grossklags.pdf.

> that the length and complexity of documents can significantly lower the notice and consent success rate achieved...
>
> Without significant improvements to notice and consent procedures for consumer programs it is doubtful that most consumers genuinely assent to the use of their desktops for advertisements, the installation of software with behavior that falls within the broad definition of spyware, or limitation of usage rights. We do not expect that there exists a one-size-fits-all solution, in particular, given the increasing popularity of mobile and small-screen devices. Notice and consent involves many stakeholders. Companies are urged to improve their information dissemination practices and regulation may carefully adjust misaligned incentives in the market place. But improved notice procedure will likely result also in a more substantive obligation for users to read notices.

Also relevant to consumer reading behavior is an "experiment" performed by PC Pitstop. That organization hypothesized that hardly any consumer reads EULAs. To demonstrate this hypothesis, PC Pitstop offered to pay money to anyone who read the company's EULA and wrote to PC Pitstop:

> …included a clause in one of its own EULAs that promised anyone who read it, a "consideration" including money if they sent a note to an email address listed in the EULA. After four months and more than 3,000 downloads, one person finally wrote in. That person, by the way, got a check for $1,000 proving, at least for one person, that it really does pay to read EULAs.[2]

Installation Behavior

In *Noticing Notice: A Large-Scale Experiment on the Timing of Software License Agreements*,[3] Nathaniel S. Good, Jens Grossklags, Deirdre K. Mulligan, and Joseph A. Konstan studied the problem of how individuals made decisions to install spyware. They found:

> Our study of 222 users showed that providing a short summary notice, in addition to the End User License

---

[2] Larry Magid, *It Pays To Read License Agreements*, PC Pitstop (n.d.), available at http://www.pcpitstop.com/spycheck/eula.asp.

[3] Available at http://www.ischool.berkeley.edu/~jensg/research/paper/Grossklags07-CHI-noticing_notice.pdf.

> Agreement (EULA), before the installation reduced the number of software installations significantly. We also found that providing the short summary notice after installation led to a significant number of uninstalls. However, even with the short notices, many users installed the program and later expressed regret for doing so. These results, along with a detailed analysis of installation, regret, and survey data about user behaviors informs our recommendations to policymakers and designers for assessing the "adequacy" of consent in the context of software that exhibits behaviors associated with spyware.

Presentation of Price

In *...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay*,[4] Tanjim Hossain and John Morgan demonstrate that how price is presented can affect consumers' decision making. They argue that: "consumers pay attention to the list price while neglecting the 'add-on' price associated with shipping or with the gratuity. Therefore, they systematically underestimate the total out of pocket price paid for the items being purchased or consumed and hence the seller is able to sell more of a given item than had that firm simply listed a single total price." This work has direct relevance to how negative options can be marketed to mask fees and other charges to consumers. Hossain and Morgan found:

> Many firms divide the price a consumer pays for a good into two pieces-the price for the item itself and the price for shipping and handling. With fully rational customers, the exact division between the two prices is irrelevant-only the total price matters. We test this hypothesis by selling matched pairs of CDs and Xbox games in a series of field experiments on eBay. In theory, the ending auction price should vary inversely with the shipping charge to leave the total price paid constant.
>
> Contrary to the theory, we find that charging a high shipping cost and starting the auction at a low opening price leads to higher numbers of bidders and higher revenues when the shipping charge is not excessive.

Also relevant to consumers' perception of product cost is Oren Bar-Gill's *Bundling and Consumer Misperception*.[5] Bar-Gill argues: "Consumer misperception of the costs and benefits associated with a product or service is prevalent. It can be the product of

---

[4] 6 Advances in Economic Analysis & Policy 3 (2006), available at http://faculty.haas.berkeley.edu/rjmorgan/eBay.pdf.
[5] Published in American Law & Economics Association Annual Meetings (2006).

imperfect information or imperfect rationality (or both). It can be independent of any action taken by sellers. It can be instigated by sellers. And it can be mitigated by sellers." Bar-Gill's paper offers policymakers a framework for evaluating whether bundling of products affects consumers' understanding of costs.

Intertemporal Decision Making

Bar-Gill's paper also discusses how the presentation of costs over time can affect the consumers' perception of price.[6] His discussion is directly relevant to the marketing of negative options, which present price information over time in such a way that may contribute to consumer misperception:

> One special type of bundling that directly responds to consumer misperception is the intertemporal bundling achieved through multi-period subscriptions. In intertemporal bundling the only difference between the bundled components is the timing. For example, a year-long subscription—to a magazine, a wireless or landline phone service, an ISP, or a health club—provides the same service every month throughout the year. Why are multi-period subscriptions so common? Why not sell only single-period products or services? Consumer misperception provides an answer…
>
> Many subscription services charge a one-time subscription fee as described in the preceding health club example. Other subscription services charge a per-period fee, but follow a fee schedule very different from per-period marginal cost pricing. In particular, many subscription services charge different per-period prices for different periods within a multiperiod subscription. Particularly common is the practice of offering a low price, or even a zero price, for an introductory period. One explanation for this practice is that sellers are exploiting consumer misperception—this time, consumer underestimation of future use. When signing on to a yearlong subscription service with a "two-month free" introductory offer, some consumers think that they will end the subscription after the first two months. Not all of these consumers actually end their subscription after two months. Put differently, sellers may be responding to consumers' underestimation of the length of the period during which they would need or want a subscription with the specific seller. If consumers

[6] *See also* Stefano DellaVigna and Ulrike Malmendier, *Paying Not to Go to the Gym* (Apr. 15, 2005).

> underestimate the effective subscription period, then sellers in a competitive market will backload their price as much as possible.
>
> This too is an example of intertemporal bundling. The cement holding the bundle together is the cost of switching from one seller to another or simply the cost of terminating the subscription. It is consumers' underestimation of these switching costs that explain the viability of a below marginal cost introductory fee. Since switching costs keep the bundle together, it is not surprising that sellers do not make a special effort to reduce these costs. Perhaps the notorious ten minute "please hold for the next available representative" wait that must be endured to cancel a subscription is not merely the result of a seller's attempt to economize on the size of its customer service department. More direct measures designed to increase switching costs are lock-in clauses and termination fees. If consumers underestimate the cost of switching or, equivalently, underestimate the likely length of their subscription, sellers who fail to take advantage of switching cost-induced bundling and offer only per-period marginal cost pricing will not survive in a competitive market.

Transaction Costs

The Internet is a sophisticated communications medium that radically reduces transaction costs between individual nodes (like the telephone did over hundred years ago). Transaction costs play less of a role and make the scenario that Professor Katz described somewhat artificial and maybe not at the heart of the issue. In 1987, Malone et al. explained three different effects of information technology on the exchange process.[7] These three effects suggest that Katz's focus on transaction costs is more suitable for offline transactions rather than the different situation online.

In *Technological Advances, Consumer Transaction Costs, and Consumer Welfare*, Rajeev K. Tyagi explains, that lowering transaction costs may even cause detriment to the consumer:

---

[7] First, the communication effect: increase information flux per unit of time; second, the electronic integration effect: easier linkage between buyer and seller; and third, the electronic brokerage effect: the contracting process between seller and buyer becomes more efficient and effective. Thomas W. Malone, Joanne Yeates, and Robert I Benjamin, Electronic Markets and Electronic Hierarchies, 30 Communications of the ACM 484 (1987).

> …Using a model in which consumers are simultaneously heterogenous in their transaction costs and in their marginal valuation of product quality, we highlight two mechanisms that can cause such reductions in consumer transaction costs to lower consumer surplus and reduce consumer share of the total social surplus. We also show that when a seller can invest to reduce consumer transaction costs, consumers may be better off with higher transaction costs. Finally, the paper shows how such reductions in consumer transaction costs can either lower or raise consumer prices.

Conclusion

Again, we appreciate the Commission's efforts in addressing challenges in the negative option marketing landscape. We hope these materials are helpful in supplementing the record.

Respectfully submitted,

/s

Chris Hoofnagle
Senior Staff Attorney
Samuelson Law, Technology & Public Policy Clinic

/s

Jens Grossklags
Ph.D. Candidate
School of Information

Empirical studies on software notices to inform policy makers and usability designers

Jens Grossklags, Nathan Good

School of Information, University of California Berkeley,
102 South Hall, 94720 Berkeley, United States
{jensg, ngood}@ischool.berkeley.edu

This version: February 6, 2007

Keywords: HCI, privacy, security, legal issues, End User License Agreement, notice

1 Introduction

Human-Computer Interface (HCI) research is of increasing importance to security researchers as well as policy makers. As Internet access has become more prevalent, many issues that previously concerned only sophisticated technical users are now issues affecting the public at large. Issues surrounding digital privacy, copyright, electronic voting, notice & consent, and location-based systems are being pushed into the public policy arena because of advances in technology. Public policy advocates have traditionally accessed academic research as one means of understanding a problem, and HCI research provides a deeper understanding of the many technological issues discussed today. Indeed, many of the recent issues with new technologies have roots in problems that HCI has dealt with for years. For example, usability issues have caused security and privacy concerns for a broad range of issues such as electronic voting machines [Bederson et *al.*, 2003], the sharing of private personal information over P2P networks [Good and Krekelberg, 2003], phishing attacks [Dhamija et *al.*, 2006; Yee and Sitaker, 2006], and email message encryption [Whitten and Tygar, 1999].

Our research focuses on the primary means that security and privacy related information is currently communicated to the end user: the software notice and license agreements. We find software with potentially unwanted consequences and risks such as Spyware and Adware to be a particularly appealing field of study. We observe that in the marketplace millions of programs are installed bundled with advertisements and privacy-invading technologies [AOL/NCSA, 2004; Earthlink, 2005]. Many of these installations are made without any notice and consent procedures (e.g., through drive-by downloads), however, a surprisingly large number of programs are installed through deliberate user action and involving some form consent process. Users desire the functionality of programs they download, but frequently seem ill-informed about potential risks and negative consequences of installations. Indeed, the reason that spyware is difficult to accurately define is that the same piece of software may be considered unacceptable spyware by one user, an

acceptable trade for other services by another, or a valuable personalization system or notifier by a third. Because of this user-centered definition of what constitutes spyware, for some portion of software that meets the definition of spyware, it seems inappropriate to adopt an outright ban. Early efforts to combat spyware—much like anti-virus software efforts—measured their success based on how infrequently the software was installed. While such a measure can help provide security, it may also limit users' access to certain software combinations by denying them the opportunity to trade some privacy, speed, or attention for services or information they actually value. Imagine if your computer "protected" you by preventing you from ever transmitting your credit card information over the Internet; it would perhaps reduce your vulnerability to identity theft, but would at the same time deny you the benefits of shopping online. As response to usage restrictions due to security software (e.g., firewalls, anti-spyware) users might experience frustration. Left with their dissatisfaction users will often disable security technologies and, therefore, reduce overall security of the system.

In the case of spyware, it is not simply that the monitoring or notifications themselves may be valuable, but anecdotal evidence suggests [Delio, 2004], and our previous study confirms, that some users are willing to install spyware when the desired application with which it is bundled is of perceived high utility, and a comparable product without spyware is unavailable or unknown to the user [Good et *al.*, 2005]. In other words, at least in situations where users are unaware of other options, they are willing to give up some privacy, screenspace, or bandwidth as "payment" for an unrelated service or product they value. Accordingly, managing spyware requires that we engage the user in controlling their desktop instead of assuming we can simply do it for them.

We want to add that our works ties into an already existing public policy debate on balance of power between consumers and commercial entities. There has been a longstanding debate on the appropriateness of standard form contracts in business interactions with consumers. Already in 1943 a law review article criticized sellers for the harsh terms and conditions that are forced upon consumers in a take-it-or-leave-it fashion in many contracts [Kessler, 1943]. Many commentators have discussed the conflict of interests between the consumer and the business entity. On the one hand, businesses strive for monetary earnings but want to minimize potential liabilities out of transactions conducted in the marketplace. Accordingly, the typical vendor software license has much less to do with the licensing of technology than it does with the creation of multiple revenue streams flowing from the user to the vendor and the elimination or minimization of most forms of accountability from the vendor to the user [Overly and Kalyvas, 2004]. On the other hand, users want to benefit from the functionality of a program and other aspects that create hedonic and intangible values while limiting privacy, security and other risks of the interaction. Further, users want to reduce the effort involved in making sound decisions; standard form contracts help in an overwhelming number of situations to reduce transaction costs for businesses and consumers. In fact, it has been estimated that 99% of all commercial contracts are standard form contracts [Slawson, 1971].

Generally, economic forces should help to balance consumer desires and concerns with business interests. However, a recent research study supported the view that market conditions are generally uncorrelated with contract terms (for example, by

asking how price and market concentration determine the harshness of contract terms). The study also indicated that license terms on average provide less consumer protection than the Uniform Commercial Code baseline regulations [Marotta-Wurgler, 2005].

Despite this criticism it must be noted that standardization of contracts creates also many important benefits for consumers in the marketplace through cost reduction and allowing consumers to limit their attention to terms of the deal they particularly care about. End User License Agreements also serve as an important medium of communication between businesses and lawyers. However, in absence of simpler and more conspicuous modes of communication to the consumer (e.g., short notices [Good et al., 2007]) these agreements also serve as important information source for download and installation decisions. Our prior research suggests that users are often even uninformed about aspects of a program they genuinely are concerned about (such as pop-up advertisements and spyware). The result is unwanted installations of programs that are later regretted [Good et *al.*, 2005]. The current paper aims to explain this disconnect between consumer wishes and their market choices in more detail.

To this effect our research task is focused on evaluating the readability and usability of End User License Agreements (EULAs) that represent the legal state of the art of informing users and obtaining user consent. We present preliminary results from an empirical study of 50 popular consumer programs on the accessibility and readability of the associated EULAs. We also briefly discuss results from a user study involving 64 users in program installation tasks. Users were observed during their interaction with an experimental program installation environment. We recorded their reading behavior, decisions to complete or cancel an installation and their responses to post-experimental surveys.

Both studies are significant extensions of our prior work [Good et *al.*, 2005]. On the one hand, we discussed in our first paper the readability metrics of only 5 programs that we randomly selected. The current study gives a more thorough overview of the notice of consent practices for an important sample of 50 consumer programs that are the most popular freeware/shareware or free-to-test versions across multiple functional categories. On the other hand, we also conducted a more thorough experimental analysis. In Good et *al.* [2005] we reported results of an in-depth user study on notices with a small sample set of 30 users across three experimental conditions. Many questions were left open and in need of further experimentation to determine or substantiate results. In this paper we will only be able to cover initial results from the new study due to space limitations but want to give the reader sufficient insights on the current experiment and the results.

Our goal is to facilitate interdisciplinary research on informed consent, and facilitate the creation of usable law based on usable technologies. The emergence of location-based systems, camera phones, RFID & sensor networks demand further research on how to better notify consumers about their privacy rights. In an increasingly networked society, new technologies and methods of distributing media raise issues regarding how content providers and consumers view ownership of digital content. Efforts to understand these complexities and inform security researchers as well as policy makers could benefit directly from participation of HCI practitioners as well as from prior research in this area.

2 Empirical study of End User License Agreements

As the data set, we chose Download.com's top 50 most downloaded software programs[1]. Download.com is a leading source for primarily free or free-to-try consumer software downloads covering major software vendors as well as small distributors but the program offerings are not necessarily representative of all consumer programs available.[2]

Related to our study Kucera et al. [2005] reported on the prevalence of spyware in a similar sample of download.com's most popular programs.[3] When defining spyware narrowly as programs that surreptitiously collect personal information from computers linked to the Internet the authors confirmed the existence of spyware for three of those programs. The current policy of the distributor does not allow for software including viruses or spyware[4], however, the website does not clearly define these terms.

Recently, the Anti-Spyware Coalition formulated a broader characterization of spyware (and other potentially unwanted software).[5] Their definition includes technologies deployed without appropriate user consent and/or implemented in ways that impair user control over:

- Material changes that affect their user experience, privacy, or system security;
- Use of their system resources, including what programs are installed on their computers; and/or
- Collection, use, and distribution of their personal or other sensitive information.

Our focus in this study is on analyzing the readability of license terms distributed with typical software available to consumers. We defer the content analysis of these agreements and a technical analysis to later stages of our research. It is to be expected that many of the terms are unremarkable and of little concern to the user [Schechter, 2000]. However, we note that our preliminary analysis suggests that the programs included in our sample included terms (including privacy implications, restrictions of usage and legal rights, distribution of adware) that are likely to the detriment of many consumers and may overlap with a broader definition of spyware.

The software programs were grouped by the category and subcategory types given by Download.com: Internet (26%), Audio & Video (22%), Security & Anti-Spyware (20%), Games (10%), Utilities (10%), Design & Photo (2%), Desktop Enhancements (2%), Developer Tools (2%), and Networking (2%).

---

[1] We chose the Top 50 most downloaded software programs for the week ending April 9, 2006.

[2] They report to serve "over 27,000 publishers representing 35,000 products and 132 countries around the world, our product library generates over 2.3 million downloads each day." http://www.upload.com/1200-21_5-5139590.html, visited February 5, 2007.

[3] Kucera et *al.* obtained data for the week ending January 12, 2003.

[4] See, for example, http://www.upload.com/1200-21_5-5081541.html?tag=fd, visited February 5, 2007.

[5] http://www.antispywarecoalition.org/documents/DefinitionsJune292006.htm

2.1 Timing and Presentation of the End User License Agreement presentation to the user

For each software program, we initiated the downloading and installation process, and stopped the process at the point where we encountered a EULA. We copied the EULA that appeared on-screen and canceled the download at this point, and thus did not capture any additional terms that may have been presented to the user after this point. If we did not encounter a EULA during the installation process or after program installation we expanded the search to the distributors' website. See Figure 1 for a typical display situation of a EULA during the installation process.

We observed that the terms were presented at different stages during the installation process for different programs: e.g., before the installation had begun or after the installation process. Knight Online 1.299 showcased a so-called "first-run notice"[6] that occurs the first time a (or potentially each) user starts the program.

Common sense regarding notice and consent would dictate that information to users should be provided before an installation is initiated or completed. In research reported elsewhere we investigate the impact of the timing of notices more thoroughly [Good et *al.*, 2007].

More problematic from an accessibility standpoint is the omission of notice during the installation process. Adobe Reader and Irfan View, had EULAs only on their websites. It is doubtful whether users would search for these terms if not included in the installation dialog. One further significant difference appeared between the two programs. Irfan View's installer was directly accessible at download.com's website. Users interested in Acrobat were redirected to Adobe to initiate download[7]. Under the 'download' button on Adobe's site the EULA was accessible by clicking on a link "By downloading software from the Adobe web site, you agree to the terms of our license agreements […]".

Software providers differ in the type of notice and presentation of notice they provide to the user. From a legal perspective these installation scenarios introduce different challenges and likely impose different consequences on the user. For example, courts have started to differentiate between different modes of presentation of notice when they decide whether a user is bound by terms. See, for example, Casamiquela [2002] who discusses caselaw and legal theory on browsewrap versus clickwrap agreements[8].

---

[6] A recent report by Microsoft [Microsoft, 2006] distinguishes between Just-In-Time, First Run, Installation Time, and Out-of-The-Box notices. Out-of-the-Box notices were not observable from our download.com sample. We did not test for Just-In-Time notices that occur in the moment before sensitive data is transmitted or some other potentially harmful or unwanted action is undertaken by the program. The majority of the programs featured Installation Time notices.

[7] We expect that many users have access to Adobe Acrobat's installation file also without visiting Adobe's website.

[8] Clickwrap agreements include situations in which a software vendor requires users to click "I agree" or similar buttons or click-check radio buttons or boxes to signify consent. A license is likely to be characterized as browsewrap if only a link to the terms is available to the user instead of the complete license.

We were unable to locate a license agreement for Limewire, Limewire (Mac), and Morpheus on the respective company websites or during the installation. There is anecdotal evidence that file-sharing companies refrain from using EULAs in an effort to limit possible liability for contributory infringement, as the presence of a license agreement would establish an ongoing relationship with the customer.

Figure 1 End User License Agreement presentation during installation (McAfee AntiVirus vso_10027_en-us-30day)

While the typical presentation of a EULA follows the pattern observable in Figure 1 access to the agreement is not always obvious. For example, the installation dialogue displayed in Figure 2 only links to the read me file that, however, contains a contractual document.[9] We believe that this access regime is from a user point of view totally unexpected.

EULAs were often presented in a format that limits users in gaining a quick overview over the terms covered. For example, the notice screen displayed in Figure 1 allows the user to only review about 50 words at a time without scrolling. The complete notice, however, is 5500 words long.

[9] Excerpt from readme file: "THIS SOFTWARE IS PROVIDED AS IS WITHOUT WARRANTY OF ANY KIND, WHETHER EXPRESS OR IMPLIED, INCLUDING WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE WHICH ARE HEREBY DISCLAIMED. […]"

Figure 2 Need for Speed Most Wanted PC demo installation dialogue

2.2 Length of EULAs

For the software programs which had EULAs, the average EULA length was 2752 words (std. dev. = 2228.8), corresponding to about 11 pages of double-spaced text (see Figure 3). Assuming that the reading difficulty of the EULA is average as reported in psychology research, the average reading time for the EULA is about 13 minutes[10]. The shortest EULA (Little Fighter) was 111 words, the longest (Adobe Reader) was 9313 words, corresponding to approximately 41 pages of double spaced text and an average reading time of 47 minutes. It is likely that the length of time it will take an average consumer to *understand* the EULA is even longer than 47 minutes.

Practically, the average EULA is even longer than what is shown in Figure 3. Many of the EULAs have website links to additional information and terms that are incorporated into the EULA, such as Terms of Use, Terms of Service, Privacy Policies, and third party EULAs. For this study we did not review any of these additional linked documents. Further, to fully understand what they were agreeing to, the user would also have to research various statutes and rules that are mentioned within the text of the EULAs, such as the "Commercial Arbitration Rules of the American Arbitration Association". In one particularly egregious example, Good et *al.* [2003] evaluated a KaZaA EULA, and noted that it contained 17458 words in the

[10] Assuming an average reading rate of 200 words/minute. Lewandowski et *al.* [2003] found an average reading rate for college students of 189 words/minute when subjects were given oral reading probes measuring words read correctly per minute (WRCM). Younger students and elderly citizens will likely read and comprehend slower on average. One reviewer correctly observed that it would be advisable to measure reading speeds specifically for EULAs. We have so far not conducted the required experiment.

EULA itself, 4 hyperlinks to outside sites and policies, 78 locations of third parties and policies, and 5 opt-out options, and would take an average reader approximately 88 minutes to read.

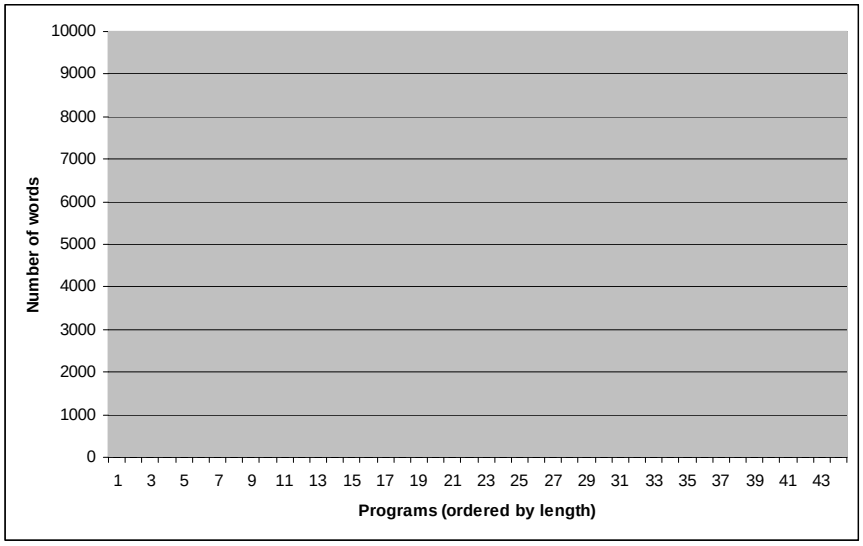


Figure 3 Length of End User License Agreements

2.3 Simple readability measures

We analyzed Flesch-Kincaid Reading Level [Kincaid et *al.*, 1975], and Flesch Reading Ease levels [Flesch, 1948] for the EULAs in the program sample. The Flesch-Kincaid Reading Level uses average sentence length and average number of syllables per word to give a rough measure[11] of a document's readability. The scores range from 1.0 to 12.0, corresponding to the reading level of an average student in grades 1 through 12, respectively. The scores do not go higher than 12.0, and thus the Flesch-Kincaid Reading Level cannot provide information as to whether the reading level is actually higher than a 12th grade reading level. 63% of the EULAs scored 12.0, the highest score possible. The average score was 11.2 (std. dev. = 1.6), with scores ranging from 5.7 to 12.0. Because the scores cannot go higher than 12.0, and

[11] One critic of the Flesch-Kincaid models noted that "to measure readability, coherence and comprehensiveness of a text, more than surface features need to be taken in consideration than surface features alone. Quantitative and qualitative factors like the number of anaphora, number of overlapping text segment, vocabulary difficulty, sentence and text structure, concreteness and abstractness, are equally needed. It is the sum of these and other factors that constitutes cohesion." University of Memphis Institute of Education Sciences, http://cohmetrix.memphis.edu/cohmetrixpr/readability.html , visited on May 11, 2006.

because of the large percentage of EULAs with the maximum score, the average score of 11.3 is likely skewed lower than it should be.

The Flesch Reading Ease also uses average sentence length and average number of syllables per word to give a rough measure of a document's readability. The Reading Ease scoring scale ranges from 0 to 100, with a higher score corresponding to easier reading ease.

As a rule of thumb, scores of 90-100 are considered easily understandable by an average 5th grader. 8th and 9th grade students could easily understand passages with a score of 60-70, and passages with results of 0-30 are best understood by college graduates. *Reader's Digest* magazine has a readability index of about 65, Time magazine scores about 52, and the Harvard Law Review has a general readability score in the low 30s. This test has become a U.S. governmental standard. Many government agencies require documents of forms to meet specific readability levels. Most states require insurance forms to score 40-50 on the test.

The average Reading Ease score was 35.7 (std. dev. = 10.7), with a low of 18.5 (WinZip), and a high of 69.8 (Mario Forever). Fully 89% of the EULAs scored under 50, and only 1 EULA (Mario Forever) scored in the ideal range when for writing for the general population (60-70)[12]. As no EULA scored at either extreme end of the range, the Reading Ease score is likely a better measure of readability than the Reading Level score.

Readability studies were conducted for the domain of privacy notices. Jensen and Potts [2004] found an average reading ease of 34.2 for popular entertainment websites and 36.5 for health care sites. Breese and Burman [2005] found a reading ease of 42.2 for privacy practices of 185 institutions listed in the 2004 US News & World Report's 'best hospitals' issue. Hochhauser [2001] found an average reading ease of 34 for 60 financial privacy notices (and a level of 39 for 31 HIPAA notices [Hochhauser, 2003]).

2.4 Discussion

If privacy or security risks are disclosed in the EULA then the length and reading ease will directly impact users' comprehension and decision making. However, if consumers cannot understand the terms to which they are ostensibly agreeing, have they really formed a valid contract with the company, or do they have a duty to read?

It appears that the EULAs in our sample are characterized by sizable length and low readability. Additionally, they are hampered by user interface issues that make it hard for users to find and access the EULAs and/or attain a quick overview over important terms.

Hochhauser [2001] suggests that several language and presentation modifications can be undertaken to improve readability and understanding. For example, the use of active everyday language, short explanatory sentences in bulleted lists, avoiding imprecise language including double negatives and effective highlighting of important terms can contribute to reader's improved decision making.

---

[12] *See e.g.*, http://www.diabetesvoice.org/issues/2004-09/ Diabetes-related_websites_are_they_readable.pdf.

But grammatical simplification of contracts will not solve all comprehension problems. Research by, for example, Masson and Waldron [1994] demonstrates that the success of simplification of sentence structure etc. is hampered through the complexity of the legal concepts that are at the heart of online notices. Not only legal concepts are hard to understand. Acquisti and Grossklags [2005] discuss consumers' limited knowledge and understanding of privacy and security risks. Further, misaligned economic incentives limit distributors' desire to improve EULA terms (see, for example, Vila et al. [2004]).

EULAs are often written from lawyers for lawyers. However, in absence of better information sources consumers have to rely on them to make successful a priori choices about software programs that are often difficult to uninstall and might significantly impact security, privacy, usage rights and performance.

Some commentators have discussed the role of experts, consumer advocates and user-to-user recommendations as a tool to improve decision making. For example, Hillman [2006] argues that mandatory display of license terms on Web sites will improve access of consumer protection organizations. However, he cautions that the improved accessibility might backfire (at consumer rights) if terms still do not receive added scrutiny, or are not read more often compared to the current notice regime. Download.com alone distributes 35000 programs - it appears unlikely that even all somewhat popular programs do receive enough scrutiny.

In future work we aim to more closely research interface aspects of EULA presentation to the user. We are also interested in analyzing the contents of these agreements to a greater extent.

3 Experiment

Below we report survey results and basic reading time measures observed in the experimental part of the project.

3.1 Experimental setup

Our experimental setup consisted of an experimental portion, followed by two surveys. Subjects were given a unique number, and sheet outlining the basic scenario of the experiment. All of the experiments and surveys were done by each subject independently on a computer located in a laboratory with dividers. As the user passed each portion of the experiment, the application would record the actions and provide the next portion of the experiment.

The experimental portion of our framework was designed to mimic the experience of installing software applications, but also allows us to modify the notice and consent process encountered. We constructed a windows application in C# that would not only depict the installation process as realistic as possible, but also log all user actions (e.g., buttons clicked, time per screen) during the study. Additionally, the application we constructed would provide a launching pad that could dynamically configure each subject's experience based on their user number we provided at the beginning of the experiment. At any time, a user may cancel the installation and return to the landing

screen to start with the next program. Additionally, users may move back and forth between screens as in typical installation programs by hitting the back key. Users were given a user id, which was matched up against a list of acceptable identifiers and associated with a treatment and a program ordering.

A representation of the framework architecture is presented in Figure 4.

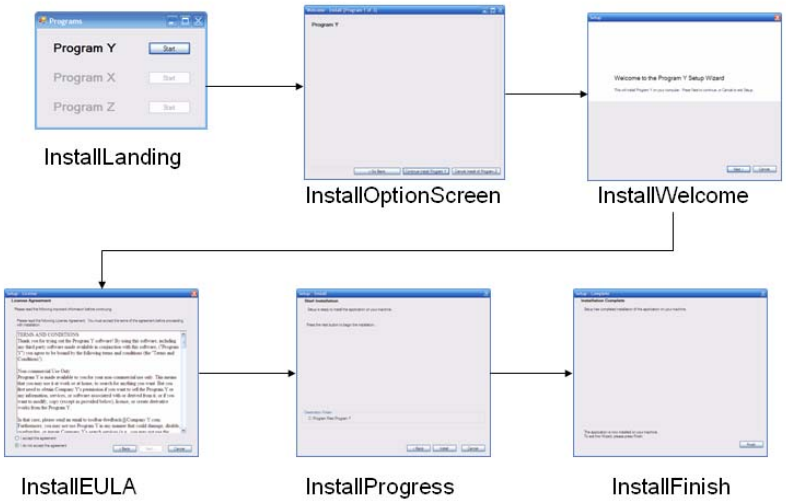


Figure 4 Framework architecture of experiment

We selected popular consumer programs from our previous study [Good et *al.*, 2003] to facilitate comparability of the results and user experience. We chose a browser toolbar, a weather information service and a file sharing application. For the experiment each brand name was removed and replaced with a generic title. The experimental program titles and descriptions are listed below:

- Program X – Weather Information Program
- Program Y – Browser Toolbar
- Program Z – File Sharing Program

We also replaced the brand names and other identifying information in the EULA statements with the above generic titles. The ordering of the programs was randomized within the subject pool.

All three programs vendors disclose in the End User License Agreements that they take significant influence on the user's desktop experience. They differ in the disclosed impact on privacy and security. Some aspects of these programs fall within the broader definition of spyware and adware.

3.2 Subjects

64 subjects participated in the treatment that we discuss below[13]. Subjects were paid $20 for their participation, and were recruited by a university service with access to a subject pool of several thousand students.

On average we had a young and very computer-experienced group of users. For example, More than 80% stated that they maintained their home computer themselves.

3.3 Survey results

Only very few users reported reading EULAs often and thoroughly when they encounter them (1.4%). Members of a larger group categorize themselves as those who often read parts of the agreement or browse contents (24.8%). However, 66.2% admit to rarely reading or browsing the contents of EULAs, and 7.7% indicated that they have not noticed these agreements in the past or have never read them.

Supporting these results, Jensen and Potts [2004] report that for a university service standalone website requiring registration only 0.24% of over 50000 users visited the site's privacy policy. Another software provider reported from an experiment in which a $1000 cash prize was offered in the EULA that was displayed during each software installation, yet the prize was only claimed after 4 months and 3,000 downloads of the software [PC Pitstop].

In contrast to EULA statements, food labels and credit card statements have been subject to substantial standardization and simplification. However, complete information about food ingredients and consequences of signing up for a new credit card are difficult to present to the user in a unified format and labels always need to be selective. Different states take different approaches towards what warnings and information are useful for consumers in their decision making. Similarly, consumer perceptions and reading behavior varies widely across the population. Individuals' health concerns are a strong driver for reading behavior. For example, Kreuter et *al.* [1997] found that patients with high blood pressure searched labels for sodium information, however, did not investigate other ingredients more often than the rest of the population.

With respect to software installations the presence of individual differences in reading behavior and other behaviors suggests that personalized solutions have promise. Analogously, consumers with certain allergies are insufficiently supported by many current food labels. Some Web users might be well-served by the current notice and EULA system, or would be with short summary notices. Others seem likely to ignore such notices and might be willing to accept more restrictions on their installation (e.g., longer delays sequences of confirmations, or approval from another individual) in order to reduce their own risk and later regret. There are many paths to explore in this direction. We also note that a state-by-state approach is unworkable for

[13] Until now we have completed three laboratory experiments with a total of 240 users. Different results of these experiments are reported in Good et *al.* [2007] in which we focus on short notices and the timing of notice presentation to the user (see this paper for more details on the user population and experimental setup).

program downloads from the Internet. Therefore, enforcement action will likely be needed from the federal government or agencies such as the FTC.

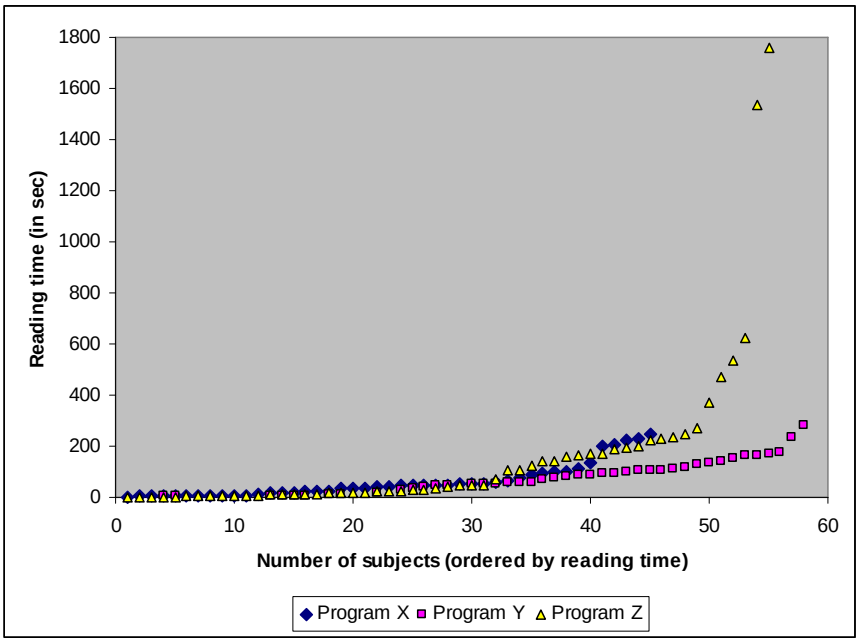


Figure 5 Reading time for End User License Agreement Screen for the three different programs (in sec)

3.4 Reading behavior in the experiment

In this paper we report data for individuals that installed programs X, Y, and/or Z leaving us with 45, 58, 55 observations for the respective programs.[14] On average individuals resided on the screen that showed the complete End User License Agreement in a scrollbox for one or two minutes (Program X: 59.7 sec, 66.4 std.dev.; Program Y: 64.9 sec, 64.4 std.dev.; Program Z (with outliers): 162.6 sec; 323 std.dev.; Program Z (without 2 outliers): 106.6 sec, 141.0 std.dev.). More than 55% of the experimental subjects spent less than one minute on this screen. Only 3.7% deliberate on this screen for more than 5 minutes. It appears the installation of the filesharing program Z caused more individuals to slow down in their reading behavior. We plotted the reading times for the three different programs in Figure 5. The time

[14] Subjects that canceled installation did not always progress through the installation routine until they were able to review the EULA.

required to pass through the EULAs is 14 min, 10 min, 14 min for Program X, Y, and Z, respectively.[15]

We were also interested in the time individuals spent on the EULA screen in comparison to the other parts of the installation dialogue. Since this screen was the only one that contained important information about the program we would expect the ratio between the two measures to be below one. The other screens prompted individuals to merely click to continue.

Up to 32.8% of the users spent more time clicking through screen without important information resided compared to the EULA screen (Program X: 71.1%; Program Y: 67.2%; 74.5%).[16]

3.5 Discussion

The results serve as a benchmark for reading behavior if individuals are unaffected by brand recognition, message framing and sophisticated user interface design techniques. It is not a reading speed test. Rather the study provides insight into the distribution of reading times across a reasonably-sized subject group in a controlled laboratory context. Surprisingly, even without particular knowledge about the programs' terms concerning privacy, security and usage rights and without time pressure almost no subjects spend enough time on the EULA screen to pass through the notice agreement. In contrast, Hillman [2004] reported that one third of the law student respondents to his survey would more likely read notices if the vendor is unknown.

Well-known limitations of laboratory studies apply also to our experiment. We cannot prove that individuals would behave exactly in the same way outside of the laboratory, but we expect similar behavior. Our subject pool consisted mainly out of young and computer-literate college students. We believe them to be a natural target audience for the type of programs in the study. Other demographical groups are likely to demonstrate slightly different behaviors, for example, older people often report higher privacy concern and might act accordingly.

We have already hinted at several reasons for the absence of careful reading. Notices are often impenetrable, written in legalese, too long. Further many consumers do not expect them to contain useful information that helps them to make better decisions [Good et al., 2005]. Consumers feel that it is not worth it to read notices [Hillman, 2006; Vila et *al.*, 2004].

---

[15] Again, using an average reading speed of 200 words/minute [Lewandowski, 2003].

[16] This cut-off level is somewhat arbitrary, but we posit that the reading time on the EULA screen should be, in general, a multiple of the time spent on basically content-free screens that merely state a generic program name and progress of the installation process.

4. Conclusion

We have presented results on readability and presentation of EULAs from 50 popular free or free-to-try programs available for download on a distribution page. We suggest that the length and complexity of documents can significantly lower the notice and consent success rate achieved.[17] According to readability expert Mark Hochhauser [2003], the length of legal documents often creates information overload leading to increased stress, impaired judgment and helplessness. This effect is particularly strong for older readers. Moreover, rewriting these documents in simple language is often impossible [Hochhauser, 2003] and the underlying legal concepts might still be too hard to understand for interested readers [Masson and Waldron, 1994]. All these effects appear particularly strong in EULAs since their length and the range of issues covered in them is beyond, for example, Web privacy notices. We suggested in public FTC hearings that federal authorities should revisit their basic approach to benchmarks with respect to industry self-regulation to create reliable standards for consumers to rely upon [Turow et al., 2006].

We also observed different presentation styles and timing of notice display. This is an additional source of confusion to Web users who will not expect to find important legal information, for example, only on the company's Web site or buried in a read me file. In our current work we are particularly interested in the influence of different but conspicuous timing modes. In treatments not discussed in this paper we explicitly modified the notice experience for the user so that especially designed short notices would appear either at the start or the end of the installation dialogue in addition to the long-form EULA [Good et *al.*, 2007].

Without significant improvements to notice and consent procedures for consumer programs it is doubtful that most consumers genuinely assent to the use of their desktops for advertisements, the installation of software with behavior that falls within the broad definition of spyware, or limitation of usage rights. We do not expect that there exists a one-size-fits-all solution, in particular, given the increasing popularity of mobile and small-screen devices. Notice and consent involves many stakeholders. Companies are urged to improve their information dissemination practices and regulation may carefully adjust misaligned incentives in the market place. But improved notice procedure will likely result also in a more substantive obligation for users to read notices.

Acknowledgements

We are greatly indebted to Susheel Daswani for constructing the experimental framework. Part of this work is conducted jointly with Joe Konstan, Deirdre Mulligan and Becca Shortle. We also thank Chris J. Hoofnagle, Ira Rubenstein and the anonymous reviewers for their valuable feedback and suggestions. This work is supported in part by the National Science Foundation under ITR award ANI-0331659.

[17] This also includes an informed decision not to install a particular program.

References

1. Acquisti, A. and Grossklags, J.: Privacy and Rationality in Individual Decision Making,. IEEE Security and Privacy, Vol. 3, No. 1, January/February (2005) 26-33.
2. AOL and National Cyber Security Alliance: AOL/NCSA online safety study, (October 2004). http://www.security.iia.net.au/downloads/safety_study_v04.pdf
3. Bederson, B.B., Lee, B., Sherman, R.M., Herrnson, P.S., and Niemi, R.G.: Electronic voting system usability issues. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2003) 145 – 152.
4. Breese, P., Burman, W.: Readability of Notice of Privacy Forms Used by Major Health Care Institutions. Journal of the American Medical Association 293 (2005) 1593-1594.
5. Casamiquela, R.J.: Contractual Assent and Enforceability in Cyberspace. 17 Berkeley Tech. L.J. (2002) 475-495.
6. Delio, M. (2004) "Spyware on My Machine? So What?" Wired News, December 06, 2004. Available at: :http://www.wired.com/news/technology/0,1282,65906,00.html
7. Dhamija, R., Tygar, J.D., and Hearst, M.: Why Phishing Works. Proceedings of the SIGCHI conference on Human factors in computing systems (2006) 581 - 590.
8. Earthlink: Earthlink spy audit: Results complied from Webroot's and Earthlink's Spy Audit programs (2005). http://www.earthlink.net/spyaudit/press.
9. Flesch, R.: A new readability yardstick. Journal of Applied Psychology, 32 (1948) 221-233.
10. Good, N., and Krekelberg, A.: Usability and privacy: a study of Kazaa P2P file-sharing. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2003) 137 – 144.
11. Good, N., Dhamija, R., Grossklags, J., Aronovitz, S., Thaw, D., Mulligan, D., and Konstan, J.: Stopping Spyware at the Gate: A User Study of Privacy, Notice and Spyware. Proceedings of the Symposium On Usable Privacy and Security (SOUPS 2005), Pittsburgh, PA, July 6-8 (2005) 43-52.
12. Good, N., Grossklags, J., Mulligan, D., and Konstan, J.: Noticing Notice: A large-scale experiment on the timing of software license agreements, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2007), San Jose, CA, April 28 - May 3 (2007).
13. Hillman, R.A.: Online Boilerplate: Would Mandatory Website Disclosure of E-Standard Terms Backfire. 104 Mich. L. Rev. (2006) 837-856.
14. Hochhauser, M.: Compliance v Communication. Clarity: Journal of the International Movement to simplify legal language, Vol. 50 (November 2003) 11-19.
15. Hochhauser, M.: Readability of HIPAA Privacy Notices (2003). Available at: http://benefitslink.com/articles/hipaareadability.pdf
16. Hochhauser, M.: Lost in the Fine Print: Readability of Financial Privacy Notices (2001). Available at: http://www.privacyrights.org/ar/GLB-Reading.htm
17. Jensen, C., and Potts, C.: Privacy policies as decision-making tools: An evaluation on online privacy notices. In CHI 2004 Connect: Conference Proceedings: April 24-29, Vienna Austria: Conference on Human Factors in Computing Systems (2004) 471–78. New York: Association for Computing Machinery.
18. Kessler, F.: Contracts of Adhesion – Some Thoughts About Freedom of Contract. 43 Columbia Law Review (1943) 629.
19. Kincaid, J., Fishburn, R. Jr., Rogers, R., and Chissom, B.: Derivation of New Readability Formulas for Navy Enlisted Personnel. CNTECHTRA Research Branch Report (1975) 8-75.
20. Kreuter, M.W., Brennan, L.K., Scharff, D.P., and Lukwago, S.N.: Do nutrition label readers eat healthier diets? Behavioral correlates of adults' use of food labels. American Journal of Preventive Medicine, Vol. 13, No. 4 (1997 Jul-Aug) 277-83.

21. Kucera, K., Plaisent, M., Bernard, P., and Maguiraga, L.: An empirical investigation of the prevalence of spyware in internet shareware and freeware distributions. Journal of Enterprise Information Management, Vol. 18, No. 6 (2005) 697 – 708.
22. Lewandowski, L.J., Codding, R.S., Kleinmann, A.E., and Tucker, K.L.: Assessment of Reading Rate in Postsecondary Students. Journal of Psychoeducational Assessment, Vol. 21, No. 2 (2003) 134-144.
23. Marotta-Wurgler, F.: Competition and the quality of standard form contracts: An empirical analysis of software license agreements. New York University working paper (2005).
24. Masson, M.E.J., and Waldron, M.A., Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting, Applied Cognitive Psychology, 8 (1994) 67-85.
25. Microsoft Corporation: Privacy Guidelines for Developing Software Products and Services. (October 10, 2006).
26. Overly, M., and Kalyvas, J.R.: Software Agreements Line by Line: A Detailed Look at Software Contracts and Licenses & How to Change Them to Fit Your Needs. Aspatore Books (2004).
27. PC Pitstop (2005) "It pays to read EULAs." Available at http://www.pcpitstop.com/spycheck/eula.asp (last accessed February 5, 2007)
28. Schechter, R.E.: The Unfairness of Click-On Software Licenses. 46 Wayne L. Rev. (2000) 1735-1803.
29. Slawson, W.D.: Standard Form Contracts and Democratic Control of Law Making Power. 84 Harvard Law Review (1971) 529-566.
30. Turow, J., Hoofnagle, C., Mulligan, D., Good, N., and Grossklags, J. (2006) Consumers & Privacy In the Coming Decade, Session on Communicating with Consumers in the Next Tech-ade - The Impact of Demographics and Shifting Consumer Attitudes, Public Hearings on Protecting Consumers in the Next Tech-ade, Federal Trade Commission, Washington D.C., November 6 - 8, 2006.
31. Vila, T., Greenstadt, R., and Molnar, D.: Why We Can't Be Bothered To Read Privacy Policies: Models of Privacy Economics as a Lemons Market. In Economics of Information Security, L. J. Camp and S. Lewis (Eds.), Springer-Verlag (2004) 143-153.
32. Whitten, A. and Tygar, J.D.: Why Johnny can't encrypt: A usability evaluation of PGP 5.0. Proceedings of the 8th USENIX Security Symposium (1999) 169 -- 184.
33. Yee, K.-P., and Sitaker, K.: Passpet: Convenient Password Management and Phishing Protection. In Proceedings of the Symposium on Usable Privacy and Security (2006) 32 – 43.

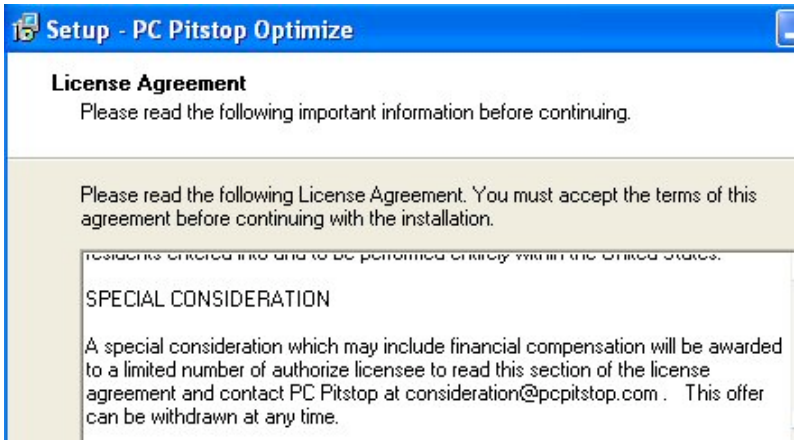
It Pays To Read License Agreements

by Larry Magid

I have a deal for you. In exchange for a free piece of software that helps you keep track of your passwords and other log on information, I'm going to install other programs on your PC that will track your web surfing and display advertising that pops-up on your screen. There will also be other types of ads on your computer based on information we collect.

Does that sound like a good deal to you? Well, if you're one of the many Windows users who have installed eWallet software from Gain Publishing that's exactly what you agreed to do. But you already know that because you read the End User License Agreement or "EULA" that was available prior to installing the program. You did read it right? Of course you did; before you could install the software you had to check a box certifying that you read the agreement. Legally speaking, that's the same thing as signing a contract with pen and ink.

OK, let's be honest. You didn't really read the EULA. How do I know? Because hardly anyone does. To prove that point, PC Pitstop included a clause in one of its own EULAs that promised anyone who read it, a "consideration" including money if they sent a note to an email address listed in the EULA. After four months and more than 3,000 downloads, one person finally wrote in. That person, by the way, got a check for $1,000 proving, at least for one person, that it really does pay to read EULAs.



Is anyone reading this? PC Pitstop offered a financial incentive in its EULA, and it took four months before anyone responded.

Although this is not a scientific sample, it does prove a point. People don't read EULAs. When we download and install software, we're usually in a hurry to take advantage of whatever it offers. That EULA is just one more thing to spend time on, and we're not just talking about a couple of

minutes. The December 2004 End User License agreement that accompanies eWallet and other programs from the GAIN network is 2,550 words long--that's seven printed pages.

To its credit, not everything in Gain's EULA is in legalese. You don't have to pay a Harvard law graduate $300 an hour to understand the first paragraph:

> "GAIN Publishing offers some of the most popular software available on the Internet free of charge ("GAIN-Supported Software") in exchange for your agreement to also install GAIN AdServer software ("GAIN"), which will display Pop-Up, Pop-Under, and other types of ads on your computer based on the information we collect as stated in this Privacy Statement. We refer to consumers who have GAIN on their system as 'Subscribers.' "

The rest of GAIN's EULA is also pretty clear. If you take the time to read it, you'll realize that you're giving the company permission to install software that "collects certain non-personally identifiable information about your Web surfing and computer usage." This, according to the agreement, "includes the URL addresses of the Web pages you view and how long you view Web pages; non-personally identifiable information on Web pages and forms including the searches you conduct on the Internet; your response to online ads; Zip code/postal code; country and city; standard web log information and system settings; what software is on the computer."

Gator's eWallet EULA is contained in this small box but it is 7 printed pages long.

So what's the harm in collecting "non-personally identifiable information?" After all, isn't that done all the time? Well there are certainly examples of such collection. Many legitimate web sites, for example, keep track of the number of visitors and where they go the site. This information is used to inform advertisers about a site's popularity and to give the site owners a better understanding of what parts of the site are doing well and what sites are now. Advertisers, of course, want to know how many people have viewed their ad as well as "clickthrough" rates and other information.

But there is a big difference between collecting non-personal information about what visitors are doing on your own site and tracking "the URL addresses of the (other) Web pages you view and how long you view Web pages."

Real live brick and mortar department stores, for example, do collect statistics about what sections of the store people are visiting, how long they spend there and what they buy. It's basic research. But imagine if you visited a store one day and they planted a bug on your person that followed you around to all the other stores you visited? While they were at it, they tracked your reading behavior, what TV shows you watched and maybe even who you talked with. They're not writing down your name, but they are following you around. Would this be legal? It might be, if they had you sign a contract specifically allowing it before they let you in the store.

And by the way, they're not just following you around. They're also getting in your way, making it harder for you to walk from place to place.

Making it harder to start your car and slowing it down once you start it. They might even cause you to stumble now and then. That's a lot like spyware and adware; it takes up hard drive space, memory, and other resources. Also, it can significantly degrade your Internet connection because spyware is going out over the Internet to get information to display and, in some cases, sending out information from your PC. In other words, it is using your resources--resources that you paid for.

GAIN is far from the only company that asks you to "sign" an agreement with serious implications. Marketscore, which bills itself as an Internet marketing research company, offers a service which, it claims, can speed up your web surfing and protect you from viruses. Whether or not it actually speeds up your service is debatable, but one thing is for sure. If you read Marketscore's privacy policy you'll learn that the company "monitors all of your Internet behavior, including both the normal web browsing you perform, and also the activity you may have through secure sessions, such as when filling a shopping basket or filling out an application form that may contain personal financial and health information."

The company says that it has all sorts of procedures in place to "restrict the third party's use of the information we provide." That's all well and good, but even if the company is as sincere and diligent as it says it is, things can change. And, if the company does decide to change its policy on how it handles personally identifiable information, it "will notify you by posting proposed changes to this Privacy Statement and on our web site." Those changes "will be effective immediately upon such posting.

And don't think that can't happen. Even if the current owners are committed to keeping information private, there is no guarantee that the company won't be sold. If it goes bankrupt, there is even the possibility of your information being sold to pay off creditors.

You may wonder whether these licenses are legal. Most of them do hold up in court as long as they are reasonably clear, according to Parry Aftab, an attorney specializing in Internet privacy and security law (www.aftab.com). "The courts have said that if you click on something saying 'I agree' then it's legal consent." There are exceptions however. "If it's not legally clear enough, you haven't given consent to anything because there is no meeting of the minds. It goes back to basis of contract law from 500 years ago. You have to both agree on what you are agreeing on." In other words, if the agreement is incomprehensible, it may be unenforceable, according to Aftab.

Another exception has to do with minors. "Kids," according to Aftab, "are under state contractual age which is sometimes 16 and sometimes 18. If the site requires the person to make an affirmative representation that they are over the age of 18, it may keep the company out of trouble but it's still not enforceable."

This is an important distinction because a lot of spyware and adware is bundled with programs that are marketed to children and teenagers.

The fact that a EULA might not be legally enforceable is of little solace because it is being enforced on you whether you like it or not. Once the program is installed on your PC, the damage is being done and it doesn't even matter if the contract that you or your child agreed to may be invalid. Simply by using your computer, you're upholding your part of that contract by giving up information.

Attorney Aftab says that even though the courts have ruled on the legality of EULAs, there are still some grey areas that need to be ironed out. And, of course, the courts are basing their rulings on current law. There are some in Congress who alarmed at the growth of spyware and a number of bills have been discussed that could impact the way these EULAs are written, agreed to and enforced.

In the mean time, it's "user beware." A click of the mouse, like a stroke of

a pen, can get you into a heap of trouble. Be careful, be aware and read those EULAs.

Larry Magid is a syndicated technology columnist and broadcaster for two decades, and contributes to CBS News, the New York Times, U.S. News & World Report and other publications.

Noticing Notice: A Large-Scale Experiment on the Timing of Software License Agreements

Nathaniel S. Good[1], Jens Grossklags[1], Deirdre K. Mulligan[2], Joseph A. Konstan[3]

[1]School of Information, UC Berkeley
102 South Hall, Berkeley, CA
{ngood,jensg}@ischool.berkeley.edu

[2]Boalt School of Law, UC Berkeley
346 Boalt Hall, Berkeley, CA
dmulligan@law.berkeley.edu

[3]Department of CS&E, University of Minnesota
200 Union Street SE, Minnesota, MN
konstan@cs.umn.edu

ABSTRACT

Spyware is an increasing problem. Interestingly, many programs carrying spyware honestly disclose the activities of the software, but users install the software anyway. We report on a study of software installation to assess the effectiveness of different notices for helping people make better decisions on which software to install. Our study of 222 users showed that providing a short summary notice, in addition to the End User License Agreement (EULA), before the installation reduced the number of software installations significantly. We also found that providing the short summary notice after installation led to a significant number of uninstalls. However, even with the short notices, many users installed the program and later expressed regret for doing so. These results, along with a detailed analysis of installation, regret, and survey data about user behaviors informs our recommendations to policymakers and designers for assessing the "adequacy" of consent in the context of software that exhibits behaviors associated with spyware.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces: *interaction styles, standardization, user-centered design*
J.4 [**Social and Behavioral Sciences**]: *psychology*
K.4.1 [**Computers and Society**]: Public Policy Issues – *privacy and regulation*
K.5.2 [**Legal Aspects of Computing**]: Governmental Issues – *regulation*

Author Keywords

Privacy, Security, Notice, End User License Agreement, Timing, Spyware

INTRODUCTION

Spyware, broadly defined, is fundamentally a challenge for HCI research as much as it is a technical one. At its core, fully disclosed spyware presents users with a trade-off: users gain the functionality of software in exchange for giving up private data, tolerating advertising messages, or both. While some examples of spyware are primarily fraudulent, others disclose the functionality of the spyware in a manner similar in form to the disclosure practices generally found in the software industry. Individual users have different needs and tolerances, and in the absence of public policy limiting their choices, the disclosures provided by spyware vendors would provide the basis for individuals to effectuate their policy choices in the marketplace. In an ideal market users would make decisions to install software, including spyware, where the trade was in their interest.[1] At times law constrains individual choice based on externalities or other market failures, or normative decisions about the values at issue. In the U.S., with respect to privacy and other issues commonly dealt with in mass-market software contracts there is little constraint on the substantive terms with respect to privacy, reliability, or security that can be presented to consumers.

It is not an ideal world. Study after study shows that people unwittingly install malicious or unwanted software [4][13]. It is easy to identify reasons for this disconnect. Certainly some of it is due to the software not disclosing what it does. Equally certainly, most users don't bother to read the lengthy and legalistic End User License Agreements (EULAs) or Privacy Agreements[27]. Even if the EULA is accurate, and the user reads it, the agreements may be so long and confusing as to prevent meaningful knowledge and consent. Some users may mistakenly believe that their operating system, antivirus software, or other precautions will protect them. But there are other reasons. Users may be too eager to use the software to be concerned about spyware—at least at that moment. And users who have just selected the action to install may be too committed to that

[1] We want to emphasize that we do not wish to downplay the problems associated with malicious software that fails to disclose its true behavior; as we discuss below, there are legal remedies for such deception. However, in this work we restrict our attention on supporting users in making correct decisions when faced with disclosed trade-offs.

action to suddenly change course (just as Norman suggests that immediate confirmation of deleting a file is useless or worse [22]).

Our work builds upon a previous study of 31 subjects that showed that short summary notices, as a supplement to EULAs, have promise in helping users identify which software they feel comfortable installing [18]. We studied 222 subjects, observing their installation behavior in one of three information conditions: 1) an ordinary EULA, 2) a short summary notice before installation additional to the EULA, 3) a short summary notice (with an opportunity to uninstall) immediately after installation additional to customary EULA. We also surveyed users about their behavior in both computer use, more general use of legal documents (e.g., signing such documents without reading them) and actions regarding several online risks. The results of this study provide significant opportunities for designing software systems that better support users in protecting themselves against unwanted spyware, and might even generalize to a broader set of "in-the-moment" decisions.

The fact that users do not read EULAs may appear to be a truth in need of little proof. But for the current policy debates this finding is important. The courts, absent fraud or unconscionability, largely hold individuals responsible for the terms of legal agreements regardless of this reality. However, because a defining element of spyware is the context—in particular the consent experience—around its acquisition state and federal regulatory agencies and the private sector are developing new policy that establishes procedures aimed at providing an "adequate" consent experience. While reflective of HCI in some respects there has been little transfer of knowledge or prior research to determine the likely effect of these enhanced procedural rules. Thus, there is much to be gained from a cross-disciplinary conversation around the HCI contributions toward these reforms.

RELATED WORK

As some experimental research demonstrates, users stated privacy preferences do not always align with their behavior [2][23]. For small monetary gains (e.g., a free program) or product recommendations, users are willing to trade off their privacy and/or security [2][23]. Moreover, users are more likely to discount future privacy/security losses if presented with an immediate discount on a product [2]. Notices often fail to dissuade individuals from making decisions that contradict their own clearly stated preferences [23][27].

HCI practitioners have been concerned about privacy concerns on the web in general, and recent work in HCISec (HCI and security)[2] has been concerned with cookie management [15][16], spyware [24], phishing [12], as well as online privacy notices [10][20] and incidental privacy issues with web browsing [19] and filesharing [17].

HCI practitioners are uniquely positioned to contribute to the conversation on designing more effective notices by, for example, improving on timing and visualization. However, contributions to notice design are challenging because users are simply trained to ignore consent documents. This challenge of attracting attention to important events is not a new one in HCI. A number of researchers are studying the effects of notification systems in computing. They examine the nature of interruptions and people's cognitive responses to work-disruptive influences. Notification systems commonly use visualization techniques to increase information availability while limiting loss of users' focus on primary tasks [11][26][25]. From control room and cockpit indicators to desktop notifiers, substantial research has been devoted to identifying visual and auditory displays that attract attention and to designing interaction sequences that prevent automatic dismissal of information.

Work on privacy notices for web sites spans several different areas. P3P[3] and privacy bird[4] were popular efforts to inform users about a Web site's privacy preferences, as well as to give users control over the types of information exchanged with interaction partners. The P3P design called for web designers and companies to provide easy to read, privacy statements in a standard format that is usable, for example, by P3P-aware browsers to communicate this information to the end user.

Recent research by Karat *et al.* [21] aims at providing design methodologies and tools to assist in the creation of more usable privacy policies that can be verified by automated techniques. A main objective is to achieve consistency between notices, as well as better compliance with emerging privacy standards.

Friedman *et al.* [15] work towards improving interfaces for informed consent through implementing value sensitive design methodologies. Their design approach targets users' comprehension and suggests methods to facilitate consent between the user and the application based on shared knowledge and trust.

Early work with privacy in HCI was focused on the notion of feedback and control, introduced by Bellotti and Sellen [6]. The concept of feedback and control suggests that users are given ample feedback on the actions a system is taking, whether it is video taping someone or sending information to a third party, and that users are given adequate means of

[2] See the HCISec Bibliography for the most important contributions to this field. http://www.gaudior.net/alma/biblio.html

[3] W3C Platform for Privacy Preferences Initiative. Platform for Privacy Preferences (P3P) Project. http://www.w3.org/P3P.

[4] http://www.privacybird.com/

controlling the flow of this information, such as being able to turn off recording or specify which information is sent to third parties.

The concept of feedback and control is related to the legal explanations of informed consent. Informed consent emphasizes that end users must receive notice (signs, readable language, etc.), and must be able to provide consent to the action. Essentially, to use a term in contracts, both parties have to establish a "meeting of the minds" where they are both consenting to and are agreeing to the same shared set of knowledge.

Courts typically enforce EULAs. Courts have enforced shrinkwrap agreements that purport to bind users to EULA terms that appear on software packaging simply because the user opened the package.[5] Courts typically find that installing or using the software is sufficient to establish acceptance of EULA terms even when users are not required to click "I Agree."[6]

There is a growing body of literature questioning the courts generally unquestioning and superficial review of the context of contract formation, specifically around notice and consent. Within the legal literature and policy circles questions about the adequacy of consent, in particular the form, content, presentation and timing of disclosures in relation to programs that exhibit behaviors associated with spyware, are being raised [14][7][8][28][29][5][3][1].The Federal Trade Commission and the State Attorney Generals are challenging the courts' laissez-faire attitude towards contract formation demanding heightened procedural, and to a lesser extent substantive, protections for contract formation in the context of spyware enforcement actions. The rules they are establishing in this context will likely inform the agencies, and in time the courts, views on contract formation with respect to downloadable software in general.

Given the connection between the privacy and security decisions of individual users and the overall security of the network, the questions about externalities bear particular attention. If we are to rely on a private contractual approach to privacy in the U.S. we need to make sure that private choices don't undermine collective security and that users are capable of understanding and making the privacy and security decisions necessary to protect their interests.

Our report adds to the growing literature on HCI and security/privacy but also makes important connections to the ongoing legal and policy reforms around notice and consent.

---

[5] *Bowers v. Baystate Technologies, Inc.* , 320 F.3d 1317 (Fed. Cir. 2003).

[6] See Tarra Zynda, Note, *Ticketmaster Corp. v. Tickets.com, Inc.: Preserving Minimum Requirements of Contract on the Internet*, 19 Berkeley Tech. L.J. 495, 504-505 (2004).

EXPERIMENTAL DESIGN

Experimental Setup

Our experimental setup consisted of an experimental portion, followed by two surveys. Subjects were given a unique number, and sheet outlining the basic scenario of the experiment. All of the experiments and surveys were done by each subject independently on a computer located in a laboratory with dividers. As the user passed each portion of the experiment, the application would record the actions and provide the next portion of the experiment. We describe the details of each portion of the experiment below.

Experimental Framework

The experimental portion of our framework was designed to mimic the experience of installing software applications, but also allows us to modify the notice and consent process encountered. Previous experiments showed us that pop-up windows and warnings are quickly ignored by users who are accustomed to click through them. In order to have users "notice" the notice conditions, we decided to build them into the install experience.

We constructed a windows application in C# that would not only depict the installation process as realistic as possible, but also log all user actions (e.g., buttons clicked, time per screen) during the study. Additionally, the application we constructed would provide a launching pad that could dynamically configure each subject's experience based on their user number we provided at the beginning of the experiment. Users were given a user id, which was matched up against a list of acceptable identifiers and associated with a treatment and a counter-balanced program ordering.

We constructed two surveys, which could be accessed from the application launching pad. After the experimental portion was completed, users could click on the survey buttons to answer each respective survey. When both surveys were completed, the user was returned to the launching pad, and told that the experiment was completed.

Notice Treatments

Our design consisted of three notice conditions: two treatments with customized short notices that included abbreviated EULAs for the programs plus the original EULA (all without brand information) and one control condition that only consisted out of the original EULA (again all without brand information):

PRE - Short notice before installation presented on the Install Option Screen plus the original EULA on the EULA screen;

POST - Short notice after installation on the Post Install Warning Screen plus the original EULA on EULA screen; and

CONTROL/None - No short notice at all, but with the original EULA on the EULA screen.

Each notice condition was integrated into three programs with consistency maintained for each portion of the controlled experiment by providing similar screens, but changing the content of the information in key screens for the different programs. Figure 1 below details the screens involved in the installation process for each user.

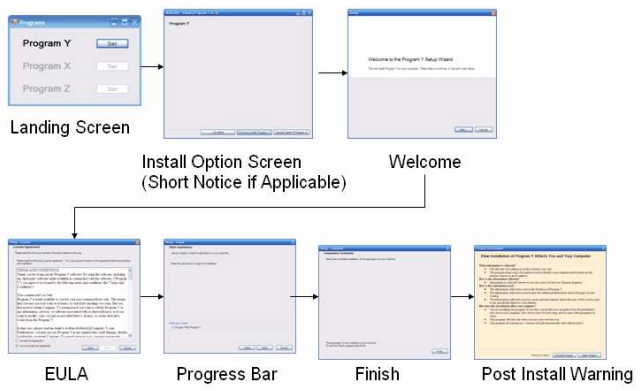


Figure 1 Process of installation screens in experiment

The Post Install Warning screen only occurs when a user is in the post notice condition. However, the Install Option Screen displays the Short Notice in the Pre condition, but appears also in the other two treatment conditions without specific information. At any time, a user may cancel the installation and return to the landing screen to start with the next program. Additionally, users may move back and forth between screens as in typical installation programs by hitting the back key.

We selected programs from our previous study [18] to facilitate comparability of the results and user experience. We chose a browser toolbar, a weather information service and a file sharing application. For the experiment each brand name was removed and replaced with a generic title. The program titles and descriptions are listed below:

- Program X – Weather Information Program
- Program Y – Browser Toolbar
- Program Z – File Sharing Program

To summarize, we ran a 3x3 mixed methods study, consisting of 3 *between*-subjects factors and 3 *within*-subjects factors. The between subjects factor were the notice conditions (None/Control, Pre, Post) and the within-subjects factors were the programs (Filesharing, Weather Service and Browser Toolbar). Within subjects factors were counter-balanced within the population. We want to add that it is of methodological interest to us to understand the relative strengths and weaknesses between the small scale user study and the current large experiment with hundreds of users.

Notice Construction

Our short notices were designed by distilling the long EULAs from three programs included in our study. We used the same short notices as we constructed in a previous study. Each notice condition that included a short notice (Pre and Post) had the same text for a specific program. The only difference between each treatment was the timing where each notice was shown. Examples of how the short notice for Program Z would appear during the experiment in the pre-notice and post-notice treatments are presented in Figure 2 and Figure 3, respectively.

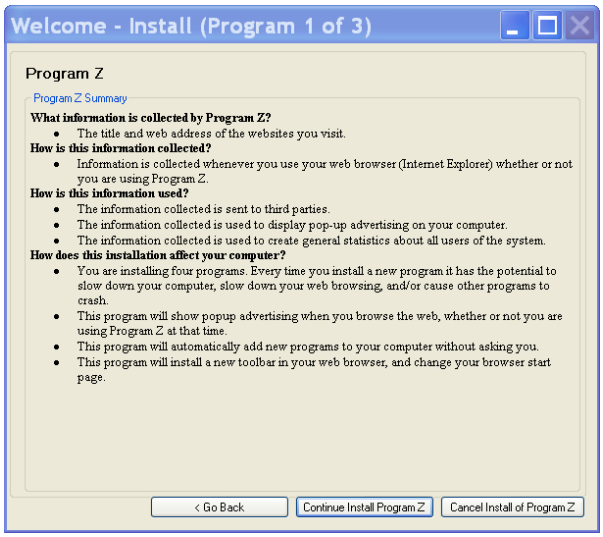


Figure 2 Pre-installation short-notice (would appear on Install Option Screen in Pre-Notice treatment)

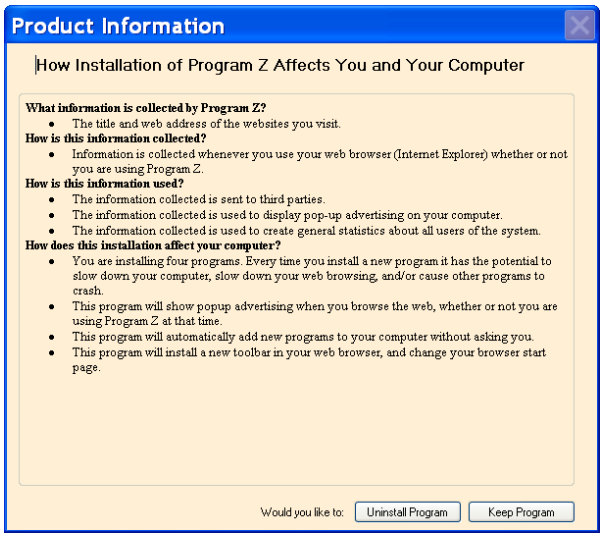


Figure 3 Post-installation short-notice (would appear on Post Install Warning Screen in Post-Notice treatment)

Surveys

Our study included two surveys that were presented to the user after the installation experiment. The surveys included different question types, for example, open ended, Likert

scales, and simple yes/no questions. The purpose of the survey was to understand the subjects' concerns regarding a representative selection of online risks and stated behaviors related to software notices, as well as to determine how the users perceived the programs in the experiment as well as whether or not they regretted the actions they performed in the experiment after being provided with an (additional) chance to review the short notice.[7]

The first survey consisted of demographic and behavioral information, while the second survey consisted of questions regarding the experimental experience. In total, we anticipated that the two surveys plus the experiment section could be passed by the average subject in about one hour.

Table I Self-reported behavior regarding online risks

	Total	
	Mean	Std. Deviation
Try Functionality	4.41	2.64
Research on Web	5.14	2.64
Once Installed wont remove	2.24	1.72
Accept Popups for free stuff	3.36	2.42
Install programs that look interesting	3.55	2.35
Install only if I know exactly what it does	6.41	2.06

RESULTS

Subjects

240 subjects participated in the study, of which we were forced to remove some entries due to missing data leaving us with 222. Subjects were paid $20 for their participation, and were recruited by a university service with access to a subject pool of several thousand students. Our subjects are divided into three treatment groups: 64 users in the control condition, 80 in the pre-notice condition and 78 in the post-notice condition. We used chi-squared to analyze differences between the discrete variables of install and regret, and ANOVA to analyze the differences between the continuous variable of time.

64.2% percent of our subjects were female. 39.5% indicated their age as less than 20 years-old. An additional 57.7% were between the ages of 20 and 25. The dataset also includes a small group of 2.7% over 25 years of age. On average we had a very computer-experienced group of users. For example, 85.2% stated that they maintained their home computer themselves.

[7] Of course, subjects in the control group and those in the post notice treatment that canceled early had not seen the notice before.

Attitudes towards online risks

We asked users to rate concerns on a scale of 1-9, with higher numbers expressing more concern. Subjects expressed high concern towards 5 different risk types and nuisances often encountered in online interactions: identity theft, spyware, viruses, pop-up advertisements, and privacy intrusions.

Surprisingly, our young subject pool was somewhat less alarmed about identity theft and privacy compromises, compared to being subject to spyware attacks and pop-up advertisements. Possible damages caused by viruses topped the list.

We employed *k*-means multivariate clustering techniques to classify subjects according to their risk attitudes. Hierarchical clustering (single linkage) preceded the data analysis. We selected the best partitioning using the Calinski-Harabasz criterion [9]. We derived two distinct clusters: a first group with a substantially higher degree of unease about online risks along all measured dimensions (62.2%) and a second less worried and comparatively smaller group (37.8%).

Self-reported behavior regarding online risks

Our subjects are forthcoming about their good computer hygiene practices (see Table I; values are reported on scale from 1-9). On average, while they are interested in trying new content, they report to somewhat often research programs on the Web before using them, claim that they only install program when they are well informed about them, and report that they hardly leave them installed if found undesirable. They rarely agree with the statement that free software in exchange for intrusive advertisements is acceptable. Only few wholeheartedly admit that they would frequently download and install programs that look interesting.

Self-reported reading practices for legal documents

Only very few users reported reading EULAs often and thoroughly when they encounter them (1.4%). Members of a larger group categorize themselves as those who often read parts of the agreement or browse contents (24.8%). However, 66.2% admit to rarely reading or browsing the contents of EULAs, and 7.7% indicated that they have not noticed these agreements in the past or have never read them.

Table II Self-reported reading practices for legal documents

	Total	
	Mean	Std. Deviation
Financial Privacy Notices	5.97	2.78
Read Web Privacy Notices	4.24	2.28
Read Shrinkwrap Licenses	3.81	2.25

In Table II we report on subjects' reading behavior for other important notices (values are reported on a scale from 1-9 from "never read" to "always read". Web privacy notices and shrinkwrap licenses are read less frequently in comparison to, for example, financial privacy notices. Less related to our field of investigation, we found that food nutrition labels and credit card statements are read almost twice as often as shrinkwrap licenses by our subject group (means of 6.8 and 7.3, respectively).

Table III Occurrences of canceled installations for each screen

Treatment	Cancellation Screen	Program X	Program Y	Program Z
None	Install option	15.6%	9.4%	9.4%
	Welcome	15.8%	0.0%	0.0%
	EULA	4.7%	0.0%	4.7%
	Install Progress	0.0%	0.0%	0.0%
	TOTAL CANCELED	32.1%	9.4%	9.4%
Pre	Install option	69.2%	28.2%	69.2%
	Welcome	1.2%	0.0%	0.0%
	EULA	0.0%	0.0%	0.0%
	Install Progress	0.0%	0.0%	0.0%
	TOTAL CANCELED	70.4%	28.2%	69.2%
Post	Install option	13.8%	2.5%	11.3%
	Welcome	0.0%	0.0%	1.3%
	EULA	11.3%	1.3%	5.0%
	Install Progress	1.3%	0.0%	0.0%
	Post Install Warning	51.3%	28.6%	58.8%
	TOTAL CANCELED	77.7	32.4%	76.4%

Behavior in the experiment: Installation

Chi-squared tests showed that both notice conditions had significantly lower instances of installation than the control condition ($p<.001$). This effect is robust independent of whether the unit of investigation are the whole treatment groups or individual programs ($p<.001$)

This result demonstrates that the short notice treatments had a significant behavioral impact on subjects. It also supports what we have seen in previous studies and have observed in the field - users that are presented with the omnipresent overly long and complex presentation of EULAs are prone to installing applications more often. As we saw in our previous small-scale ecological user study[18], the toolbar application was most frequently installed, followed by the file sharing application and finally the weather information service. For the control treatment we attribute the difference between programs to a combination of two effects: preference for a program type (e.g., toolbar vs. filesharing client) vs. desirability of contractual terms. Users seem to be able to discriminate between programs even without additional cues such as brand information and familiar user interface design. The results we present in following sections rest on the variation of treatment variables.

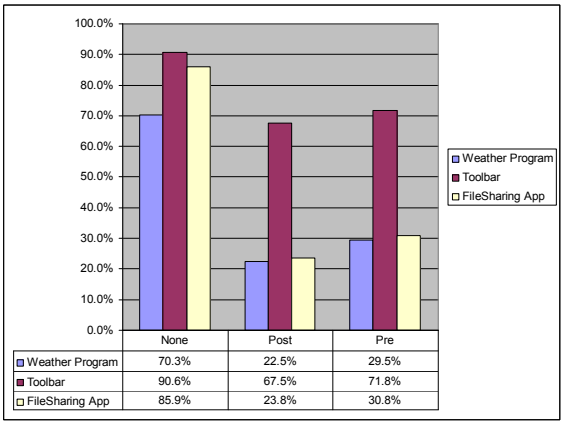


Figure 4 Programs installed by users

Behavior in the experiment: Cancellations/Uninstalls

Interestingly, of those that canceled the installation in the control and post-notice treatments the preferred action (always more than 50%) was to leave the program immediately on the very first screen (that is Install option screen). Only between 0% and 42.8% of subjects ended up visiting the EULA screen. This result is of importance for program developers interested in increasing their installed base. It seems that although many people might enter an installation they often will immediately leave even without gathering further information.

Table III reports the percentage of all individuals that canceled on particular screens for the three different treatments. It demonstrates the dominance of the short notice screens (Install Option for pre-notice and Post Install Warning for post-notice) in comparison to the EULA screen.

Note that the short notices contained information from the original EULA, however, presented in a unified and abbreviated format. Therefore, the data for the post notice treatment clearly demonstrates the inadequacy of the long and complex EULA. All subjects that canceled on the Post Install Warning screen have seen the original EULA on the screens they passed to reach the warning screen. But only on the short notice screen they decided to cancel.

Behavior in the experiment: Timing and reading notices

Pre-notice

In our experiment there is only one screen per program that was visited by everybody independent of treatment condition and whether they installed or canceled the installation at some point of the process. This is Install Option Screen. In Table III we have already shown that

many subjects decided to cancel at this point of installation. It, however, is also interesting to note that there are significant differences in time spent by the subjects at this point of the experiment.

Table IV Time in sec for Install Option Screen (Program X)

Treatment	Installation completed	Mean time (sec)
None	Yes	1.9
	No	13.5
	Significance result	p<0.0037
Pre	Yes	3.4
	No	59.2
	Significance result	p<0.0000
Post	Yes	0.5
	No	8.3
	Significance result	p<0.0002

First, individuals that decided not to install a program are significantly slower than installers for all three treatments (see Table IV).[8]

Second, not surprisingly, individuals that do not install the programs spent more time on the Install Option Screen if they are members of the pre-notice treatment group compared to those in the control or post-notice treatments (with at least $p<0.05$ for all programs, but stronger for most). They are obviously paying attention to the notice.

Third, for those that installed the programs there is no such statistically significant difference noticeable. It should be added that installers in the pre-notice condition also passed quicker through the Install EULA screen than the control and post-notice group; that is, they did not pass quickly through the pre-notice screen with the intention of reading the actual EULA carefully.[9] Accordingly, they are consistently quicker than others. Or, to phrase it differently, they are ignorant towards notices.

We conclude that one main difference between installers and those that decline the program offerings is the time they spent deliberating at the start of the program installation. Even for the two treatments where no pre-notice was displayed on the Install Option Screen non-installers are considerably slower. One interpretation is that more deliberate individuals take additional time to study the short notice.

[8] Result does not hold for program y in the control treatment. However, the group of non-installers is extremely small for this program which makes it an outlier case.

[9] The group mean comparison test is significant for program x and z ($p<0.005$). For program y the differences have the expected direction, however, are not significant.

Table V Comparison of pre and post-notice reading time

Program installed	Cancellation Screen	Time for Program X (in sec)	Time for Program Y (in sec)	Time for Program Z (in sec)
Yes	Mean Pre-notice	3.4	8.5	0.0
	Mean Post-notice	14.5	35.6	15.6
	Significance	p=0.01	p<0.0000	p<0.0000
No	Mean Pre-notice	59.2	81.0	44.2
	Mean Post-notice	30.2	37.1	30.8
	Significance	P<0.0000	p<0.0006	p<0.005

Post-notice

Finding a natural comparison standard for the reading time in the post-notice treatment is more difficult since the Post Install Warning Screen appeared only in this treatment. We believe that comparing pre-notice and post-notice reading time is the most natural approach.

Table V strongly supports the finding that individuals who eventually installed a program passed slower through the Post Install Warning screen compared with the Install Option Screen. However, subjects that did not install a particular program took more time reading the pre-notice.

Assuming that increased reading times improve consumer decision-making this demonstrates a conundrum. On the one hand, we observed for the pre-notice that only non-installers read the notice (or even become aware of the notice). This is different for the post-notice where reading times even for subjects that completed installations are significantly distinct from zero. On the other hand, if subjects became aware of the pre-notice they spent a considerably longer amount of time absorbing the information which usually led to a cancellation of the installation process.

From a behavioral point of view it appears that subjects are very much willing to cancel an installation at the beginning of the process if they are adequately informed about the terms of the transaction. However, the risk is that they are too involved in the flow of conducting the necessary installation steps in order to notice the additional warning terms.

In contrast, the post-notice serves to slow down a majority of the individuals. It seems that subjects at this time of the installation process are able to notice and digest further information; that is, they have left the flow state. However, reading the notice does not necessarily result in the uninstallation of the program. One interpretation is that subjects have made an emotional investment into the program installed. Or they might be interested in trying the program even if they dislike its terms since it is already

installed at this point. As a result, not all users are willing to reverse their decisions. Another potential explanation is that subjects who keep the program feel that it adequately reflects their preferences for a consumer program. The distinction between these hypotheses is left for future work, but we present further evidence on this question in the next section.

Correlation Survey and Experiment: Regret

In Post Survey 2, we showed all subjects the short notice for each program, and asked them if they would install the program described in the short notice or not. We used this measure as a means of calculating the user's regret. In the post case, we calculated regret after users had seen the post notice, and had made the decision to keep or uninstall the program. We determined that users would have two types of regret, regret that they installed a program (and would like to remove it) and regret that they chose not to install a program (and they would like to install it). The second case we expected to be less common.

Overall regret was high for programs that users installed. Only in the case of program Y, the toolbar, do we see that over 50% of the users were happy with their installation choices. Regret is still very high for the programs that users consider the least usable, namely program X. In the best case, the pre notice, 70% of the users still regret installing the application. Although short notices may help, there is still much room for improvement.

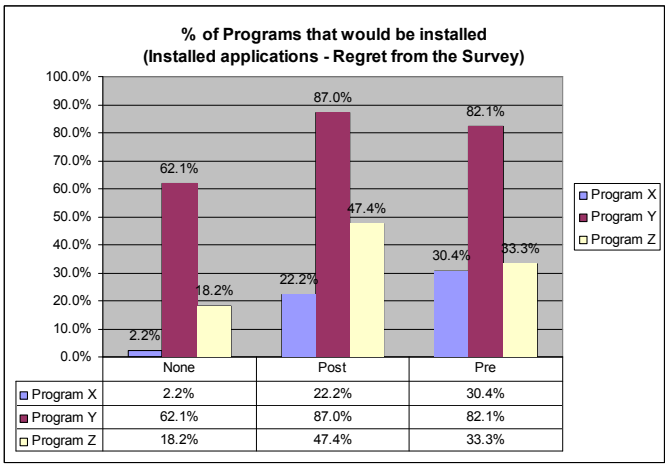


Figure 5 Graph of percent of users who were happy with their installation choices

Reading and Regret

The high regret we saw lead us to wonder as to whether regret was different in cases where users actually spent more time on the notice screens, either the short notices and/or the EULAs themselves. We decided to analyze cases where users had made it to at least one notice screen (EULA and/or short notice), and compared the time that users spent on each screen to their stated regret. By using time as an implicit measure of reading we were able to determine if the notice reading time had an effect on regret.

Overall, we found a strong difference between the control treatment and the pre treatment in terms of regret ($p<.05$). We found that in most cases, users who spend more time reading the short notices in both the pre and post conditions had less regret. Details of regret in each treatment condition are given below.

Regret and the Short Notice Condition

Not surprisingly, in the short notice case, we saw a significantly lower level of regret ($p<.05$) for users who spent more time reading. Users who had less regret spent on average 20 – 30 seconds more per notice, approximately double the average time in most cases.

Regret and the Post Notice Condition

At a first glance, the post notice condition seemed similar to the control condition. Because the post condition comes after users look at the EULAs, the combined post notice and EULA time may be dominated by behaviors we see in the EULA only case. For this reason, we ran another case where we separated the EULA time from the post notice time, and looked at how time on the post notice related to regret.

In this case, we found that the post noticed behaved similarly to the short notice, but the effect was not as strong across all program types. Users who spent more time on the post notice had significantly lower regret for the least desirable program, program X, but not across the other program types. We found this to be interesting because users in the post notice condition had to decide to keep the program or uninstall it. It was also the last notice that the users saw, so we were surprised that post notice had still had high cases of regret. It may be the case that for some users, after they have committed to an installation they feel they have some investment in the program or momentum and would like to continue to install. In the future, we plan to use more sophisticated modeling techniques to derive more comprehensive and powerful explanations of these kinds of user behavior.

Regret and the Control Condition

We found that in the control condition, users had a high amount of regret, whether they spent time reading the EULA or not across most programs. The control case, users had significantly higher values of regret ($p<.001$) for programs x and z. and a moderately higher value than the pre condition in program y ($p<.10$). In some cases, users who read more had significantly higher cases of regret than those that read less ($p<.05$). There is evidence that the EULAs could be confusing and misleading [18]. One user from our survey said she was "befuddled by the language" another mentioned that "they're often very very long [and] not easy reading, either." Users also mentioned that if they

have concerns, they look for certain terms such as "pop ups" or other things that may adversely affect their computer. One user mentioned "[I would look at them] if [they were] precise and clear, and the agreement is short, so it's not too time-consuming. And the words are keywords, so I could just browse it very quickly at a glance. Besides, I will read it when the program alerts me the bad consequences of not reading the agreements." This result emphasizes the need to have common terms across software licenses, especially for cases that deal with issues users are generally concerned about (performance, pop-ups, monitoring, etc).

Summary of Reading & Regret Results

From our analysis of reading and regret, we have learned that if we can get users to spend time reading the notices, they may experience significantly less regret. In this case, it is important for HCI practitioners to determine what can be done in terms of interfaces to get more users to slow down and read notices.

DISCUSSION

Observations and Implications

Four observations have very clear design implications for software installation systems and for efforts in the public and private sector to make the consent experience meaningful.

First, the experiment validates the use of short summary notices as a mechanism for reducing the installation of unwanted software. There are many ways in which such notices could be provided, ranging from legal solutions (where the use of such notices could be necessary for documenting informed consent) to technical and business ones (e.g., the creation of subscription services that provide such "installation reviews" for users). Efforts at state and federal regulatory agencies to simplify and highlight core software behaviors and draw attention to particular terms appear promising based on our research.

Second, the effectiveness of post-install notices suggests an alternative strategy for reducing unwanted spyware – delaying the actual irrevocable installation of software. Users might be well-served by systems that "pretend" to install software, then warn about the consequences before really completing the installation. (This approach could be a variant of the "to finish installation, you must reboot" barrier.) Or in some cases, it may even be worth preventing immediate use of software to provide a period of reflection. Efforts in the private sector to create virtual machines or sand boxes of sort that would allow consumers to test out software without allowing it fully onto their machine appear promising.

Third, from the regret data in the pre- and post-experimental conditions, we know that substantial regret exists even with these short notices. Accordingly, it is important to continue to explore other remedies to the spyware problem, including legal protections, technical protections, and interfaces that intercept the problem before the installation decision is made. Indeed, Google's warning that forces confirmation from people following a link to certain web sites (primarily cracking-related) could be adapted, or better yet, tools tied to ratings services could label links to software with indicators of the negative consequences of its use. It also points to the need for users to be provided with simple means to restore their machines to pre-installation state. Recent spyware enforcement actions have focused on this requirement.

In general, our research conclusions support the additional procedural constraints the FTC and State AGs are placing on spyware vendors. Given the contextual and individually subjective decisions about what is spyware our research would support the expansion of these protections to a broader range of software installations. The question is how broad a range is appropriate given that enhanced notices about everything is likely to undermine the utility and effectiveness of these "express consent" procedures where users face the greatest risks.

Finally, the presence of individual differences in reading behavior and other behaviors correlated with spyware installation suggests that personalized solutions have promise. Some users are well-served by the current system, or would be with short summary notices. Others seem likely to ignore such notices and might be willing to accept more restrictions on their installation (e.g., longer delays sequences of confirmations, or approval from another individual) in order to reduce their own risk and later regret. There are many paths to explore in this direction. To the extent that the overall security of the network is influenced by the decisions of users some of who ignore the processes established to engage them in good decision making, it is worth asking whether some private choices to tolerate spyware—particularly spyware that creates opportunities for others to remotely assume control of computers—are just too damaging to the network to remain in the realm of private choice.

FUTURE WORK

The research reported here opens as many questions as it resolves. It is our goal in future work to better understand the factors that lead individuals to install spyware, and how those factors vary in different demographic groups (including older users) and in different situations. We recognize the limitations of a laboratory study, and are hopeful that it will be possible to conduct more extensive studies of software installation, and more general questions of a personal computer's life cycle, on computers installed in individual homes and offices. Further, we believe that appropriate notice can help reduce installation of unwanted spyware, but also recognize that "appropriate" may vary by individual. We would particularly welcome further research into possible negative effects of excessively long

and impenetrable EULAs, and other explorations into interfaces for more effectively presenting the relevant information to users for meaningful, informed, consent.

ACKNOWLEDGEMENTS

We wish to thank Microsoft for funding our work. We also appreciate the help of Tye Rattenbury, Frances Tong and Xin Wang concerning statistical analysis. We are greatly indebted to Susheel Daswani for constructing the experimental framework. Finally, we thank Becca Shortle, Chris J. Hoofnagle, Ira Rubenstein and the anonymous reviewers for their valuable feedback and suggestions. This work is supported in part by the National Science Foundation under ITR award ANI-0331659.

REFERENCES

1. Abrams, M., M. P. Eisenhauer, and L.J. Sotto Letter to Federal Trade Commission. March 29, 2004. Re: alternative forms of privacy notices, project no. P034815. Hunton & Williams: The Center for Information Policy Leadership.
2. Acquisti, A., and J. Grossklags. 2005. Privacy and rationality in individual decision making. IEEE Security & Privacy 3(1): 26–33.
3. Anti Spyware Coalition, Anti Spyware Coalition Definitions and Supporting Documents, Working Report (June 29, 2006), available at http://www.antispywarecoalition.org/documents/documents/ASCDefinitionsWorkingReport20060622.pdf
4. AOL and National Cyber Security Alliance. 2004. AOL/NCSA online safety study, (October). http://www.security.iia.net.au/downloads/safety_study_v04.pdf
5. Bellia, P. L. Spyware and the Limits of Surveillance Law, 20 Berkeley Tech. L.J. 1283 (2005)
6. Bellotti, V. and A. Sellen. 1993. Design for Privacy in Ubiquitous Computing Environments. In Proceedings of The Third European Conference on Computer Supported Cooperative Work (ECSCW'93). Milan, Italy: Kluwer Academic Publishers.
7. Blanke, J. M. "Robust Notice" and "informed Consent:" the Keys to Successful Spyware Legilsation, 7 Coum. Sci & Tech. L. Rev. 2 (2006)
8. Buenaventura, M. A. Teaching a Man to Fish: Why National Legislation Anchored in Notice and Consent Provisions is the Most Effective Solution to the Spyware Problem, 13 Rich. J.L. & Tech. 1 (2006)
9. Calinski, R.B. and Harabasz, J. 1974. "A Dendrite Method for Cluster Analysis," Comm. in Statistics, vol. 3, pp. 1–27.
10. Cranor, L.F., J. Reagle, and M. S. Ackerman.1999. Beyond concern: Understanding net users' attitudes about online privacy. In Ingo Vogelsang and Benjamin M. Compaine, eds. The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy. Cambridge, Massachusetts: The MIT Press, p. 47-70
11. Cutrell, E., M. Czerwinski, and E. Horvitz. 2001. Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. Proceedings of Interact 2001: IFIP Conference on Human-Computer Interaction, Tokyo, Japan. http://research.microsoft.com/~cutrell/interact2001messaging.pdf.
12. Dhamija, R., Tygar, J. D., and Hearst, M. 2006. Why phishing works. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). ACM Press, New York, NY, 581-590.
13. Earthlink. 2005. Earthlink spy audit: Results complied from Webroot's and Earthlink's Spy Audit programs, http://www.earthlink.net/spyaudit/press.
14. Federal Trade Commission, Monitoring Software on Your PC: Spyware, Adware, and Other Software, http:www.ftc.gov/os/2005/03/050307spywarerpt.pdf
15. Friedman, B., Howe, D., and Felten, E. 2002. Informed Consent in the Mozilla Browser: Implementing Value Sensitive Design. In Proceedings of the 35th Annual Hawaii international Conference on System Sciences (Hicss'02)-Volume 8 - Volume 8 (January 07 - 10, 2002). HICSS. IEEE Computer Society, Washington, DC, 247.
16. Goecks, J. and Mynatt., E.D. 2005. Supporting Privacy Management via Community Experience and Expertise, Proceedings of 2005 Conference on Communities and Technology, p. 397-418.
17. Good, N. S. and Krekelberg, A. 2003. Usability and privacy: a study of KaZaA P2P file-sharing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003). CHI '03. ACM Press, New York, NY, 137-144.
18. Good, N., Dhamija, R., Grossklags, J., Thaw, D., Aronowitz, S., Mulligan, D., and Konstan, J. 2005. Stopping spyware at the gate: a user study of privacy, notice and spyware. In Proceedings of the 2005 Symposium on Usable Privacy and Security (Pittsburgh, Pennsylvania, July 06 - 08, 2005). SOUPS '05, vol. 93. ACM Press, New York, NY, 43-52
19. Hawkey, K. and Inkpen, K. M. 2006. Keeping up appearances: understanding the dimensions of incidental information privacy. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). ACM Press, New York, NY.
20. Jensen, C., and C. Potts. 2004. Privacy policies as decision-making tools: An evaluation on online privacy notices. In CHI 2004 Connect: Conference Proceedings: April 24-29, Vienna Austria: Conference on Human Factors in Computing Systems 6(1): 471–78. New York: Association for Computing Machinery.
21. Karat, C., Karat, J., Brodie, C., and Feng, J. 2006. Evaluating interfaces for privacy policy rule authoring. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM Press, New York, NY, 83-92.
22. Norman, D. A. The Design of Everyday Things, 1988.
23. Spiekermann, S., J. Grossklags, and B. Berendt. 2001. E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior. In Proceedings of the Third ACM Conference on Electronic Commerce, Association for Computing Machinery (ACM EC'01), 38-47. New York: ACM Press.
24. Stiegler, M., Karp, A. H., Yee, K., Close, T., and Miller, M. S. 2006. Polaris: virus-safe computing for Windows XP. Commun. ACM 49, 9 (Sep. 2006), 83-88.
25. Trafton, J. G., E. M. Altmann, D. P. Brock, and F. E. Mintz. 2003. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. International Journal of Human Computer Studies 58(4): 583–603.
26. Van Dantzich, M., R. Daniel, E. Horvitz, and M. Czerwinski. 2002. Scope: Providing awareness of multiple notifications at a glance. Proceedings of Advanced Visual Interfaces 2002, Trento, Italy.
27. Vila, T., R. Greenstadt, and D. Molnar. 2004. Why we can't be bothered to read privacy policies: Models of privacy economics as a lemons market. In Economics of Information Security. Vol 12 of Advances in Information Security, eds. L.J. Camp and S. Lewis, 143-154. Boston: Kluwer Academic Publishers.
28. Wayne R. Barnes, Rethinking Spyware: Questioning the Propriety of Contractual Consent to Online Surveillance, 39 U.C. Davis L. Rev 1545 (2006)
29. Winn, J. Contracting Spyware by Contract, 20 Berkeley Tech. L.J. 1345 (2005)

Advances in Economic Analysis & Policy

Volume 6, *Issue* 2 2006 *Article* 3

FIELD EXPERIMENTS

...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay

Tanjim Hossain* John Morgan†

*Hong Kong University of Science & Technology, tanjim@ust.hk

†University of California, Berkeley, morgan@haas.berkeley.edu

Copyright ©2006 The Berkeley Electronic Press. All rights reserved.

...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay*

Tanjim Hossain and John Morgan

Abstract

Many firms divide the price a consumer pays for a good into two pieces—the price for the item itself and the price for shipping and handling. With fully rational customers, the exact division between the two prices is irrelevant—only the total price matters. We test this hypothesis by selling matched pairs of CDs and Xbox games in a series of field experiments on eBay. In theory, the ending auction price should vary inversely with the shipping charge to leave the total price paid constant. Contrary to the theory, we find that charging a high shipping cost and starting the auction at a low opening price leads to higher numbers of bidders and higher revenues when the shipping charge is not excessive. We show that these results can be accounted for by boundedly rational bidding behavior such as loss-aversion with separate mental accounts for different attributes of the price or disregard for shipping costs.

KEYWORDS: internet auctions, e-retailing, field experiments, price framing, mental accounting, loss aversion

---

*We thank two anonymous referees, Mike Baye, Raj Chetty, Richard Engelbrecht-Wiggans, Stefano DellaVigna, Antonio Rangel, Jason Shachat, Ken Steiglitz, Steve Tadelis, Jeremy Tobacman and seminar participants at Harvard, SITE conference, UC Berkeley, UC Santa Barbara and Washington University in St. Louis for helpful comments. The second author also thanks Charles Thibeaux for his able research assistance in conducting earlier pilot experiments. All remaining errors are our own. The first author gratefully acknowledges the financial support of the Woodrow Wilson Foundation. The second author gratefully acknowledges the financial support of the National Science Foundation. This paper was previously circulated under the title, "A Test of the Revenue Equivalence Theorem Using Field Experiments on eBay."

Erratum

The equation for bidder i's payment if she wins on page 5 should read $P_i(r,o,s) = \max_{j \neq i} \beta_j(r,o,s) + s$. Consequently, the subscript on β and γ in the proof of Proposition 1 should be j instead of i. These are typographical errors, and none of the results and conclusions of the paper are affected.

A common marketing practice, particularly in e-retail, is to divide the price a consumer pays for a good into two pieces—the price for the item itself and the price for shipping and handling. A similar practice has been prevalent in certain sectors of the offline retail sector as well. Many items advertised on television and "not sold in stores," such as Ginsu knives as well as other products sold by the companies K-tel and Ronco, highlight the price for the item itself and then (typically in smaller type) add on a shipping and handling charge. A similar practice occurs routinely in restaurants. In the US, it is common to automatically add an 18% gratuity to the list price for parties of eight or more. Outside the US, such a gratuity (although at a lower percentage) is often automatically added to bills for parties of any size. In theory, the practice of dividing a price into these two pieces should have little effect on overall demand for a good. A perfectly informed and fully rational consumer will merely add together the two parts of a price to obtain the total out of pocket price for an item and then determine whether or not to buy based on this total price.

While for some cases there may be an important cost reason to keep the "list" price and the shipping and handling charge separate. For instance, a customer buying a consumer electronics product from an e-retailer will often be given a choice in the speed in which an item is shipped. Since faster shipping does indeed cost the e-retailer more to provide, it makes sense that the total price is divided in this fashion. Similarly, in book retailing online, the cost to the company of shipping the second and additional items to the same address is typically lower than the cost to ship the first item and the shipping and handling schedule will often reflect this nonlinearity.

In other instances, however, there is no cost-based reason for the division of the price into two parts. For example, it is clear that if there is only one way in which the item is to be shipped or one cannot really even separate the item (say, an item from a menu) from the service (actually serving the item on the menu to the customer at the restaurant), the two part price is not selected for cost-based reasons.

It is, however, anecdotally suggested that dividing the price into two parts leads to an increase in demand. The informal argument suggests that consumers pay attention to the list price while neglecting the "add-on" price associated with shipping or with the gratuity. Therefore, they systematically underestimate the total out of pocket price paid for the items being purchased or consumed and hence the seller is able to sell more of a given item than had that firm simply listed a single total price. Mental accounting (Kahneman and Tversky, 1984 and Thaler, 1985) offers one way to formalize this intuition. Here, the idea is that consumers may keep separate mental "accounts" for each of the components of the provision of an item, such as an account for the item itself and a separate account for the shipping of the item. Under this framework, demand under the

"total price" need not be the same as the demand when the price is decomposed into prices for the various accounts. To estimate how various price "frames" affect demand in a posted price market setting using field data is difficult. One would need a significantly large data set on demand resulting from different price frames. That is, since the two prices making up the frame are fixed, one would have to use variation in quantity demanded to test for an effect. The online auction site, eBay, however, offers an attractive alternative for testing whether variation in the price frame matters. In a typical eBay auction, the quantity available for sale as well as the shipping and handling charge are fixed while the market clearing price for the item (exclusive of shipping) is determined endogenously through eBay's auction process. By relying on this price variation rather than quantity variation, one is able to much more readily determine the effect of changes in the price frame.

Exploiting this observation, we conducted 80 auctions on eBay. Forty of these auctions were for various popular music CDs while the remaining 40 auctions were for video games for Microsoft's Xbox gaming console. The items auctioned were chosen to be "simple" in the sense that all of these items are readily available at retail stores; thus differential expertise or information about the quality of a given item among bidders would seem to play little role. Further, since these items tend to be of short-lived popularity, a bidder's estimate of resale value is unlikely to be an important consideration either. We auctioned all of these items using the standard (in many cases the default) rules governing auctions on eBay—again in hopes of isolating the effect of changing the price frame from other effects that might be present in more complicated settings such as auctions of unique items where expertise might be important.

Auctions on eBay are, in effect, English auctions run over seven day periods.[1] Hence, if a bidder knows his or her value for an item, her optimal bidding strategy is simply to bid up to maximum willingness to pay in that auction as her final bid.[2] When a bidder has outside options (future auctions or a retail market), a bidder should bid up to the point where she is just indifferent between winning the auction at her bid amount and the value of the outside option. In determining whether to bid at all, a bidder must determine whether her willingness to pay exceeds the value of the "effective reserve price" in an auction. The effective reserve price is the minimum amount any bidder has to pay in the event that no other bidders bid in an auction. For eBay, the effective or total reserve price is simply the sum of the opening bid amount and the shipping and handling charge. A key implication of auction theory is strategic equivalence.

[1] The interested reader should see Roth and Ockenfels (2002) for additional details about the specifics of bidding on eBay.
[2] For simplicity, we will use valuation and willingness to pay interchangeably.

Two eBay auctions with the same effective reserve price are strategically equivalent in the sense that a bidder's effective maximum bid (her bid on eBay plus the shipping charge) should be invariant to changes in the composition of the effective reserve price. An example helps to fix ideas. Suppose that a bidder had a maximum willingness to pay of $40 for the delivery of an Xbox game to her home via first-class mail. Suppose that shipping and handling were free for the item. Then the bidder should be willing to bid up to her true willingness to pay, $40. If, however, the auction had a shipping and handling charge of $5 for the item, then the bidder should bid only up to $35 in the auction. Notice that her effective bid—the amount that she would pay to the seller in the event that she won the auction at her amount bid—is exactly the same, $40, in the two auctions. Thus, the two auctions are strategically equivalent. Moreover, strategic equivalence in turn implies that the expected revenues to the seller will be the same under variations in the shipping charge keeping the effective reserve level fixed—that is, these auctions are also predicted to be revenue equivalent as well. Since the revenue equivalence theorem is widely considered to be the cornerstone of auction theory, our experiments should also be of interest in that literature.[3]

In our experiments, we auctioned 4 copies each of 10 popular music CDs and 10 Xbox games. Half of these auctions were run with a $4 effective reserve and the other half with an $8 effective reserve. For a given reserve amount, we auctioned off matched pairs of CDs where we varied the level of the opening bid and the shipping cost by $4 while holding fixed the overall reserve level.

Our main findings are as follows: When the effective reserve was $4, auctions with a low ($0.01) opening bid and high ($3.99) shipping charges attracted more bidders, earlier bidding, and yielded higher revenue for both CDs and Xbox games. With $8 effective reserve, the low opening bid was ($2) and high shipping charge ($6) generated higher revenue for Xbox games. In these cases, charging a significant portion of the price as shipping increased revenue. For CDs, the $8 effective reserve represents over 50% of the retail price of the item. Here, we find no systematic difference in the number of bidders attracted to the auction or revenues as a function of how the effective reserve is allocated between opening bid and shipping charges. An institutional detail of eBay may account for this non-result: A shipping charge of $6 is uncommon in eBay auctions of music CDs while is not uncommon for auctions of Xbox games. For

[3] The main interest in the revenue equivalence theorem centers around differences in the auction form (i.e., first-price versus second-price), which we are not testing. Nonetheless, for the conditions of our experiments, revenue equivalence is predicted to hold regardless of whether bidders have private values or know how many other bidders are competing, are symmetric, and so on. We tried, however, to choose conditions approximating the usual statement of the theorem (see, for example, Proposition 3.1 in Krishna, 2002) to offer the theory its best chance of succeeding.

all other treatments in the experiment, the shipping prices charged were well within the norms observed on eBay.

The nearest antecedent to our paper is Katkar and Reiley (2005), who conducted auctions for Pokémon trading cards on eBay to test the equivalence between public and secret reserve prices. List and Lucking-Reiley (2000), who examine whether the auction theory prediction of demand reduction in multiple unit uniform price auctions is observed in the data, also use field experiments. Lucking-Reiley (1999) and Reiley (2005) conducted field experiments examining the implications of varying auction form and reserve prices using collectible trading cards on Usenet.

The literature on auction laboratory experiments is vast (see, for instance Davis and Holt, 1993; and Kagel, 1995 for surveys of the literature). The main focus on this literature is to study the effects of variation on auction form (i.e. first-price versus second-price, static versus dynamic) as well as bidder information structure (i.e., independent versus correlated signals about value) on auction outcomes. Our main finding that a seller usually earns greater revenue by setting a lower opening bid and a higher shipping charge is similar to the findings of Morwitz, Greenleaf, and Johnson (1998). They show that, relative to a standard first price auction, bidders effectively bid more aggressively when the winner of the auction has to pay 115% of her bid. Unlike in our field experiments where the total cost is the price from the auction plus shipping fee, the total cost to a winner cannot be so easily calculated in that experiment.

Within the empirical literature on e-retailing, our work is related to Ellison and Ellison (2004), who show that "obfuscation" in computer memory is a common (and apparently profitable) retailing strategy. Smith and Brynjolfsson (2001) find that in online book retailing consumers are more sensitive to variation in shipping charges than to variation in price. Again, this is contrary to the notion that only the total price matters in determining demand but in a direction opposite to our findings. There are several key differences between Smith and Brynjolfsson and our work. First, Smith and Brynjolfsson can only observe consumers' click behavior (last clicks more precisely) rather than actual purchases. Second, as we mentioned above, there are economic reasons for variation in shipping charges in book retailing as well as nonlinearity in shipping schedules that make these prices arguably more salient than in our setting.

The remainder of the paper proceeds as follows: In section 2, we sketch the theory leading to strategic equivalence. In section 3, we detail the experimental methodology. In section 4, we present the results of the experiments. In section 5, we consider a number of possible explanations for our findings. We conclude the paper in section 6.

1. Theory

Consider a second-price sealed bid auction for some object, which one can think of as a stylized version of an eBay auction.[4] In this auction, a stochastic number of bidders, $i = 1, 2,\ldots, N$ submit bids b_i for an object. The high bidder pays the second-highest bid amount plus a shipping charge, s. Further, there is a minimum opening bid, o. A bidder must bid at least o to participate in the auction. Further, if only a single bidder bids, then that bidder pays the amount of the effective reserve price, $r = o + s$. Suppose that for an auction with reserve price r, opening bid o and shipping charge s, Bidder i makes equilibrium bid $\beta_i(r, o, s)$. If that bidder wins, then her payment is:

$$P_i\left(r,o,s\right) = \max_{j\neq i} \beta_i\left(r,o,s\right) + s.$$

Next, consider a variation where the effective reserve price remains fixed but o and s are varied to o' and s' respectively. That is, there exists some $\varepsilon \neq 0$ such that $o' = o + \varepsilon$ and $s' = s - \varepsilon$.

Proposition 1 *Suppose $\beta_i(r, o, s)$ is an equilibrium bidding profile in a second-price auction under (r, o, s). Then, $\gamma_i(r, o, s) = \beta_i(r, o, s) + \varepsilon$ is an equilibrium in a second-price auction under (r, o', s').*

Proof. Notice that the conditions in which each bidder participates are identical under (r, o, s) as under (r, o', s'). Further, under the γ bidding strategies, the conditions in which bidder i wins the auction are identical under (r, o, s) as under (r, o', s'). Finally, notice that if bidder i wins, here expected payment is

$$\begin{aligned} P_i\left(r,o',s'\right) &= \max_{j\neq i} \gamma_i\left(r,o',s'\right) + s' \\ &= \max_{j\neq i} \left(\beta_i\left(r,o,s\right) + \varepsilon\right) + s - \varepsilon \\ &= \max_{j\neq i} \beta_i\left(r,o,s\right) + s \\ &= P_i\left(r,o,s\right), \end{aligned}$$

which is identical under (r, o, s) as under (r, o', s') . Therefore, if the β strategies comprise an equilibrium under (r, o, s) then the γ strategies comprise an equilibria under (r, o', s'). ■

[4] The basic intuition for the main result in this section can be readily extended to other possible models of eBay such as English auctions or dynamic second-price auctions.

In words, Proposition 1 shows that variation in the composition of the effective reserve price between the opening bid and the shipping charge leads to strategically equivalent bidding strategies. This immediately implies:

Corollary 1 *Fix an effective reserve level r. Then all eBay auctions under r are revenue equivalent.*

Finally, since we vary the effective reserve level in the experiments, we restate the following well-known result from auction theory (See Krishna (2002)):

Proposition 2 *Consider an eBay auction where bidders have private values. Raising the effective reserve price (regardless of its composition): reduces the number of expected bidders, decreases the probability of a sale, and increases expected revenues conditional on a sale being made.*

2. Procedures

We wanted to auction goods where multiple units of the same good are identical and where markets are thick enough so that our auctions were unlikely to have a marked effect on market prices for these goods overall. First, we chose personal entertainment goods: music CDs and Xbox games as the categories in which we want to participate. Then, we did a survey of eBay sales in each of these markets to select items that are currently popular within each category. By choosing relatively popular CDs, we would expect smaller variance in the sale price from random fluctuations in demand for the good. Further, we hoped that the thickness of the market would disguise the fact that we are running a field experiment. This alleviates the worry that bidders might behave differently when they are aware that they are participating in an experiment.

A number of studies have found the existence of reputation effects on sales price for eBay auctions (see, for instance, Resnick and Zeckhauser, 2002). To control for this effect, we created unique seller accounts for each of the goods we sold. Each of these seller accounts had a zero feedback rating. Thus, the sellers were all identical from a bidder's perspective.

For each type of good (music CDs and Xbox games), we ran four treatments. The first two treatments were constructed as follows: In treatment A, we set an opening bid of $4 and set the shipping and handling cost of the item at $0. In treatment B, we set an opening bid of $0.01 and a shipping cost of $3.99. We will refer to treatments A and B collectively as "low reserve price" treatments. The second set of two treatments were as follows: In treatment C, the opening bid was set at $6 and the shipping cost at $2. Finally, treatment D set the opening bid at $2 and the shipping and handling cost at $6. We will refer to

treatments C and D collectively as "high reserve price" treatments. For bidders, treatments A and B are strategically equivalent and treatments C and D are strategically equivalent.

For all treatments, we used an identical description of the good and stipulated the exact method of shipping (USPS first-class). The exact wording used in each of the auctions is given in the Appendix. To control for the possibility that local bidders might assume that shipping and handling charges might be avoided by picking up the item in person, we responded to all bidder queries along these lines that the shipping and handling charge was non-negotiable and that in-person pickup was not possible. Upon receipt of payment, items were shipped to the winning bidder exactly as described.

Our first round of experiments consisted of auctioning four copies of each of the 10 music CDs. Each of these CDs was purchased new (and sealed) from Amazon.com. In every case, we used the standard seven day closing rules on eBay, varying only the opening bid and the shipping and handling amounts as described above. We ran the low reserve price treatments for auctions scheduled to end on 11/19/2001 and 11/20/2001 (a Monday and a Tuesday, respectively). For the auctions scheduled to end on Monday, we randomly chose whether to run treatment A or B with equal probability. The following day, we posted auctions for the same 10 CDs using the opposite treatment. Thus, for each CD, treatments A and B occur (in random order) on consecutive days. The reason for adopting this procedure is that we would expect there to be little difference in demand for auctions for a given CD that are 24 hours apart. Further, if there is a systematic Monday versus Tuesday effect on revenues, our randomization should avoid confounding this with a treatment effect. The following week, we performed an identical procedure for the high reserve price treatments. All of the auctions ended between eight and ten p.m. eastern standard time.

Our second round of experiments, which consisted of auctioning four copies of each of 10 Xbox games, occurred in March 2002. Each of these games was purchased new and sealed from Amazon. We chose the timing of this round of experiments to avoid the Christmas holiday, when we expected demand might be different from other times of the year. We followed exactly the same procedures as with the music CDs. The next section discusses the results from these 80 auctions.

3. Results

3.1 Overview

Tables 1 and 2 summarize the results from music CD and Xbox games auctions respectively. The column "Bids" presents the total number of bids and "Bidders"

is the number of distinct bidders who placed at least a bid in that auction. For the low reserve price treatments, all but one of the CDs auctioned were sold; whereas under the high reserve price treatments, 5 out of the 20 CDs went unsold. This is broadly consistent with the prediction that higher reserve prices are more likely to lead to failures to sell. Conditional on being sold, the average final sale price (including shipping) for the low reserve price treatments was $9.30. Under the high reserve price treatments, the final price conditional on a sale was $12.21. This too is consistent with theoretical prediction that prices conditional on a sale occurring are higher under a higher reserve price. Finally, sale prices on eBay tend to be well below the retail price. The retail prices for each of the CDs represent the cost to us from purchasing from Amazon, allocating the total shipping cost equally over all 40 music CDs. Including shipping costs, the average price we paid was $14.82 per CD; thus, winning bidders in our auctions received "bargains" compared to buying at Amazon. This can be explained by the fact that many consumers with relatively lower valuations shop on eBay to find great deals.

In all treatments of Xbox games auctions, every Xbox game was sold as Table 2 shows. The average final sale price (including shipping fees) for the low reserve price treatments was $37.47. Under the high reserve price treatments, the final price conditional on a sale was $39.01. This is consistent with the theoretical prediction that prices conditional on a sale occurring are higher under a higher reserve price. This revenue improvement prediction, however, assumes that at least the higher reserve price is inside the support of the bidders' private valuations. If both reserve prices are below the lowest possible valuation, they will generate the same expected revenue conditional on a sale. For Xbox games, even $8 is significantly lower than the pre-tax retail price of $49.99. Thus, a very small fraction of bidders may have valuation below either of the reserve prices. Not surprisingly, although the mean revenue is higher for $8 reserve, this difference is not statistically significant. As with the music CDs, sale prices on eBay tend to be below the retail price. Including shipping charges, the average price we paid was $51.07 per Xbox game; thus, winning bidders in our auctions received "bargains" paying about 72% of the cost at Amazon.

Table 1. Overview of Music CD Auctions

Low Reserve Treatments:

Total Reserve = $4	**Treatment A**			**Treatment B**		
CD	**Revenue**	**Bids**	**Bidders**	**Revenue**	**Bids**	**Bidders**
Music	5.50	4	2	7.24	6	4
Ooops! I Did it Again	6.50	3	3	7.74	10	4
Serendipity	8.50	5	4	10.49	8	4
O Brother Where Art Thou?	12.50	7	7	11.99	7	4
Greatest Hits - Tim McGraw	11.00	11	8	15.99	12	8
A Day Without Rain	13.50	7	6	14.99	9	6
Automatic for the People	0.00	0	0	9.99	5	3
Everyday	7.28	3	3	9.49	9	7
Joshua Tree	6.07	3	3	8.25	6	3
Unplugged in New York	4.50	3	3	5.24	5	2
Average	*7.54*	*4.6*	*3.9*	*10.14*	*7.7*	*4.5*

High Reserve Treatments:

Total Reserve = $8	**Treatment C**			**Treatment D**		
CD	**Revenue**	**Bids**	**Bidders**	**Revenue**	**Bids**	**Bidders**
Music	9.00	3	3	8.00	1	1
Ooops! I Did it Again	0.00	0	0	0.00	0	0
Serendipity	12.50	5	4	13.50	5	4
O Brother Where Art Thou?	11.52	5	5	11.00	7	5
Greatest Hits - Tim McGraw	18.00	9	3	17.00	15	7
A Day Without Rain	15.50	11	7	16.00	10	5
Automatic for the People	0.00	0	0	0.00	0	0
Everyday	10.50	6	4	13.50	6	4
Joshua Tree	8.00	1	1	11.10	7	4
Unplugged in New York	8.00	1	1	0.00	0	0
Average	*9.30*	*4.1*	*2.8*	*9.01*	*5.1*	*3*

Table 2. Overview of Xbox Auctions

Low Reserve Treatments:

Total Reserve = $4	**Treatment A**			**Treatment B**		
Game	**Revenues**	**Bids**	**Bidders**	**Revenues**	**Bids**	**Bidders**
Halo	34.05	13	9	41.24	11	9
Wreckless	44.01	7	7	33.99	12	7
Circus Maximus	40.99	8	7	39.99	10	6
Max Payne	36.01	24	10	36.99	16	10
Genma Onimusha	41.00	17	8	32.99	14	11
Project Gotham Racing	37.00	13	6	38.12	11	8
NBA 2K2	42.12	20	10	42.99	15	9
NFL 2K2	26.00	14	5	33.99	9	9
NHL 2002	36.00	10	8	37.00	8	8
WWF Raw	33.99	13	8	40.99	21	11
Average	*37.12*	*13.9*	*7.8*	*37.83*	*12.7*	*8.8*

High Reserve Treatments:

Total Reserve = $8	**Treatment C**			**Treatment D**		
Game	**Revenues**	**Bids**	**Bidders**	**Revenues**	**Bids**	**Bidders**
Halo	40.01	10	8	43.00	14	10
Wreckless	35.00	11	6	36.00	10	7
Circus Maximus	39.00	8	7	42.53	13	10
Max Payne	37.50	27	8	42.00	24	11
Genma Onimusha	36.00	18	11	37.00	13	8
Project Gotham Racing	35.02	13	7	40.01	12	10
NBA 2K2	41.00	26	7	45.00	15	10
NFL 2K2	33.00	15	8	40.10	21	10
NHL 2002	36.00	16	12	41.00	18	14
WWF Raw	37.00	17	10	44.00	27	11
Average	*36.95*	*16.1*	*8.4*	*41.06*	*16.7*	*10.1*

Finally, we check for the presence of "day of the week" effects in the data. Under the null hypothesis that there are no systematic differences between Mondays and Tuesdays, the revenue ranking for each pair of CDs or games should be the outcome of a fair coin flip. We find that for 20 out of 38 matched pairs of auctions, higher revenue was obtained on Tuesdays compared to Mondays; thus we do not find evidence of a systematic day of the week effect.[5]

We now look more closely into revenue ranking and effect of different treatments on the number of bidders.

3.2 Revenues

Table 3 compares revenues under the low reserve treatments. Notice for music CDs, the average revenue under treatment B is $10.14 compared to only $7.54 in treatment A. Of course, part of this difference is accounted for by the fact that one of the CDs under treatment A did not sell. If we exclude this CD from the averages, the average under treatment B still exceeds that under treatment A by $1.79, or about 21% of the price CDs sold for under treatment A. For Xbox games, average revenues under treatment B still exceed those under treatment A, but by a smaller margin, only $0.71 (or 2% of the revenues under treatment A). Taken together, however, this suggests that there might be systematic revenue differences between the two treatments.

More formally, we can test the null hypothesis of revenue equality against the one-sided alternative that B outperforms A using a binomial test. This essentially involves a count of the number of auctions in which treatment B yielded higher revenues than did treatment A. In 9 out of 10 matched pairs of auctions for music CDs, this is the case. In 7 out of 10 matched pairs of auctions for Xbox games treatment B outperforms A. Thus, for 16 out of 20 matched pairs of auctions with low reserve prices, we find that treatment B outperforms treatment A. Using a one-sided binomial test, then we get a p-value of 0.005. Alternatively, one could use a Wilcoxon signed-ranks test to test the null hypothesis of no treatment effect against a two-sided alternative. Performing this test, we obtain a test statistic of $z = 2.249$, which is significant at the 97.5% level. We can obviously reject that treatments A and B are revenue equivalent in favor of the hypothesis that treatment B generates higher revenue. Thus, we get significant increase in revenue for auctions with a higher shipping and handling fee than auctions with free shipping even though the auctions end within one day of each other. This may suggest that the trade volume on eBay is large enough that these auctions are quite independent even though they overlap in time.

[5] In the high reserve price treatment, the CDs "Oops!…I did it again" and "Automatic for the People" did not sell under either treatment; hence there is no revenue ordering for these items. The "day of the week" effect is also insignificant if we look at music CDs and Xbox games separately.

Indeed, since we chose music CDs and Xbox games that are popular on eBay, many auctions of identical objects went on at the same time.

Table 3. Revenues from Low Reserve Treatments

CD Title	Revenues under Treatment A	Revenues under Treatment B	B - A	Percent Difference
Music	5.50	7.24	1.74	32%
Ooops! I Did it Again	6.50	7.74	1.24	19%
Serendipity	8.50	10.49	1.99	23%
O Brother Where Art Thou?	12.50	11.99	-0.51	-4%
Greatest Hits - Tim McGraw	11.00	15.99	4.99	45%
A Day Without Rain	13.50	14.99	1.49	11%
Automatic for the People	0.00	9.99	9.99	
Everyday	7.28	9.49	2.21	30%
Joshua Tree	6.07	8.25	2.18	36%
Unplugged in New York	4.50	5.24	0.74	16%
Average	*7.54*	*10.14*	*2.61*	*35%*
Average excluding unsold	*8.37*	*10.16*	*1.79*	*21%*

Xbox Game Title	Revenues under Treatment A	Revenues under Treatment B	B - A	Percent Difference
Halo	34.05	41.24	7.19	21%
Wreckless	44.01	33.99	-10.02	-23%
Circus Maximus	40.99	39.99	-1.00	-2%
Max Payne	36.01	36.99	0.98	3%
Genma Onimusha	41.00	32.99	-8.01	-20%
Project Gotham Racing	37.00	38.12	1.12	3%
NBA 2K2	42.12	42.99	0.87	2%
NFL 2K2	26.00	33.99	7.99	31%
NHL 2002	36.00	37.00	1.00	3%
WWF Raw	33.99	40.99	7.00	21%
Average	*37.12*	*37.83*	*0.71*	*2%*

Table 4 compares revenues under the high reserve treatments. Here, one sees a distinct difference in the ranking of average revenues between CDs and Xbox games. Excluding the two CDs that went unsold under both treatments, the average revenues under treatment C are $11.63 compared to $11.26 under treatment D.

Table 4. Revenues from High Reserve Treatments

CD Title	Revenues under Treatment C	Revenues under Treatment D	D - C	Percent Difference
Music	9.00	8.00	-1.00	-11%
Ooops! I Did it Again	0.00	0.00	0.00	
Serendipity	12.50	13.50	1.00	8%
O Brother Where Art Thou?	11.52	11.00	-0.52	-5%
Greatest Hits - Tim McGraw	18.00	17.00	-1.00	-6%
A Day Without Rain	15.50	16.00	0.50	3%
Automatic for the People	0.00	0.00	0.00	
Everyday	10.50	13.50	3.00	29%
Joshua Tree	8.00	11.10	3.10	39%
Unplugged in New York	8.00	0.00	-8.00	-100%
Average	*9.30*	*9.01*	*-0.29*	*-3%*
Average excluding unsold	*12.15*	*12.87*	*0.73*	*6%*

Game Title	Revenues under Treatment C	Revenues under Treatment D	D - C	Percent Difference
Halo	40.01	43.00	2.99	7%
Wreckless	35.00	36.00	1.00	3%
Circus Maximus	39.00	42.53	3.53	9%
Max Payne	37.50	42.00	4.50	12%
Genma Onimusha	36.00	37.00	1.00	3%
Project Gotham Racing	35.02	40.01	4.99	14%
NBA 2K2	41.00	45.00	4.00	10%
NFL 2K2	33.00	40.10	7.10	22%
NHL 2002	36.00	41.00	5.00	14%
WWF Raw	37.00	44.00	7.00	19%
Average	*36.95*	*41.06*	*4.11*	*11%*

Thus, for music CDs, we do not find that increasing shipping costs and reducing the opening bid one-for-one is revenue enhancing. However, this difference in revenues is reversed if we only look at revenues from the seven observations where the same CD sold under both treatments. Here, we find that treatment D yields an average revenue of $12.87 compared to $12.15 under treatment C. Regardless, the presence of the treatment effect on revenues is tenuous at best. In 4 matched pairs of auctions, treatment D had higher revenue

than treatment C, in another 4 matched pairs, this revenue ranking is reversed. Finally, in 2 auctions, the CD did not sell under either treatment, so the revenues are the same. Moreover, the test-statistic for Wilcoxon signed-ranks test is 0. Clearly, one fails to reject the null hypothesis of revenue equivalence at any level for this data; that is, a treatment effect is absent.

For Xbox games, the situation more closely resembles that in Table 3. Average revenues under treatment D are higher by $4.11 compared to treatment C. This is an 11% difference in revenues. Further, in 10 out of 10 matched pairs of auctions, treatment D yields higher revenues than treatment C. The test-statistic for Wilcoxon signed-ranks test is 3.364 implying that we can reject the null hypothesis of revenue equivalence in favor of the one-sided alternative at any significance level.

The key feature that the low reserve price treatments and the high reserve price treatment with Xbox games share is that the reserve is less than 27% of the retail price. In contrast, for CDs under the high reserve price treatments, the reserve represents 53% of the retail price. Thus, the data suggests that the theoretical prediction of revenue equivalence fails systematically for relatively low reserve prices but not for relatively high reserve prices. In section 4, we will return to this idea and offer some ex post theoretical rationales drawn from behavioral economics to try to explain this difference.

Table 5 examines the effect of raising the total reserve price across treatments. Recall that treatment C and D are constructed by adding $2 to the opening bid and $2 to the shipping cost to treatments A and B, respectively. Having already established that pooling treatments A and B is not justified, we study revenues under C versus A (Table 5) and D versus B (Table 6) to get at reserve price effects.

Looking at CDs in table 5, we earlier observed that the likelihood of not selling the item was higher under treatment C (2 items unsold) compared to treatment A (1 item unsold). Conditional on a sale, revenues should be higher under treatment C compared to A. As the table shows, conditional on the sale of the item, average revenues under C are $11.63 versus $8.61 under treatment A. Further, in 7 of 8 auctions where the CDs were sold under both treatments, revenue was higher under treatment C than treatment A. For these 8 CDs, we obtain a z-value of 2.380 using the Wilcoxon signed-ranks test accepting the hypothesis that for low shipping treatments, higher effective reserve generated higher revenue conditional on sale at the 99% confidence level.

All items sold under both treatments for Xbox games, so there was no observable difference in the probability of a good going unsold. Interestingly, the average revenues under the high reserve treatment, treatment C, are lower than under treatment A. There is no systematic difference in the revenues for a given game across treatments. Treatment C yields higher revenues than A for 4 of 10

matched pairs, whereas A yields higher revenues than C for 5 out of 10 matched pairs. Revenues are exactly equal for the game NHL 2002. Thus, one cannot reject the null hypothesis of no treatment effect for this data.

Table 5. Effects on Revenues of Changes in Reserve Price: A vs C

CD Title	Revenues under Treatment A	Revenues under Treatment C	C - A	Percent Difference
Music	5.50	9.00	3.50	64%
Ooops! I Did it Again	6.50	0.00	-6.50	-100%
Serendipity	8.50	12.50	4.00	47%
O Brother Where Art Thou?	12.50	11.52	-0.98	-8%
Greatest Hits - Tim McGraw	11.00	18.00	7.00	64%
A Day Without Rain	13.50	15.50	2.00	15%
Automatic for the People	0.00	0.00	0.00	
Everyday	7.28	10.50	3.22	44%
Joshua Tree	6.07	8.00	1.93	32%
Unplugged in New York	4.50	8.00	3.50	78%
Average	*7.54*	*9.30*	*1.77*	*23%*
Average excluding unsold	*8.61*	*11.63*	*3.02*	*35%*

Game Title	Revenues under Treatment A	Revenues under Treatment C	C - A	Percent Difference
Halo	34.05	40.01	5.96	18%
Wreckless	44.01	35.00	-9.01	-20%
Circus Maximus	40.99	39.00	-1.99	-5%
Max Payne	36.01	37.50	1.49	4%
Genma Onimusha	41.00	36.00	-5.00	-12%
Project Gotham Racing	37.00	35.02	-1.98	-5%
NBA 2K2	42.12	41.00	-1.12	-3%
NFL 2K2	26.00	33.00	7.00	27%
NHL 2002	36.00	36.00	0.00	0%
WWF Raw	33.99	37.00	3.01	9%
Average	*37.12*	*36.95*	*-0.16*	*0%*

Table 6 compares revenues under treatments B and D. Looking at CDs, again we find a higher probability that the item goes unsold with a high reserve price. This happens three times under treatment D and never under treatment B. Conditional on the item being sold, average revenues are higher ($12.87) under

treatment D than under treatment B ($11.21). Further, conditional on the item being sold, treatment D yields higher revenues in 6 out of 7 matched pairs. Results are quite similar for Xbox games. Average revenues are higher for treatment D compared to treatment B, and, indeed, in 10 out of 10 matched pairs, D yields higher revenues than B.

Table 6. Effects on Revenues of Changes in Reserve Price: B vs D

CD Title	Revenues under Treatment B	Revenues under Treatment D	D - B	Percent Difference
Music	7.24	8.00	0.76	10%
Ooops! I Did it Again	7.74	0.00	-7.74	-100%
Serendipity	10.49	13.50	3.01	29%
O Brother Where Art Thou?	11.99	11.00	-0.99	-8%
Greatest Hits - Tim McGraw	15.99	17.00	1.01	6%
A Day Without Rain	14.99	16.00	1.01	7%
Automatic for the People	9.99	0.00	-9.99	-100%
Everyday	9.49	13.50	4.01	42%
Joshua Tree	8.25	11.10	2.85	35%
Unplugged in New York	5.24	0.00	-5.24	-100%
Average	*10.14*	*9.01*	*-1.13*	*-11%*
Average excluding unsold	*11.21*	*12.87*	*1.67*	*15%*

Game Title	Revenues under Treatment B	Revenues under Treatment D	D - B	Percent Difference
Halo	41.24	43.00	1.76	4%
Wreckless	33.99	36.00	2.01	6%
Circus Maximus	39.99	42.53	2.54	6%
Max Payne	36.99	42.00	5.01	14%
Genma Onimusha	32.99	37.00	4.01	12%
Project Gotham Racing	38.12	40.01	1.89	5%
NBA 2K2	42.99	45.00	2.01	5%
NFL 2K2	33.99	40.10	6.11	18%
NHL 2002	37.00	41.00	4.00	11%
WWF Raw	40.99	44.00	3.01	7%
Average	*37.83*	*41.06*	*3.24*	*9%*

3.3 Bidders

Going back to Table 1, we look at how bidding behavior differed in each of the treatments for the music CD auctions. First, notice that treatment B averages about 0.6 more bidders per auction than treatment A. Treatment D averages only 0.2 more bidders than treatment C; thus, whatever effect transferring the reserve from the opening bid to shipping and handling fee has in inducing bidder participation, it is attenuated in the high reserve treatments. Similar analysis for Xbox games can be done using Table 2. Here there are an average of 8.8 bidders from Xbox games under treatment B, whereas there are only 7.8 under treatment A, which is consistent with what we saw for music CDs. An average of 1.6 more bidders bids under treatment D than under treatment C.

The comparison of number of bidders between low and high shipping treatments, on the surface, seems consistent with the revenue ranking we found in the previous section. Using the Wilcoxon signed-ranks test to test more formally for a treatment effect on the number of bidders, we find that the increase in high shipping treatments is not significant for music CDs and for low reserve treatments of Xbox games (treatments B vs. A). However, the increase is significant at 97.5% confidence-level for high reserve treatments of Xbox games (treatments D vs. C) with a test statistic of $z = 1.962$. Pooling the matched pairs of CD auctions and testing the null hypothesis that the number of bidders in high and low shipping treatments (treatments B vs. A and D vs. C) are equal, we obtain a test statistic of $z = 0.956$, which is not significant. Pooling the matched pairs of Xbox games auctions and testing the hypothesis that the number of bidders in high and low shipping treatments are equal, we obtain a test statistic of $z = 2.624$, which is significant at 99.5% level.

Next, we turn to changes in the level of the reserve price. Recall that increases in the level of the overall reserve price are predicted to decrease the number of bidders by excluding those with values between the two different reserve price levels. Comparing treatments A and C (or B and D) in Table 1, one sees evidence of this effect. Compared to treatment A, the average number of bidders under treatment C falls by 1.1 bidders. Likewise, compared to treatment B, the average number of bidders under treatment D is lower by 1.5 bidders. Qualitatively similar results do not occur for Xbox games as shown in Table 2. Compared to treatment A, the average number of bidders under treatment C increases by 0.6 bidders. Similarly, compared to treatment B, the average number of bidders in treatment D increases by 1.3 bidders. This strengthens the assertion that a very small fraction of bidders may have valuation below $8 for Xbox games. The increase in bidders can be due to the fact that high reserve auctions were held one week after the low reserve auctions.

The theory predicts that auctions attracting a larger number of bidders will have higher expected revenues. To study this question, we compare the revenues under treatments A and B for each matched pair of CDs and ask how often the auction with the higher revenues attracted (weakly) more bidders. The evidence is consistent with the theory, in 9 out of 10 auctions, the auction with the higher number of bidders obtained the higher revenue. One obtains similar results from comparing treatments A versus B for Xbox games. In this case, the treatment attracting the higher number of bidders obtained the higher revenue in 8 out of 10 auctions. Comparing treatments C versus D, the CD auction with the higher number of bidders obtained the higher revenue in 6 out of 8 cases, while in 9 out of 10 cases, the Xbox auction with the higher number of bidders obtained the higher revenue. Thus, it seems that a likely transmission mechanism for the revenue ranking we obtained above is that auctions with lower opening bid amounts succeed in attracting more bidders than those with higher opening bids, holding the total reserve price constant.

An investigation of timing of the first bid an auction and the revenue suggests a more nuanced explanation, auctions with lower opening bid amounts attract earlier bidding and this, in turn, leads to higher revenues. There is a strong correlation between receiving the earlier first bid and obtaining higher revenue. In 29 out of 38 matched pairs of auctions, the treatment receiving the earlier bid obtained the higher revenue. Thus, it seems that, holding the overall reserve price fixed, auctions with lower opening bid amounts attract earlier bidders, and this in turn leads to greater overall bidding interest in the item, which ultimately results in higher revenues.

4. Discussion

While strategic equivalence predicts that two auctions for identical items with the same total reserve price should yield the same expected revenue, we find evidence that for auctions on eBay, lowering the opening bid and increasing shipping charges while holding the total reserve price constant raises the revenue of the seller. It seems to do this by attracting earlier and more bidders to the auction. In this section, we discuss a number of possible theoretical models that might rationalize the observed results.

4.1 Loss Aversion and Mental Accounting

Two essential characteristics of a value function to evaluate outcomes of risky prospects proposed by Kahneman and Tversky (1979) and Tversky and Kahneman (1991) are reference-dependence and loss aversion. Reference-dependence suggests that an agent defines gains and losses with respect to a

reference point. Loss aversion hypothesize that negative utility received from a loss is greater in size than utility received from a gain of the same amount.

Further, Kahneman and Tversky (1984) and Thaler (1985) postulate that consumers retain separate mental accounts for different aspects of a purchase decision. Experimental evidences show that a decision maker makes different choices when presented with relevant data in different accounting formats. One plausible way this might happen in our setting is that bidders have separate accounts for shipping and for the good itself.

Combining these observations allows us one way to rationalize our findings as follows: Suppose bidders have separate mental accounts for shipping and the good itself with less emphasis on shipping. Then, the treatment with higher shipping costs will generate higher expected revenue than the treatment with lower shipping cost. However, if bidders are loss-averse with respect to some reference level, then a very high shipping cost may lead to revenue equivalence and revenue performance can even be reversed.

We offer a simple model to operationalize this intuition. Suppose a consumer with valuation of v for some object has a reference level of v_s for shipping charges. When she pays a price of p and shipping cost of s in winning an auction, her utility is:

$$U(v,p,s) = \gamma u(v - v_s - p) + \alpha(v_s - s).$$

Here α is smaller than γ and for simplicity we assume $\gamma = 1$. We further assume that $u(0) = 0$ and, if x is positive, x and $-x$ refer to subtraction and addition of wealth respectively. If the utility function $u(.)$ demonstrates loss aversion then $u(x) \leq -u(-x)$ and $u^{-1}(y) \geq -u^{-1}(-y)$ for all positive x, y. For a fixed s, the buyer resets her maximum bid to p^* where

$$p^* = v - v_s - u^{-1}\left(-\alpha u(v_s - s)\right)$$
$$\text{and} \quad p^* \leq v - v_s + u^{-1}\left(+\alpha u(v_s - s)\right).$$

Since the agent is loss-averse, if $\alpha u(v_s - s) > u(s - v_s)$ for all $s > v_s$ then a bidder's total expenditure on the object will be increasing in s for $s \leq v_s$ and decreasing in s for $s > v_s$. Hence, we can have a situation where

$$u(v - v_s - p_1 + \varepsilon) + \alpha u(v_s - s_1) \geq u(v - v_s - p_1) + \alpha u(v_s - s_1 + \varepsilon) \text{ but}$$
$$u(v - v_s - p_2 + \varepsilon) + \alpha u(v_s - s_2) \leq u(v - v_s - p_2) + \alpha u(v_s - s_2 + \varepsilon).$$

For example, suppose $s_1 = v_s$, $s_2 = 2v_s$, $p_2 = p_1 - v_s$, $\alpha > 0.5$, $\frac{\alpha}{1-\alpha} v_s \geq \varepsilon \geq v_s$ and

$$u(x) = \begin{cases} 2x & \text{for } x < 0 \\ x & \text{for } x \geq 0. \end{cases}$$

Plugging in these values, we obtain:

$$v - v_s - p_1 + \varepsilon \geq v - v_s - p_1 + \alpha\varepsilon \text{ and}$$
$$v - p_1 + \varepsilon - 2v_s \leq v - p_1 + \alpha(\varepsilon - v_s)$$

That is,

$$U(v, p_1 - \varepsilon, s_1) \geq U(v, p_1, s_1 - \varepsilon) \text{ but}$$
$$U(v, p_2 - \varepsilon, s_2) \leq U(v, p_2, s_2 - \varepsilon).$$

This implies that a bidder with valuation v bids higher in the auction with higher shipping cost in the first case and she bids higher in the auction with lower shipping cost in the second (high reserve price) case.

This hypothesis is testable. Let us assume that v_s is the same for all agents for some good. If shipping charges are smaller than v_s for treatment C but not for treatment D, treatment C may generate revenue higher than or equivalent to that of treatment D. On the other hand, if shipping charges are smaller than v_s for all treatments A through D then there will not be loss aversion effects for any treatment. Hence, treatments B and D will outperform treatments A and C respectively. It is worth mentioning that on eBay, average shipping and handling fee for new music CDs is between \$3 and \$4 and for new Xbox games is between \$4 and \$6. Thus, our findings from the music CD and Xbox games experiment are consistent with this explanation.

4.2 Salience

The eBay display makes the current second highest bid much more prominent than the shipping cost. Suppose that a fraction of the potential bidders simply ignore shipping costs when making their bids. In this case, it is still a weakly dominant strategy for each bidder to bid up to his or her expected surplus. Thus, when shipping is zero, the expected price is simply the second highest valuation for the object. When shipping is positive, a fraction of bidders discount for the amount of the shipping, whereas bidders who ignore shipping continue to bid up to their values for the object. This leads to a distribution of bids that first-order stochastically dominates the no shipping cost environment and one obtains the

revenue ranking we observe. Gabaix and Laibson (2005) show that when some agents are naïve, it might be optimal for sellers to hide information by using shrouded attributes. In our experiment, the shipping and handling fee is somewhat shrouded as it is not immediately visible when auctions are listed on eBay.

It seems plausible that the salience of shipping costs depends on two things, the ratio of shipping cost to the total cost of the item and the level of shipping cost itself. We postulate that the fraction of bidders ignoring shipping cost is given by $\lambda(\rho, s)$ where ρ is the ratio of shipping cost to the retail price of the good and s is the level of shipping cost. We also assume that $(\partial\lambda/\partial\rho) < 0$, $(\partial\lambda/\partial s) < 0$ and $\lambda(\rho, s) = 0$ if $\rho \geq \underline{\rho}$ for some $\underline{\rho}$.

Under these assumptions, we can conjecture the ancillary prediction that revenue differences, for a given total shipping cost, should be greater for Xbox games than for CDs (since ρ is smaller for Xbox) and revenue differences should be reduced for higher shipping treatments than for lower shipping treatments. First we test that the revenue difference (between the high shipping cost and low shipping cost treatments) for low reserve price treatments will be greater than that for high reserve price treatments. This is a difference-in-differences ((B – A) – (D – C)) measure and we use the standard one-sided t-statistic for testing the null hypothesis that the mean of a normal population is zero. Here the difference-in-differences in revenue for a given music CD is considered as one observation. Next we test that the revenue difference for Xbox games are greater than that for music CDs. We test this for low reserve price and high reserve price treatments separately. We treat the revenue difference (B – A or D – C) for CDs and Xbox games as two different population samples whose means we would like to compare. Since these two populations may not have the same variance, we use the one-sided Smith-Satterthwaite test for comparing the means of two processes with different variances. The degree of freedom (*DF*) for this test involves both the number of observations in each sample and the sample variances. The results of these tests, presented in Table 7, are somewhat consistent with the salience mode.

Table 7. Differences in Revenue Differences Under Various Treatments

	Test Statistic	**DF**	***t*(.05, *DF*)**
Difference-in-Differences ((B-A)-(D-C)) for CD	2.32	9	1.83
Difference-in-Differences ((B-A)-(D-C)) for Xbox	– 2.41	9	1.83
Difference in Xbox and CD Revenue Differences for Low Reserve (B-A)	– 0.89	13	1.77
Difference in Xbox and CD Revenue Differences for High Reserve (D-C)	3.73	16	1.75

Specifically, we reject (at 95% confidence level) the two alternative hypotheses that:

i) for low reserve price treatments, revenue difference for Xbox games is larger than music CDs

and

ii) for Xbox games, revenue difference is higher for the low reserve price treatment.

We can reject (at 95% confidence level) the null hypotheses in favor of the one-sided alternatives implied by the salience model that:

i) for high reserve price treatments, revenue difference for Xbox games is larger than music CDs

and

ii) for music CDs, revenue difference is higher for the low reserve price treatment.

4.3 Costly Search

Another alternative explanation postulates that the cost of search might account for our findings.[6] Suppose that bidders use the following search strategy: Sort the prices for the desired item from lowest to highest then bid only on the low price item. In that case, since auctions run under treatments B and D have lower initial listed prices than those run under treatments A and C, these auctions would be more likely to attract earlier bidders. For example, suppose there are only two auctions and one is run under treatment A and the other under treatment B. Then, treatment B would get the first two bids. Since, with the same number of bidders placing bids in these two auctions, the expected "price" is always lower in treatment B (the expected "price" plus shipping fee are equal for a given number of bidders), treatment B will receive higher number of bids on average. This translates into higher revenue for treatment B. This is also consistent with the result that auctions attracting earlier bid generated higher revenue. Notice that the revenue difference goes away if we have infinite number of bidders. The revenue difference due to costly search will be larger the thinner the market (in terms of number of bidders) is.

While this may explain the revenue ranking and number of bidders we observe in the low reserve treatments for both types of products and in the high reserve treatment for Xbox games, it does not explain why revenue equivalence holds for the high reserve treatments in the CD auctions. In fact, the market size for CD auctions with high effective reserve is relatively thin implying that the revenue difference between treatment D and treatment C is likely to be larger.

[6] We thank Richard Englebrecht-Wiggans for suggesting this possibility.

One point worth mentioning is that only 15 out of 20 such auctions resulted in sale and the smaller data set could be a driving force behind this equivalence. Costly search also does not explain why Xbox game auctions received higher number of bidders under high reserve treatments. Xbox game auctions with high reserves on average attracted 0.95 bidders per auction more than those with low reserves. This difference is statistically significant at 90% confidence level. However, there was a one-week difference between low reserve and high reserve auctions and that can be a cause behind the increase in bidders.

Finally, we offer several alternative explanations that are less successful in rationalizing our findings.

4.4 Suspicious Bidders

Another possibility is that some bidders have a preference for a more familiar auction format than a less familiar one. While free shipping is sometimes offered on eBay auctions, it is less common than auctions with positive shipping costs, so it may be less familiar to bidders. Further, bidders may view the familiarity of the auction as a signal of the quality of the seller. For instance, bidders may infer from the low shipping charge that the seller is a low quality seller. This might mean that they expect the speed of shipping to be slower (although we described the shipping method using identical language in both types of auctions) or that there is a greater possibility of not receiving the good at all (although we restricted payment to credit cards so that bidders are mostly insulated against this problem). For these reasons, it might be that fewer bidders choose to participate in less familiar auctions and that this accounts for the revenue ranking. This explanation seems less plausible in explaining the same revenue ranking for Xbox games under treatments C and D. There seems little a priori reason to suppose that bidders are less familiar with an auction with a $2 shipping charge than a $6 shipping charge. Moreover, if they are, then one should see a similar revenue ranking for music CDs under these two treatments, but this is inconsistent with the data.

There are further reasons to doubt this explanation. First, many online retailers, such as Barnes and Noble and Amazon have gone to a free shipping policy for purchases about some modest amount. Indeed, the trend in online retail seems to be toward more free shipping. So how unfamiliar is free shipping to bidders? Second, field experiments by List and Lucking-Reiley (2000) found confirmation of many of the theory predictions in comparing the uniform price auction to a 2-unit Vickrey auction, which is a very unfamiliar auction format.[7]

[7] It should be mentioned here that in List and Lucking-Reiley (2000) experiment, the numbers of bidders in different treatments were exogenously given. It is possible that were the bidders free to

4.5 Love of Winning

Suppose that bidders obtain utility simply from winning the auction itself, in addition to their surpluses. In particular, suppose that the utility from winning is a decreasing function of the difference between the winning bid and the retail price of the object. In this case, since bidders are shading down bids for positive shipping treatments, the love of winning "bonus" is greater in when shipping is positive than when it is zero. This leads to revenue non-equivalence of auctions with different shipping and handling fee that is consistent with our findings. However, this does not explain the return of revenue equivalence in the high reserve price treatment for music CDs. Moreover, if love of winning is independent of price level or fully accounts for shipping costs, then revenue equivalence once again obtains. To explain failure of strategic equivalence between first-price and Dutch auctions, Cox, Roberson and Smith (1982) suggest that bidders may gain some utility from winning in a Dutch auction because of the suspense value of the auction structure. This is, in essence, similar to the love of winning. However, Cox, Smith and Walker (1983) reject that explanation using some further experiments.

5. Conclusions

We have shown that when the effective reserve price is not a large fraction of the retail price of the item, the structure of the reserve price affects the number of bidders, the timing of bids, and, most importantly, the revenue derived from the auction. Setting a low opening bid and a high shipping and handling cost yields systematically higher revenue than doing the reverse. When the effective reserve is more than 50% of the retail value of the item, we do not observe these systematic differences. We present a number of theories that attempt to ex post rationalize our findings. One theory suggests that bidders might be loss averse and maintain separate mental accounts for shipping charges and the price of the item itself. Under such a framework, we show that an apparently revenue-neutral variation in the structure of the effective reserve price can lead to revenue differences consistent with those we find. Another potential explanation can be that some fraction of bidders simply overlook the shipping charge when making bids. We would expect that as the shipping charge becomes large relative to the final price of the item, fewer people will ignore it. This theory also leads to systematic variation in revenue as a function of the structure of the reserve price. If many bidders sort the auctions of the desired item in terms of the "current price" and then bid on the lowest price item, that will also lead to auctions with

---

choose the auction where they want to bid, they would choose to bid in auctions with a more familiar format.

high shipping charges having revenue advantage over auctions with the same effective reserve but lower shipping charges.

We began the discussion of the paper by noting that dividing a price into two components was a common marketing practice in posted price markets as well as auctions. A central contribution of this paper is to show the effect of dividing the price in different attributes in an extremely transparent setting where possible payments were made crystal clear to bidders and determined endogenously. We find important revenue differences even in this relatively ideal setting. This suggests to us that the effects we identified are likely to be quite important in posted-price retailing as well. Indeed, the main conclusion we draw is that different framing of the same price as a sum of different attributes may significantly affect consumer behavior.

Appendix

This section gives the exact wording used in placing our auctions on eBay.

CD Auctions

Title: {Name of CD here} -- Brand New in Original Shrink Wrap

Text: This superhit CD is brand new in original shrink wrap. The songs are: {List of songs here.}

I accept only Paypal or Billpoint payments.

Buyer pays shipping and handling charge of {insert shipping charge here.} I will ship via first-class mail.

I will ship to US and Canada only.

Happy Bidding.

For the free shipping treatment, this paragraph read:

There is no shipping and handling fee for this CD. I will ship via first-class mail.

I will ship to US and Canada only.

Happy Bidding.

Xbox Auctions

Title: BRAND NEW!!! {Name of Xbox Game here} XBOX GAME!!*SEALED*

Text: This superhit game for Microsoft XBOX system is brand new and is in original shrink wrap.

I accept payment only by Paypal or money orders.

Buyer pays a fixed shipping and handling charge of {insert shipping charge here}, and I will deliver by USPS first-class mail as soon as the payment is received. Payment should be sent within 10 days of the end of the auction.

I will ship to US and Canada only.
Happy Bidding.

For the free shipping treatment, this paragraph read:

There is no shipping and handling fee for this Xbox game. I will deliver by USPS first-class mail as soon as the payment is received. Payment should be sent within 10 days of the end of the auction.

I will ship to US and Canada only.
Happy Bidding.

References

Cox, James C., Bruce Roberson and Vernon L. Smith (1982): "Theory and Behavior of Single Object Auctions," *Research Experimental Economics, Vol. 2*, Vernon L. Smith, editor, Greenwich, JAI Press.

Cox, James C., Vernon L. Smith and James M. Walker (1983): "A Test that Discriminates between two models os the Dutch-First Auction Non-isomorphism" *Journal of Economic Behavior and Organization*, 14, 205-219.

Davis, Douglas D., and Charles A. Holt (1993): *Experimental Economics*, Princeton, Princeton University Press.

Ellison, Glenn, and Sara F. Ellison (2004): "Search, Obfuscation, and Price Elasticities on the Internet," Working Paper, MIT.

Gabaix, Xavier, and David Laibson (2005): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, forthcoming.

Kagel, John H. (1995): "Auctions: A Survey of Experimental Research," *Handbook of Experimental Economics*, John H. Kagel and Alvin E. Roth, editors, Princeton, Princeton University Press, 501-585.

Kahneman, Daniel, and Amos Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263-291.

Kahneman, Daniel, and Amos Tversky (1984): "Choices, Values, and Frames," *American Psychologist*, 39, 341-350.

Katkar, Rama, and David H. Reiley (2005): "Public Versus Secret Reserve Prices in eBay Auctions: Results from a Pokémon Field Experiment," Working Paper, Vanderbilt University.

Krishna, Vijay (2002): *Auction Theory*, San Diego, Academic Press.

List, John A., and David Lucking-Reiley (2000): "Demand Reduction in Multi-Unit Auctions: Evidence from a Sportscard Field Experiment," *American Economic Review*, 90, 961-972.

Lucking-Reiley, David (1999): "Using Field Experiments to Test Equivalence Between Auction Formats: Magic on the Internet," *American Economic Review*, 89, 1063-1080.

Morwitz, Vicki G., Eric A. Greenleaf, and Eric J. Johnson (1998): "Divide and Prosper: Consumers' Reaction to Partitioned Prices," *Journal of Marketing Research*, 35, 453-463.

Reiley, David H. (2005): "Field Experiments on the Effects of Reserve Prices in Auctions: More Magic on the Internet," *RAND Journal of Economics*, forthcoming.

Resnick, Paul, and Richard Zeckhauser (2002): "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System.," *The Economics of the Internet and E-Commerce, Advances in Applied Microeconomics, Vol. 11*, Michael R. Baye, editor, Amsterdam, Elsevier Science, 127-157.

Roth, Alvin E., and Axel Ockenfels (2002): "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet," *American Economic Review*, 92, 1093-1103.

Smith, Michael D. and Erik Brynjolfsson (2001): "Consumer Decision-making at an Internet Shopbot: Brand Still Matters," *Journal of Industrial Economics*, 49, 541-558.

Thaler, Richard H. (1985): "Mental Accounting and Consumer Choice," *Marketing Science*, 4, 199-214.

Tversky, Amos, and Daniel Kahneman (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106, 1039-1061.

American Law & Economics Association Annual Meetings

Year 2006 *Paper* 52

Bundling and Consumer Misperception

Oren Bar-Gill
NYU Law School

This working paper site is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the publisher's permission.

http://law.bepress.com/alea/16th/art52

Copyright ©2006 by the author.

Bundling and Consumer Misperception

Oren Bar-Gill†

This Essay studies bundling of two (or more) products as a strategic response to consumer misperception. In contrast to the bundling and tying studied in the antitrust literature—strategies used by a seller with market power in market A trying to leverage its market power into market B—bundling in response to consumer misperception may occur in intensely competitive markets. The analysis demonstrates that such competitive bundling can be either welfare enhancing or welfare reducing. The Essay considers several "unbundling policies" that can protect consumers and increase welfare in markets where bundling is undesirable.

I. INTRODUCTION

Consumer misperception of the costs and benefits associated with a product or service is prevalent. It can be the product of imperfect information or imperfect rationality (or both). It can be independent of any action taken by sellers. It can be instigated by sellers. And it can be mitigated by sellers.[1]

---

† Assistant Professor of Law, New York University School of Law. This Essay greatly benefited from comments and criticisms by Barry Adler, Jennifer Arlen, Lucian Bebchuk, Omri Ben-Shahar, Bob Cooter, Chaim Fershtman, Eleanor Fox, Clay Gillette, Marcel Kahan, Ehud Kamar, Ronald Mann, Luis Rayo, seminar participants at Bar-Ilan University, NYU, and the University of Texas, and conference participants at the University of Chicago Law School conference on *Homo Economicus*, *Homo Myopicus*, and the Law and Economics of Consumer Choice. Finally, I thank the NYU School of Law for generous financial support.

[1] A seller offering a superior product has every incentive to ensure that consumers appreciate the superiority of her product. Specifically, such a seller has every incentive to correct any misperception consumers might have about the shortcoming of a competing product. But correcting consumers' misperceptions is costly. A seller might thus choose to ride the tide of consumer misperception and offer an inferior product, rather than convince consumers that a superior product justifies a higher price. The concern that sellers will often lack the incentive to educate consumers is reinforced by the public-good nature of such educational efforts. If a seller succeeds in correcting consumers' misperceptions, competitors will be quick to adapt their products to the changed demand. And the seller, who brought about this desirable change in demand, will not be able to recoup her investment in educating consumers. See, for example, Ian Ayres and Barry Nalebuff, *In Praise of Honest Pricing*, 45 MIT Sloan Management Rev 24, 28 (Fall 2003) (mentioning the "first-mover disadvantage" that occurs when companies "change the environment on their own" by supplying information to the marketplace); Howard Beales, Richard Craswell, and Steven C. Salop, *The Efficient Regulation of Consumer Information*, 24 J L & Econ 491, 503–04, 506 (1981) (finding that there is often "an undersupply of general information" about products, because when a company furnishes its consumers with information about a particular product, its competitors end up benefiting too). In a recent contribution Xavier Gabaix and David Laibson argue that under some conditions sellers will shroud unattractive product attributes even when advertising costs are zero. See Xavier Gabaix and David Laibson, *Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets*, MIT Econ Working Paper No 05-18 (2005), available online at http://ssrn.com/abstract=728545 (visited Nov 21, 2005). In some cases, signaling, for example, through warranties, can alleviate consumer misperception. See Michael Spence, *Consumer Misperceptions, Product Failure and Producer Liability*, 44 Rev Econ Stud 561, 569 (1977) (finding that product guar-

This Essay takes consumer misperception as given and studies one common strategy employed by sellers facing such misperception: the bundling strategy. "Bundling" in this Essay is used in a somewhat broader sense than is conventional in the antitrust and industrial organization literatures.[2] I define the bundling of products A and B to include any case where a consumer purchasing product A from seller X has a sufficiently strong incentive to purchase product B from the same seller. In a second significant departure from the antitrust and industrial organization literatures, I focus on bundling by sellers operating in competitive markets.

Consider first consumer misperception about the value of a product. To fix ideas assume that a consumer underestimates the amount of in-home printing she will choose to do if she had a printer at home, thus underestimating the value to her of owning a printer. Such underestimation of value and of use will also lead the consumer to underestimate the number of ink cartridges she will purchase over the life of the printer. For instance, the consumer may estimate that she will need fifty ink cartridges, when in fact she will need one-hundred cartridges.[3] The argument is that under these assumptions a seller offering printers only will find it hard to compete with a seller who bundles printers and ink, that is, who, through technological compatibility constraints and/or intellectual property protection, forces consumers who bought its printers to also purchase its ink cartridges.

The competitive advantage of the bundling seller can be explained as follows. In a competitive market a seller offering only printers will have to price its printers at the marginal cost of a printer, say $1000. Consumers who buy printers from this seller will know that they will have to buy their ink from another seller at the marginal cost of ink, say $10. Accordingly, the total cost of printing perceived by a consumer purchasing a printer from the printer-only seller is $1500 (i.e., $1000 + 50 × $10). (The actual, as opposed to perceived, total cost of printing is $1000 + 100 × $10 = $2000.)

Now consider a bundling seller. This seller may offer the same (or equivalent) printer at a below-cost price of $500 and cover its losses by charging $15 per ink cartridge. The total cost of printing perceived by a consumer purchasing a printer from the bundling seller is $1250 (= $500 + 50 × $15). (The actual, as opposed to perceived, total cost of printing is

---

antees may act as signals to consumers because they are "costly to the seller" and are "systematically related to product liability").

[2] See, for example, Jean Tirole, *The Theory of Industrial Organization* 333–35 (MIT 1988).

[3] For expositional simplicity this example and most of the analysis in the Essay assumes inelastic demand. A more general model with elastic demand is developed and studied in the appendix. The main results are qualitatively unchanged when elastic demand is introduced. See Part II.D.

$500 + 100 × $15 = $2000, as in the no-bundling case.[4]) Given consumer underestimation of ink usage, sellers in a competitive market must bundle printers and ink.

Overestimation of value and of use can similarly induce a bundling response by sellers. Consider the health club market. Sellers can charge a per-visit fee. Sellers can also offer a one-year subscription, which can be viewed as an intertemporal bundle (access to the club in period 1 is bundled with access in period 2). For consumers who overestimate the number of visits they will make to the health club the bundle/subscription will be the preferred option. Accordingly, in a competitive market health clubs who fail to offer subscriptions will be at a disadvantage.

The welfare implications of bundling depend on the type of misperception that triggers the bundling response. Absent bundling, underestimation of value leads to too little trade. In the printers and ink example this means that too few printers will be purchased. Bundling, with its accompanying backloaded pricing, generates an underestimation of cost that offsets the underestimation of value. Bundling restores efficiency. Overestimation of value, on the other hand, leads to excessive trade. Bundling exacerbates this inefficiency. Absent bundling with per-product marginal cost pricing, the overestimation of value is partially offset by the overestimation of cost. Bundling, with its accompanying front-loaded pricing, eliminates this beneficial offsetting effect.

The bundling strategy has distributional effects as well. When bundling is a response to underestimation of value and of use, high value/use consumers end up cross-subsidizing low value/use consumers. When bundling is a response to overestimation of value and of use, low value/use consumers end up cross-subsidizing high use consumers. The welfare implications of these distributional effects depends on the identity of the high value/use and low value/use groups.

Misperception of value is not the only type of misperception that can trigger a bundling response. Price misperception can similarly force bundling of the product whose price is misperceived and another product whose price is accurately perceived. The efficiency implications are straightforward. Bundling exacerbates the overconsumption problem created by underestimation of price. When overestimation of price leads to underconsumption, bundling alleviates this inefficiency.

The main goal of this Essay is to argue that competitive bundling in response to persistent consumer misperception is both predicted in theory

---

[4] In a competitive market, the total price collected by the bundling seller cannot exceed the total cost, $2,000 (= $1000 [cost of a printer] + 100 [actual cost of an ink cartridge] × $10 [cost of an ink cartridge]). Neither can the total price fall below $2,000, otherwise the seller will lose money.

and observed in practice. I therefore begin, in Part II, by developing a theory of bundling in response to consumer misperception. I then proceed in Part III to briefly consider three real world examples of misperception-based bundling. I begin with intertemporal bundling in subscription markets. I proceed to discuss the credit card market which exhibits bundling between transacting and borrowing services as well as intertemporal bundling. And I end with the cell phone market where phones/handsets are bundled with calling plans.

While mainly descriptive, the analysis in this Essay has normative and prescriptive implications. I show that bundling has both efficiency and distributional consequences. The feasibility of bundling can either increase or reduce welfare, depending largely on the type of misperception that triggers the bundling response. When bundling reduces welfare, regulation that discourages bundling may provide a valuable tool for policymakers. Part V considers several unbundling policies.[5]

This Essay studies bundling in competitive markets. It shows that consumer misperception can lead to welfare-reducing bundling even in competitive markets. The analysis thus departs from the legal and economic literatures on bundling and tying that have focused on concentrated markets.[6] An important exception is a recent article by David Evans and Michael Salinger that studies bundling in competitive markets.[7] Evans and Salinger, however, highlight the potential cost-based *efficiency* of bundling. This Essay explores the potential *efficiency* and *inefficiency* of bundling, when the bundling strategy is adopted in response to consumer misperception.

It is important to emphasize at the outset that the proposed account of bundling, and of the pricing of the bundle and its components, as a response to consumer misperception is not offered as an exhaustive or even a dominant account. There are other important explanations for the bundling strategy that have nothing to do with consumer misperception. In particular, as

---

[5] The analysis has a second policy implication. The existence of product bundles and the pricing of these bundles can be used as indicators of persistent consumer misperception in the examined market. However, bundles with similar pricing patterns may also be the product of other economic forces. Therefore, observing such bundles and pricing patterns should not be considered decisive evidence of misperception.

[6] See, for example, Michael D. Whinston, *Tying, Foreclosure, and Exclusion*, 80 Am Econ Rev 837, 838 (1990); Dennis W. Carlton and Michael Waldman, *The Strategic Use of Tying to Preserve and Create Market Power in Evolving Industries*, 33 RAND J Econ 194, 198 (2002); Barry Nalebuff, *Bundling as a Barrier to Entry*, 119 Q J Econ 159, 183 (2004).

[7] David S. Evans and Michael Salinger, *Why Do Firms Bundle and Tie? Evidence from Competitive Markets and Implications for Tying Law*, 22 Yale J Reg 37, 42 (2005). See also Richard Craswell, *Tying Requirements in Competitive Markets: The Consumer Protection Issues*, 62 BU L Rev 661, 681-87 (1982) (discussing cases in which bundling is likely to be efficient).

noted above, many bundles can be justified on cost-saving grounds. Moreover, many observed bundles can be explained by a combination of the misperception-based and cost-based accounts.

Market responses to consumers' imperfect rationality are the subject of increasing attention in the behavioral economics literature.[8] An important recent contribution by Stefano DellaVigna and Ulrike Malmendier develops a model of optimal two-part tariff pricing with consumer misperception.[9] My analysis of how bundled components should optimally be priced builds on the DellaVigna and Malmendier model. Della Vigna and Malmendier also discuss many of the markets that feature in this Essay. DellaVigna and Malmendier, however, take the existence of the bundle as given.[10] The main contribution of this Essay is to endogenize bundling—to present bundling as an important strategic response to consumer misperception. I also discuss welfare and policy implications that are not identified in previous work.[11]

Finally, the potential role of bundling as a strategic response to consumers' imperfect rationality has been recognized in two important early articles by Richard Thaler and Richard Craswell. Pioneering the field of behavioral economics, Thaler's seminal contribution shows how mental accounting (by consumers), and specifically the framing and coding of multiple gains and losses, can lead sellers to adopt a bundling strategy.[12] Craswell, working at the intersection of antitrust law and consumer protection law, identifies the viability of misperception-driven bundling in com-

---

[8] See, for example, Gabaix and Laibson, MIT Econ Working Paper No 05-18 (cited in note 1) (discussing sellers' frequent practice of shrouding information about their products when that information is not readily available to the public); Stefano DellaVigna and Ulrike Malmendier, *Paying Not to Go to the Gym* 2 (forthcoming), online at http://faculty-gsb.stanford.edu/malmendier/personal_page/Papers/gymemp05-04-20.pdf (visited Nov 21, 2005) (studying consumer behavior regarding gym membership); Sharon M. Oster and Fiona M. Scott Morton, *Behavioral Decision-Making: An Application to the Setting of Magazine Subscription Prices*, SSRN Working Paper Series 5 (unpublished manuscript 2004), online at http://ssrn.com/abstract=467222 (visited Nov 21, 2005).

[9] See Stefano DellaVigna and Ulrike Malmendier, *Contract Design and Self-Control: Theory and Evidence*, 119 Q J Econ 353, 357 (2004).

[10] An important exception is DellaVigna and Malmendier's analysis of endogenous switching costs. Id at 385. Previous analyses of subscription contracts may also be interpreted as studying endogenous intertemporal bundling. See id at 392 (collecting examples of intertemporal bundling from other industries); DellaVigna and Malmendier, *Paying Not to Go to the Gym* at 1 (cited in note 8) (modeling a choice between monthly and annual health club subscriptions); Oster and Morton, *Behavioral Decision-Making* (cited in note 8) (magazine subscriptions). Intertemporal bundling is discussed in Part IV.A.

[11] Specifically, DellaVigna and Malmendier ask how misperception affects welfare, while I ask how bundling affects welfare (taking misperception as given).

[12] See Richard Thaler, *Mental Accounting and Consumer Choice*, 4 Marketing Sci 199, 208–09 (1985).

petitive markets.[13] I generalize Thaler's and Craswell's insights and explore their welfare and policy implications.

II. BUNDLING IN RESPONSE TO CONSUMER MISPERCEPTION

A. Printers and Ink

When consumers misperceive the costs or benefits of one product, competition may force sellers to bundle this product with another product. I allow for separate pricing of the two products, but show that the competitive response to consumer misperception will often entail a single price. It is important to note that an effective bundle can exist even when two (or more) distinct products are separately priced. This bundle can be sustained through technology (i.e., compatibility constraints), law (i.e., contractual obligation), or simply economic or psychological switching/transaction costs.

Consider a competitive market for printers. Assume, however, that once a consumer buys a brand X printer she can buy ink cartridges only from X (as a matter of technological compatibility). How will sellers price their product? Or, more accurately, how will they price their two bundled products: printers and cartridges?[14]

Take a specific example. Let the cost of a printer be 1000 and the cost of an ink cartridge be 10. Assume that the seller knows that an average consumer will buy 100 cartridges over the life of the printer. (Inelastic demand is assumed for expositional simplicity. An elastic demand extension is developed in the appendix; the results are summarized in Part II.C.) The total per-consumer cost is thus 2000. If consumers are homogeneous in their printing practices,[15] and fully aware of their expected use of the printer, then a continuum of printer-ink price pairs is possible. To see this, let p_p and p_i denote the price of a printer and of an ink cartridge, respectively. The consumer thus expects to pay a total price of $P = p_p + 100 \cdot p_i$. Competition guarantees that the total price P equals the total cost to the seller, namely, $P = p_p + 100 \cdot p_i = 2000$. This competitive pricing condition is satisfied by per-product marginal cost pricing: $p_p = 1000$ and $p_i = 10$. But it is also satisfied, for example, by $p_p = 0$ and $p_i = 20$ and by $p_p = 2000$ and $p_i = 0$.

---

13 See Craswell, 62 BU L Rev at 671–81 (cited in note 7).

14 The analysis below directly applies to any market where a durable product is bundled with replacement parts or service. And, as explained below, it also applies more broadly to any misperception-based bundling of two goods, services or components.

15 The homogeneous consumer case can be viewed as homogeneity within a class of consumers that has been segmented by sellers.

Consumer heterogeneity breaks the indifference between the infinity of possible price combinations. Consider two types of consumers: high use consumers who will buy 110 cartridges on average and low use consumers who will buy 90 cartridges on average. Assume an equal number of high and low use consumers. The $p_p = 0$, $p_i = 20$ price pair is unattractive to high use consumers. These consumers expect to pay 2200 under this pricing scheme, and will thus be quick to choose a seller offering a more balanced price combination. Specifically, high use consumers will prefer per-product marginal cost pricing, under which they expect to pay 2100. Conversely, the $p_p = 2000$ and $p_i = 0$ price pair is unattractive to low use consumers. These consumers expect to pay 2000 under this pricing scheme, and will thus prefer per-product marginal cost pricing, under which they expect to pay 1900. Heterogeneity will thus lead sellers to price both their printers and their ink cartridges at marginal cost.[16] Perhaps more importantly, elastic demand also breaks the indifference between the possible price combinations and leads to marginal cost pricing (see Appendix).

B. Underestimation of Value/Use

Now assume that consumers are myopic and systematically underestimate the amount of printing they will do and thus the number of ink cartridges that they will buy. How does consumer misperception affect the above result that competition will lead to per-product marginal cost pricing?

For expositional clarity, I return to the unrealistic benchmark market where consumers are homogeneous with respect to their printing practices. Absent consumer misperception any price combination satisfying $p_p + 100 \cdot p_i = 2000$ can persist in a competitive market. What if consumers mistakenly believe that they will need 50 ink cartridges, rather than 100 cartridges—the true number of cartridges that they will use? Compare the perceived attractiveness of the three price pairs considered above. With the $p_p = 2000$, $p_i = 0$ pricing scheme the consumer will perceive a total price of 2000. With the $p_p = 1000$, $p_i = 10$ pricing scheme the consumer will perceive a total price of 1500. And with the $p_p = 0$, $p_i = 20$ pricing scheme the consumer will perceive a total price of 1000. Since sellers get the same total price under the three pricing schemes, they will choose the $p_p = 0$, $p_i = 20$ scheme.[17]

---

[16] Other equilibria where high use consumers are offered one contract/pricing scheme and low use consumers are offered another contract/pricing scheme are theoretically possible (if sellers can make low-price ink compatible only with high-price printers). In a competitive market, however, sellers have no incentive to deviate from per-product marginal cost pricing.

[17] In theory sellers might even set a negative printer price, p_p, and raise the ink price, p_i, above 20. In practice, however, negative prices pose too big a temptation for strategic behavior.

This result is robust to the introduction of consumer heterogeneity. As before, assume that there are two types of consumers: high use consumers who will buy 110 cartridges on average and low use consumers who will buy 90 cartridges on average. Introducing consumer misperception, assume that the high use consumers think they will buy 60, rather than 110, ink cartridges, and that the low use consumers think they will buy 40, rather than 90, ink cartridges.

Compare the perceived attractiveness of the three price pairs considered above. With the $p_p = 2000$, $p_i = 0$ pricing scheme both high use and low use consumers will perceive a total price of 2000. With the $p_p = 1000$, $p_i = 10$ pricing scheme high use consumers will perceive a total price of 1600 and low use consumers will perceive a total price of 1400. And with the $p_p = 0$, $p_i = 20$ pricing scheme high use consumers will perceive a total price of 1200 and low use consumers will perceive a total price of 800. Since sellers get the same total price under the three pricing schemes, they will choose the $p_p = 0$, $p_i = 20$ scheme.[18]

The preceding analysis focused on the pricing of printers and ink, taking the existence of the printer-ink bundle as given. But the formation of bundles in itself is an endogenous deliberate strategy. A main theme of this Essay is that the bundling strategy is an effective, often inevitable, response to consumer misperception. The optimal pricing analysis demonstrated the dominance of the $p_p = 0$, $p_i = 20$ scheme. This pricing scheme cannot survive without effective bundling of printers and ink. Absent such bundling, a consumer who received a free printer under the $p_p = 0$, $p_i = 20$ scheme will buy ink at the marginal cost of $p_i = 10$ from an independent ink seller. Foreseeing this dynamic no one will adopt the free printer–expensive ink tactic.

Making the $p_p = 0$, $p_i = 20$ scheme viable, however, is an attractive prospect. As shown above, if viable, this scheme dominates all other pricing schemes. Specifically, it dominates the pricing scheme that inevitably emerges absent bundling, the $p_p = 1000$, $p_i = 10$, marginal-cost-pricing scheme. Sellers thus have a powerful incentive to bundle printers and ink. Indeed, sellers employ technological compatibility constraints coupled with

---

[18] The analysis in the text considers only heterogeneity with respect to printer use. Another important dimension of heterogeneity, whose implications are not explored here, is heterogeneity with respect to the level of the bias/misperception (including, for example, the case where some consumers underestimate use and some overestimate use). For an analysis of this type of heterogeneity in related contexts, see generally DellaVigna and Malmendier, *Paying Not to Go to the Gym* (forthcoming) (cited in note 8); DellaVigna and Malmendier, 119 Q J Econ at 360 (cited in note 9) (setting a variable that represents a given consumer's estimate of his consumption of a product at a given time); see also John Haltiwanger and Michael Waldman, *Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity*, 75 Am Econ Rev 326, 328–33 (1985).

intellectual property protection to secure effective bundling of printers and ink.[19]

What are the welfare implications of the bundling of printers and ink cartridges and specifically of the $p_p = 0$, $p_i = 20$ pricing scheme that such bundling entails? It may seem, at first blush, that there are no welfare implications. After all, under all three pricing schemes— $p_p = 2000$, $p_i = 0$; $p_p = 1000$, $p_i = 10$; and $p_p = 0$, $p_i = 20$ —the average consumer ends up paying the exact same amount: 2000. But the three pricing schemes are not welfare-neutral.

First, while the *average* consumer will end up paying the same amount under the different schemes, some consumers will benefit and some consumers will lose. Table 1 lists the total (and per-use/cartridge) payments of high use and low use consumers under each of the three pricing schemes.

Table 1: The differential effect of pricing distortions on the two consumer types

	High use consumers pay	Low use consumers pay
$p_p = 2000$, $p_i = 0$	Total: 2000 (Per use/cartridge: 18.2) Perceived total: 2000	Total: 2000 (Per use/cartridge: 22.2) Perceived total: 2000
$p_p = 1000$, $p_i = 10$	Total: 2100 (Per use/cartridge: 19.1) Perceived total: 1600	Total: 1900 (Per use/cartridge: 21.1) Perceived total: 1400
$p_p = 0$, $p_i = 20$	Total: 2200 (Per use/cartridge: 20) Perceived total: 1200	Total: 1800 (Per use/cartridge: 20) Perceived total: 800

Table 1 shows that low use consumers benefit from a move to the $p_p = 0$, $p_i = 20$ pricing scheme, while high use consumers lose from such a move. This distributional effect can be seen as either good or bad, depending on the identity of the high use and low use consumers, but it is not welfare neutral.[20]

---

[19] Hall reports that printer manufacturers control over 90 percent of the ink market. See Robert E. Hall, *The Inkjet Aftermarket: An Economic Analysis* 23 (unpublished manuscript 1997) (on file with author). Also consistent with the above analysis, Hall suggests that ink cartridges are "hugely profitable." Id at 4.

[20] In fact, even if the two buyer types are identical in every aspect other than their printing practices, still the distributional effect of the different pricing schemes will not be welfare neutral. With decreasing marginal utility of money, a loss to one individual is not perfectly balanced by a commensurate gain to another individual: Consider two identical individuals (with the same level of wealth), A and B. When a sum X is transferred from A to B, the reduction in A's utility is greater than the increase in

A second welfare implication derives from the distortion in the number of printers that consumers buy. Underestimation of value/use naturally leads to the purchase of too few printers, at least absent bundling. To see this let v denote the per-cartridge value of printing to the average consumer. The total value of printing to the average consumer is thus $100v$. Since the total cost of printing is 2000 (recall that the cost of a printer is 1000 and the cost of an ink cartridge is 10), it is efficient for a consumer to purchase a printer whenever $100v > 2000$ or $v > 20$. With underestimation of value/use the perceived total value of printing is $50v$ ($< 100v$). Under the $p_p = 1000$, $p_i = 10$, marginal-cost-pricing scheme the average consumer perceives a total price of printing of 1500, and will thus purchase a printer whenever $50v > 1500$ or $v > 30$. In particular, efficient purchases will not occur whenever $20 < v < 30$.

Bundling cures this problem. With bundling coupled with the $p_p = 0$, $p_i = 20$ pricing scheme the average consumer perceives a total price of printing of 1000, and will thus purchase a printer whenever $50v > 1000$ or $v > 20$. Efficiency is restored. The underestimation of value is perfectly offset by the underestimation of total price. Bundling in response to underestimation is welfare enhancing.

C. Overestimation of Value/Use

What would be the market response to the opposite kind of misperception—to overestimation, rather than underestimation, of use? While overestimation is less likely in the printer-ink context, I continue with this example for ease of exposition.

Again, I return to the unrealistic benchmark market where consumers are homogeneous with respect to their printing practices. Assume that consumers mistakenly believe that they will need 150 ink cartridges, rather than 100 cartridges—the true number of cartridges that they will use. Compare the perceived attractiveness of the three price pairs considered above. With the $p_p = 2000$, $p_i = 0$ pricing scheme the consumer will perceive a total price of 2000. With the $p_p = 1000$, $p_i = 10$ pricing scheme the consumer will perceive a total price of 2500. And with the $p_p = 0$, $p_i = 20$ pricing scheme the consumer will perceive a total price of 3000. Since sellers get the same total price under the three pricing schemes, they will choose the $p_p = 2000$, $p_i = 0$ scheme.[21]

The optimal pricing scheme with overestimation of use—$p_p = 2000$, $p_i = 0$—is diametrically opposite to the optimal pricing scheme with un-

B's utility. For a definition and discussion of the decreasing marginal utility assumption see Richard A. Posner, *Economic Analysis of Law* 10–11 (Aspen 6th ed 2003).

[21] As in the preceding subpart, this result is robust to the introduction of consumer heterogeneity.

derestimation of use— $p_p = 0$, $p_i = 20$. The bundling of printers and ink, however, is equally necessary to support this very different pricing scheme. Absent such bundling, the consumer would purchase a printer at the marginal cost of $p_p = 1000$ from an independent printer seller and then pick-up free ink from the seller offering the $p_p = 2000$, $p_i = 0$ scheme. Anticipating this dynamic, no one will adopt the expensive printer/free ink tactic. As with underestimation of use, here too sellers have a strong incentive to make the $p_p = 2000$, $p_i = 0$ scheme viable, by bundling printers and ink. Arguably, effective bundling with $p_p = 2000$, $p_i = 0$ pricing can be more easily achieved, as compared to effective bundling with $p_p = 0$, $p_i = 20$ pricing (as was needed with consumer underestimation of use). All sellers need to do is to price discriminate, in the sale of ink, between consumers who purchased their printer from the same seller and consumers who purchased their printer from a different seller. (Of course, ink arbitrage must also be prevented: a person who purchased a printer from an expensive printer/cheap ink seller must not be allowed to buy a million ink cartridges at the low price and then resell most of these cartridges to consumers who purchased their printer from another seller.)

What are the welfare implications of the bundling of printers and ink cartridges, this time with the $p_p = 2000$, $p_i = 0$ pricing scheme? As before, while the *average* consumer will end up paying the same amount under the different schemes, some consumers will benefit and some consumers will lose. The distributional implications, however, are the reverse of those resulting from the $p_p = 0$, $p_i = 20$ scheme adopted in response to underestimation of value/use. High-value/use consumers benefit from a move to the $p_p = 2000$, $p_i = 0$ pricing scheme, while low-value/use consumers lose from such a move.

Now to efficiency: while bundling increased efficiency when adopted in response to underestimation of value/use, the opposite is true when bundling responds to overestimation of value/use. To see this, let v denote the per-cartridge value of printing to the average consumer. The total value of printing to the average consumer is thus $100v$. Since the total cost of printing is 2000 (recall that the cost of a printer is 1000 and the cost of an ink cartridge is 10), it is efficient for a consumer to purchase a printer whenever $100v > 2000$ or $v > 20$.

With overestimation of value/use the perceived total value of printing is $150v$ ($> 100v$). Overestimation of value/use naturally leads to the purchase of too many printers even under marginal cost pricing. Under the $p_p = 1000$, $p_i = 10$, marginal-cost-pricing scheme the average consumer perceives a total price of printing of 2500, and will thus purchase a printer whenever $150v > 2500$ or $v > 16.6$. In particular, inefficient purchases will occur whenever $16.6 < v < 20$. While overestimation produces inefficiency

even absent bundling, bundling exacerbates this inefficiency. With bundling coupled with the $p_p = 2000$, $p_i = 0$ pricing scheme, the consumer will purchase a printer whenever $150v > 2000$ or $v > 13.3$. Namely, the range of inefficient purchases increases from $16.6 < v < 20$ without bundling to $13.3 < v < 20$ with bundling.

Without bundling, marginal-cost pricing leads to overestimation of the total price in addition to the overestimation of value. And two overestimations are better than one: The overestimation of the total price partially offsets the overestimation of value, thus reducing the inefficiency.[22] With bundling, sellers frontload the entire cost of printing onto the price of the printer, preventing overestimation of total price by the consumer. By removing the beneficial offsetting effect of price overestimation, bundling exacerbates the excessive trade problem.

D. Elastic Demand

The preceding analysis assumed that the demand for ink is inelastic, namely that the number of ink cartridges that the consumer will purchase, as well as the number of ink cartridges that the consumer perceives that she will purchase, does not depend on the price of ink. This assumption, while useful for demonstrating the main implications of misperception-based bundling, is clearly unrealistic. In this section I describe the results of a more general elastic demand model, which is developed and analyzed in the appendix. These results confirm that the main features of misperception-based bundling, as developed in Parts II.B and II.C above, are robust to the introduction of elastic demand.

In the inelastic demand model the bundling seller chose extreme pricing schemes— $p_p = 0$, $p_i = 20$ in the underestimation case and $p_p = 2000$, $p_i = 0$ in the overestimation case. Elastic demand introduces a new factor that might lead to less extreme pricing schemes. With elastic demand, increasing (decreasing) the price of ink above (below) marginal cost leads to distortions in the amount of ink that is actually purchased (affecting also the amount of ink that the consumer thinks she will purchase). In equilibrium this inefficiency will be balanced against the forces that push the price of ink away from marginal cost. Nevertheless, the main results developed in the preceding sections continue to hold: when consumers underestimate value/use, ink will be priced above marginal cost and printers will be priced below marginal cost; and when consumers overestimate

---

[22] The overestimation of total price *completely* offsets the overestimation of value under the $p_p = 0$, $p_i = 20$ scheme. Under this pricing scheme the consumer will purchase a printer whenever $150v > 3000$ or $v > 20$.

value/use, ink will be priced below marginal cost and printers will be priced above marginal cost.

The elastic demand model also qualifies the welfare implications derived in the inelastic demand model. In the inelastic demand model, bundling in response to underestimation of value is unambiguously beneficial, since it improves the printer purchasing decision. In the elastic demand model, this improvement in the printer purchasing decision must be weighed against the distortion in the ink purchasing decision caused by the deviation from marginal cost pricing. The normative appeal of bundling is, therefore, reduced.

When consumers overestimate the value/use of printing, bundling was shown to be welfare-reducing in the inelastic demand model. This result is strengthened in the elastic demand model where distortions in the ink purchasing decision caused by the deviation from marginal cost pricing constitute an additional cost of bundling.

E. Bundling in Response to Price Misperception

Bundling is a strategic response to consumer misperception. This idea has been demonstrated above, focusing on consumer misperception about the value of the product. Another type of misperception that can induce a bundling response is price misperception.

Pure price misperception is arguably rare. Price is usually the most salient feature of a product—the one thing consumers can be expected to perceive with reasonable accuracy. Still, consumers misperceive prices, especially when these prices are not immediately due. Returning to the printers and ink example, Robert Hall found that "people buy inkjet printers without information about [the cost of replacement ink]."[23] Credit purchases provide another example. Many consumers systematically underestimate the total price they will end up paying simply because they do not understand how fast interest accrues.

Consider two products (or two components) A and B that can be separately manufactured and sold at unit costs of c_A and c_B, respectively. The value to consumers of product A is v_A, and the value to consumers of product B is v_B. Consumers are assumed to be homogeneous. Absent bundling in a competitive market product A would be sold at a price of $p_A = c_A$, and product B would be sold at a price of $p_B = c_B$. Under reasonable assumptions this would be the market outcome regardless of consumer perception or misperception.

---

[23] Hall, *Inkjet Aftermarket* at 22–23 (cited in note 19). This lack of information about price is independent of any misperception about use.

Now allow for bundling. Without loss of generality assume that consumers misperceive the price of product B to be $\hat{p}_B = \delta \cdot p_B$. I begin with underestimation of p_B, or $\delta < 1$. Absent bundling, the equilibrium will exhibit (per-product) marginal cost pricing, i.e., $(p_A, p_B) = (c_A, c_B)$. This equilibrium cannot be sustained when bundling is feasible. Sellers offering such (per-product) marginal cost pricing would lose business to a competitor that sells the two products as a bundle A-B and sets $p_A < c_A$ and $p_B > c_B$. In fact, at equilibrium $(p_A, p_B) = (0, c_A + c_B)$.

To see why, note that the perceived cost of the bundle is $\hat{P} = p_A + \hat{p}_B = p_A + \delta \cdot p_B$. Competition will force equilibrium prices (p_A, p_B) that minimize $\hat{P}$ subject to the constraint $p_A + p_B \geq c_A + c_B$ (I assume non-negative prices, $p_A, p_B \geq 0$). The $(p_A, p_B) = (0, c_A + c_B)$ result follows. The intuition underlying this result is straightforward. Every dollar subtracted from the price of product A is perceived by consumers as a \$1 benefit; every dollar added to the price of product B is perceived by consumers as a $\$\delta (< \$1)$ cost. Sellers will thus transfer as much of the price as possible from product A to product B.

It bears emphasis that a seller who fails to adopt the bundling strategy will be driven out of the market. A seller who offers product A alone will have to charge $p_A \geq c_A$. Similarly, a seller who offers product B alone will have to charge $p_B \geq c_B$. The bundled product would always appear more attractive to consumers.

Now assume that consumers overestimate the price of product B, or $\delta > 1$. Again, when bundling is feasible (per-product) marginal-cost pricing cannot be sustained in equilibrium. As before the perceived cost of the bundle is $\hat{P} = p_A + \hat{p}_B = p_A + \delta \cdot p_B$. And again competition will force equilibrium prices (p_A, p_B) that minimize $\hat{P}$ subject to the constraint $p_A + p_B \geq c_A + c_B$. But now, with $\delta > 1$, this minimization implies $(p_A, p_B) = (c_A + c_B, 0)$. A parallel intuition explains this result. Every dollar subtracted from the price of product B is perceived by consumers as a $\$\delta (> \$1)$ benefit; every dollar added to the price of product A is perceived by consumers as a \$1 cost. Sellers will thus transfer as much of the price as possible from product B to product A.

The efficiency implications are straightforward. Bundling exacerbates the overconsumption problem created by underestimation of price. Overestimation of price leads to underconsumption absent bundling. This inefficiency is completely alleviated when bundling is feasible.

III. THREE EXAMPLES

A. Subscription Services

One special type of bundling that directly responds to consumer misperception is the intertemporal bundling achieved through multi-period subscriptions. In intertemporal bundling the only difference between the bundled components is the timing. For example, a year-long subscription—to a magazine, a wireless or landline phone service, an ISP, or a health club—provides the same service every month throughout the year. Why are multi-period subscriptions so common? Why not sell only single-period products or services? Consumer misperception provides an answer.[24]

Consider a health club offering access to its facilities on a per-visit basis. In a competitive market this health club will have to set its admission fee equal to marginal cost, say $10 per visit. Assume that the average consumer overestimates her future use of the health club: instead of ten visits per year the consumer anticipates that she will make twenty visits per year.[25] Under per-period marginal cost pricing the consumer expects a total cost of $200. Now assume that a competitor offers—instead of or in addition to the per-visit fee—a yearlong subscription for a price of $100. (I assume that health clubs know the number of visits made by an average consumer.) The consumer will clearly prefer the multi-period subscription over per-period admission.

Many subscription services charge a one-time subscription fee as described in the preceding health club example. Other subscription services charge a per-period fee, but follow a fee schedule very different from per-period marginal cost pricing. In particular, many subscription services charge different per-period prices for different periods within a multiperiod subscription. Particularly common is the practice of offering a low price, or even a zero price, for an introductory period.[26] One explanation for this practice is that sellers are exploiting consumer misperception—this time, consumer underestimation of future use. When signing on to a yearlong subscription service with a "two-month free" introductory offer, some consumers think that they will end the subscription after the first two months.

---

[24] For important theoretical and empirical analyses of subscription markets and contracts from a behavioral perspective see DellaVigna and Malmendier, 119 Q J Econ 353 (cited in note 9) (setting forth a theoretical analysis supported by evidence from numerous markets); DellaVigna and Malmendier, *Paying Not to Go to the Gym* (cited in note 8) (studying health club membership contracts); Oster and Morton, *Behavioral Decision-Making* (cited in note 8) (analyzing magazine subscriptions).

[25] This is an assumption supported by casual observation as well as by more rigorous empirical analysis. See DellaVigna and Malmendier, *Paying Not to Go to the Gym* (cited in note 8).

[26] Examples include introductory offers by newspapers and magazines, credit card teaser rates, etc.

Not all of these consumers actually end their subscription after two months. Put differently, sellers may be responding to consumers' underestimation of the length of the period during which they would need or want a subscription with the specific seller. If consumers underestimate the effective subscription period, then sellers in a competitive market will backload their price as much as possible.

This too is an example of intertemporal bundling. The cement holding the bundle together is the cost of switching from one seller to another or simply the cost of terminating the subscription. It is consumers' underestimation of these switching costs that explain the viability of a below marginal cost introductory fee. Since switching costs keep the bundle together, it is not surprising that sellers do not make a special effort to reduce these costs. Perhaps the notorious ten minute "please hold for the next available representative" wait that must be endured to cancel a subscription is not merely the result of a seller's attempt to economize on the size of its customer service department. More direct measures designed to increase switching costs are lock-in clauses and termination fees. If consumers underestimate the cost of switching or, equivalently, underestimate the likely length of their subscription, sellers who fail to take advantage of switching cost-induced bundling and offer only per-period marginal cost pricing will not survive in a competitive market.[27]

The normative implications of intertemporal bundling depend on the misperception to which the bundling responds. Consider overestimation of future health club use due to overestimation of self control. Absent bundling such overestimation of value would lead to excessive initial visits to the health club by consumers who will not make sufficient return visits to secure any benefit. Bundling with a front-loaded subscription price will only exacerbate this inefficiency.

Next consider bundling in response to underestimation of use due to overestimation of termination/switching or underestimation of termination/switching costs. Absent bundling, that is, with zero termination/switching costs, the consumer would make efficient period-by-period decisions. Bundling will lead to excessive initiation of subscription contracts.

---

[27] For a formal analysis of the underestimation of switching costs and its implications see DellaVigna and Malmendier, 119 Q J Econ at 389 (cited in note 9).

B. CREDIT CARDS[28]

The credit card is a complex, multi-attribute product. In fact, the credit card is a bundle of different products and services. The credit card bundles together transacting and borrowing services. The credit card also implements intertemporal bundling, whereas borrowing now is bundled with borrowing later. The argument advanced in this Essay suggests that these forms of bundling are a strategic response to consumer misperception.[29]

Evidence suggests that consumers systematically underestimate their future credit card borrowing. They also underestimate the likelihood that they will need to consume various borrowing-related services. They underestimate the likelihood of late repayment. They underestimate the likelihood of requiring a special (and more expensive) loan beyond their specified credit limit.[30] Beyond underestimation of borrowing (and consequently of the likelihood that they will use borrowing-related services), consumers also underestimate the price of borrowing and borrowing-related services. They underestimate the speed with which interest accrues. They do not fully understand the implications of minimum payments and low (or even negative) amortization rates. And they are not aware of various computational techniques employed by issuers to increase the magnitude of interest (and related) payments.[31]

What are the implications of consumers' underestimation of borrowing and of the price of borrowing? In particular, what are the implications for the transacting service? Can the transacting service be offered independently? Imagine a card issuer who offers only transacting services—a debit card. In a competitive market this issuer would have to set a price equal to

---

[28] This subsection is based on Oren Bar-Gill, *Seduction by Plastic*, 98 Nw U L Rev 1373 (2004). Economist Lawrence Ausubel was the first to suggest that credit card pricing responds to consumer misperception. See Lawrence M. Ausubel, *The Failure of Competition in the Credit Card Market*, 81 Am Econ Rev 50, 70–72 (1991); see also David Laibson, Andrea Repetto, and Jeremy Tobacman, *A Debt Puzzle*, in Philippe Aghion, et al, ed, *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps* 228, 230–31 (Princeton 2003); Lawrence M. Ausubel, *Credit Card Defaults, Credit Card Profits, and Bankruptcy*, 71 Am Bankr L J 249, 261 (1997); Haiyan Shui and Lawrence Ausubel, *Consumer Time Inconsistency: Evidence from an Experiment in the Credit Card Market* 2–3 (unpublished manuscript 2004), online at http://ssrn.com/abstract=586622 (visited Nov 21, 2005).

[29] I do not wish to preclude other explanations, for example cost-based explanations, for the bundling of transacting and borrowing services in the credit card.

[30] Consumers also underestimate their use of different transacting-related services, for example, currency exchange pursuant to foreign purchases.

[31] For a description of these techniques see Mark Furletti, *Credit Card Pricing Developments and Their Disclosure*, Payment Cards Center Discussion Paper, Federal Reserve Bank of Philadelphia 1, 14–16 (2003), online at http://www.phil.frb.org/pcc/discussion/discussion0103.pdf (visited Nov 21, 2005) (discussing such techniques as payment allocation, compounded interest, and double-cycle interest, which help increase the size of interest payments).

the marginal cost of the transacting service offered. Now consider a credit card issuer that bundles transacting and financing services. Given consumer underestimation of borrowing the credit card issuer would respond by setting a high price for the financing service and a low price for the transacting service. In fact, credit card issuers often offer a negative price for the transacting service: the transacting consumer receives bonus points and frequent flyer miles for every dollar spent. With underestimation of borrowing the bundle offered by the credit card issuer will be more attractive to consumers than transacting and borrowing services sold separately (and at per-service marginal cost).

Consistent with this analysis, debit cards, despite their increasing popularity, are finding it difficult to compete with the credit card bundle. Absent the back-end financing and fees revenues that credit card issuers enjoy, debit card issuers cannot match the attractive short-term perks that credit card issuers routinely offer.[32] Accordingly, debit cards are quickly replacing checks, but are not as successful in supplanting credit cards.[33]

Moving on to intertemporal bundling, consumers underestimate the cost of switching cards and thus overestimate the likelihood of switching. Issuers respond by offering short-term credit at low, even zero teaser rates. Lenders offering equivalent short-term credit as a stand alone service must price the loan at cost, and thus cannot compete with an issuer that engages in intertemporal bundling.

What are the welfare consequences of the different forms of bundling observed in the credit card market? Absent bundling consumers would make accurate transacting and short-term borrowing decisions. Bundling, with its accompanying back-loaded prices, leads to excessive transacting and short-term borrowing.

---

[32] See Ronald J. Mann, *Global Credit Card Use and Debt: Policy Issues and Regulatory Responses*, U of Tex L and Econ Working Paper No. 49, 28–29 (2004), online at http://ssrn.com/abstract=509063 (visited Nov 21, 2005) (describing the disadvantages of debit cards, specifically no float, no benefits, and fewer legal protections); Sujit Chakravorti and Alpa Shah, *A Study of the Interrelated Bilateral Transactions in Credit Card Networks*, Emerging Payments Occasional Paper Series (EPS-2001-2) 1, 23 (Federal Reserve Bank of Chicago July 2001) (showing that debit cards do not offer the benefit programs, extended warranties, and free credit until the end of the billing cycle that credit cards offer).

[33] See Mann, U of Tex L and Econ Research Paper No. 49 at 31–32 (cited in note 32) (showing that debit cards have surpassed check use, but that in the United States, debit cards have had trouble infiltrating the dominant network that credit cards have in place). This is not to say that debit cards are not replacing credit cards at all. They are. See Tom Brown and Lacey Plache, *Paying With Plastic: Maybe Not So Crazy*, 73 U Chi L Rev **XX** (2006). The success of the debit card is at least partially explained by its successful bundling with other banking services—the cost of which are underestimated by consumers.

The distributional consequences of credit card bundling are also troubling. Bundling of transacting and financing with prices backloaded onto the long-term financing component implies that borrowers, consumers that are more likely to use and pay for the financing services, will cross-subsidize transactors, consumers that use their cards mainly for transacting and enjoy free service plus frequent flyer miles and bonus points. As long as borrowing is positively correlated with weaker socioeconomic status bundling leads to regressive distributional consequences. Bundling is similarly unattractive if weaker consumers are more likely to extend their low-price short-term loans into the high-priced long-term.

C. Cell Phones

As in the credit card market, common practices in the cell phone market suggest that providers of wireless communication services are responding to consumer misperception.[34] Consider the steep jump in per-minute charges when the consumer exceeds the plan limit. A recent study found that most contracts specify an increase of over 100 percent in the per-minute price, with some contracts specifying increases of 200 percent and beyond.[35] Clearly, these huge increases do not reflect a corresponding change in the provider's per-minute cost.

Arguably, the high prices set for minutes beyond the plan limit respond to consumers' underestimation of their future use of the cellular phone. Providers respond to consumer misperception by bundling airtime (i.e., talking minutes), handsets and other services such as voice mail. The high long-term prices subsidize the free phones, free voicemail, and lower short-term prices.[36] A seller that offers handsets as a standalone item would have to price these handsets at their marginal cost. Such a seller would find it difficult to compete with cellular service providers who "give away" free handsets as part of their "handset plus service" bundle.

Wireless service contracts are also an important example of multi-period subscriptions. As discussed in Part III.A, providers employ various

---

[34] See DellaVigna and Malmendier, 119 Q J Econ at 380 (cited in note 9) ("Cellular phone companies can extract profits from naiveté by setting high marginal prices for minutes beyond the monthly allowance."); Bar-Gill, 98 Nw U L Rev at 1429–30 (cited in note 28) ("As in the credit card market, competition might be pressuring providers of wireless communication services to exploit consumers' imperfect perceptions of the future."); Barry Nalebuff & Ian Ayres, *Why Not?: How to Use Everyday Ingenuity to Solve Problems Big and Small* 178–79 (Harvard Bus Sch 2003) (discussing how cell phone carriers exploit consumer misperceptions).

[35] See DellaVigna and Malmendier, 119 Q J Econ at 380–81 (cited in note 9).

[36] See Nalebuff and Ayres, *Why Not?: How to Use Everyday Ingenuity to Solve Problems Big and Small* at 179 (cited in note 34) (noting that competition in the cell phone market focuses on the short-term, free phone dimension).

tactics designed to sustain this intertemporal bundling though increased switching costs. This bundling argument explains the common lock-in clause, which ties the consumer to the specific provider for as long as three years.[37] The lock-in clause targets consumers' underestimation of the cost of lock-in. Specifically, consumers may underestimate the many contingencies that would induce them to end the contract earlier: the appearance of a more attractive offer from another provider; a change in their need for wireless services; or an unanticipated financial hardship that renders the monthly cellular phone bill too painful to bear.

As with high prices for minutes beyond the plan limit, it seems difficult to justify the lock-in clause on cost grounds. In some industries, fixed costs may justify a lock-in clause. It is unlikely, however, that per-consumer fixed costs alone can explain the lengthy lock-in clauses observed in the cell phone industry. Moreover, even if fixed costs are substantial, lock-in is not the obvious response. Why not simply charge an upfront fee? Lock-in clauses are common because consumers underestimate the cost of lock-in. In the cell phone market lock-in clauses do double duty. First, they sustain the intertemporal bundle. Second, they support the bundling of cellular service plans with handsets and other services. The revenues generated by these lock-in clauses, together with the revenues generated by high prices for minutes beyond the plan limit, pay for the free phones and other short-term perks offered by cell phone service providers.

What are the welfare implications of bundling in the cell phone market? If consumers overestimate the value of cellular communications (or at least overestimate the value of sticking to a specific provider), they will buy too many handsets even absent bundling. Bundling only exacerbates this problem.

The cell phone market exemplifies the dynamic interaction between consumer perceptions and seller/provider reactions. Recent innovations such as rollover minutes and flexible, no lock-in contracts suggest that at least some consumers have more accurate perceptions about the long-term costs of the wireless service. The evolution of consumer perception is driven by independent learning by consumers and by providers' advertisement campaigns. Understanding consumer perception may help predict market outcomes. Conversely, market outcomes can serve as indicators of consumer (mis)perception.

---

[37] Providers offer different short-term perks to tempt consumers into choosing service plans with longer term commitments. See DellaVigna and Malmendier, 119 Q J Econ at 380–81 (cited in note 9); Bar-Gill, 98 Nw U L Rev at 1429–30 (cited in note 28).

IV. POLICY IMPLICATIONS

The preceding analysis suggests that sellers often respond to consumer misperception by bundling the misperceived product (or component) with another, accurately perceived product (or component). The analysis further suggests that such bundling can be either welfare reducing or welfare enhancing. When bundling exacerbates the adverse effects of consumer misperception, regulation designed to discourage bundling may be desirable.[38]

In noncompetitive markets the antitrust prohibition on tying serves as a direct unbundling policy.[39] One possibility is to extend this prohibition against bundling to competitive markets. In at least two contexts such an extension may have already occurred. First, where a base-good seller operating in a competitive market (for the base good) attempts to bundle the base good with aftermarket parts or service, the Supreme Court has suggested that antitrust tying law may apply.[40] Second, The Magnuson-Moss warranty legislation of 1975 restricts sellers' ability to bundle warranted goods with other goods regardless of the level of competition in the relevant market.[41] Given the severity of this remedy, however, it should probably be used, if at all, only in extreme cases where the bundling practice is obviously harmful and where alternative policies are ineffective.

A less blunt unbundling policy is to promote competition on each component of the bundled product. If a consumer who bought a printer from seller A can buy ink cartridges from seller B, seller A would not be able to set low (below marginal cost) printer prices and high (above marginal cost) ink prices. This example suggests standardization as a potential solution to the bundling problem.[42]

---

38 I focus explicitly on unbundling policies. Other policies may be equally effective. For example, regulators can directly target the misperception that gives rise to the bundling response.

39 See, for example, the anti-tying provisions of the Sherman and Clayton Acts, 15 USC §§ 1, 14 (2000). For a recent analysis of tying in the printer market, see *Independent Ink, Inc v Illinois Tool Works, Inc*, 396 F3d 1342 (Fed Cir 2005), cert granted, 125 S Ct 2937 (2005) (discussing patent tying through a contract requiring the buyer to purchase toner cartridges only from the seller of the printer).

40 See *Eastman Kodak Co v Image Technical Services, Inc*, 504 US 451, 479 (1992) (denying defendant's motion for summary judgment on the allegation that the defendant's practice of limiting independent service organizations' access to materials that were necessary to partake in the machine repair service market constituted an illegal tying arrangement in violation of the Sherman Act, even though the base good market was competitive); Carl Shapiro, *Aftermarkets and Consumer Welfare: Making Sense of* Kodak, 63 Antitrust L J 483 (1995) (summarizing and criticizing the *Kodak* rule). However, given the specific circumstances of the *Kodak* case and the way the *Kodak* rule has been interpreted in subsequent decisions, the Supreme Court's willingness to extend tying law to competitive markets should not be overstated.

41 15 USC § 2302(c) (2000). See Craswell, 62 BU L Rev at 662 (cited in note 7).

42 Of course, standardization is not without cost. In particular, it may hinder innovation.

Focusing on intertemporal bundling, the use of bundling tactics can be discouraged by reducing switching costs. The legal guarantee of cell phone number portability is an example of a policy aimed at increasing competition by reducing switching costs.[43] Limiting sellers' ability to use early termination penalties in subscription contracts is another example of a competition fostering, unbundling policy.

Disclosure regulation may also serve as an unbundling policy.[44] If sellers bundle printers and ink in response to consumer misperception about future use, regulation requiring sellers to provide "total cost of ownership" information may effectively prevent bundling.[45] If a seller must advertise, in addition to the printer's stand alone price, an inclusive price that adds the average cost of ink over the life of the printer, consumers will be less inclined to buy cheap printers that are bundled with expensive ink. If a mortgage lender or a credit card issuer is required to calculate for the consumer and explicitly state the total (or expected) interest and fee payments over the life of the loan, then consumers will be more likely to balance this total cost information against the short term perks offered by the lender or issuer on a bundled product.[46]

V. CONCLUSION

Bundles are everywhere. Durables are bundled with parts and service. Diagnostic services are bundled with treatment services.[47] Products are bundled with selling services (e.g., showrooms and knowledgeable salespersons).[48] Michael Spence, in a seminal contribution, argued that almost every product "should be thought of as a bundle of characteristics."[49] In the modern world economy these bundled characteristics should be broadly

---

43 See 47 CFR § 52.23 (2005) (requiring carriers to allow for phone number portability).

44 For an excellent discussion of disclosure regulation as a potential response to bundling see Craswell, 62 BU L Rev at 689–94 (cited in note 13).

45 The EPA's "Energy Star" program is a leading example. On the benefits of "total cost of ownership" information see Nalebuff and Ayres, *Why Not?* at 177 (cited in note 34) (claiming the EPA's Energy Star program gives consumers total-cost-of-ownership information, which allows them to "add in the expected energy costs to the initial purchase price" of many products, and that "[c]onsequently, the program is said to save consumers $5 billion annually"). The special disclosure required under Section 106 of the Magnuson-Moss Warranty—Federal Trade Commission Act when bundling durables and service is another case on point. See Pub L No 93-617, 88 Stat 2188 (1975), codified at 15 USC § 2306 (2004). Ideally such disclosure regulation can achieve the same results as direct price regulation—a much more obtrusive policy.

46 See Bar-Gill, 98 Nw U L Rev at 1403–04 (cited in note 28). Bar-Gill also discusses the limits of disclosure. Id at 1418.

47 See Beales, Craswell, and Salop, 24 J L & Econ at 515 (cited in note 1).

48 On resale price maintenance and the recovery of selling costs see, for example, Lester G. Telser, *Why Should Manufacturers Want Fair Trade?*, 3 J L & Econ 86, 104 (1960).

49 Spence, 44 Rev Econ Studies at 561 (cited in note 1).

defined to include contractual provisions and potentially independent products.

The motivations for bundling are numerous: from leveraging of monopoly power, to product differentiation to simple cost saving. This Essay explored another motivation for bundling. It presented bundling as a strategic response to consumer misperception. The welfare and policy implications of bundling depend on the motivation for the observed bundling. Monopoly leveraging is bad. Cost saving is good. Bundling in response to consumer misperception can be either good or bad.

This Essay provided some tools for the policymaker to identify misperception-based bundling, and to ascertain when such bundling is welfare reducing. It then considered various regulatory responses and unbundling policies. The difficulty in identifying the motivation (or motivations) for an observed bundle, coupled with the difficulty in evaluating the welfare implications of bundling even when its underlying motivation is revealed suggests regulatory caution. For this reason the most attractive unbundling policies are those that facilitate the smooth operation of markets—through reduced switching costs and the provision of information—rather than the more heavy handed policies that directly prohibit bundling or attempt to fix the price of the bundle or its components.[50]

APPENDIX: ELASTIC DEMAND

This appendix extends the analysis in Part II of the article by introducing elastic demand.

A. Framework of Analysis

Returning to the printers and ink example, let $i \geq 0$ denote the amount of ink (for example, the number of ink cartridges) to be used by the consumer, which also represents the amount of printing that the consumer will do. Let $v(i) > 0$ denote the value of printing and assume decreasing marginal utility from printing: $v'(i) > 0$ and $v''(i) < 0$. (To guarantee an internal maximum I also assume that $v'(0) = \infty$ and $\lim_{i \to \infty} v'(i) = 0$.) Introduc-

---

[50] See Colin Camerer, et al, *Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism,"* 151 U Pa L Rev 1211, 1212 (2003) (proposing "an approach to evaluating paternalistic regulations and doctrines . . . [called] 'asymmetric paternalism' . . . [which] creates large benefits for those who make errors while imposing little or no harm on those who are fully rational"); Cass R. Sunstein and Richard H. Thaler, *Libertarian Paternalism Is Not an Oxymoron*, 70 U Chi L Rev 1159, 1160, 1162 (2003) (proposing "a form of paternalism, libertarian in spirit, that should be acceptable to those who are firmly committed to freedom of choice on grounds of either autonomy or welfare. . . . Libertarian paternalism is a relatively weak and nonintrusive type of paternalism, because choices are not blocked or fenced off.").

ing consumer misperception, let the perceived value of printing be $\hat{v}(i) = \delta \cdot v(i)$, where $\delta < 1$ represents underestimation and $\delta > 1$ represents overestimation. The per-unit costs of printers and ink are c_p and c_i, respectively. The per-unit prices of printers and ink are p_p and p_i, respectively.

I study a two-period model. At T=0 the consumer decides whether to purchase a printer. At T=1 the consumer decides how much to print, and accordingly how much ink to purchase. Misperception affects only the T=0 printer purchase decision, not the T=1 ink purchase decision. At T=1 the consumer learns her true value of printing, $v(i)$. The model is solved backwards, starting from the T=1 ink purchase decision.

If the consumer bought a printer at T=0, then at T=1 she will buy an amount of ink $i^*(p_i)$ that solves $\max_i \langle v(i) - p_i \cdot i \rangle$. The First-Order Condition (FOC) $v'(i^*(p_i)) = p_i$ implicitly defines $i^*(p_i)$.

I now examine the consumer's T=0 decision whether to purchase a printer. First note that due to misperception at T=0 the consumer *thinks* that at T=1 she will buy an amount of ink $\hat{i}^*(p_i)$ that satisfies $\hat{v}'(\hat{i}^*(p_i)) = p_i$ or $v'(\hat{i}^*(p_i)) = p_i / \delta$. Comparing the FOCs that define $i^*(p_i)$ and $\hat{i}^*(p_i)$ leads to the following observation.

Observation 1:

(1) If the consumer underestimates the value of printing, i.e., if $\delta < 1$, then at T=0 the consumer underestimates the amount of ink that she will buy at T=1: $\hat{i}^*(p_i) < i^*(p_i)$.

(2) If the consumer overestimates the value of printing, i.e., if $\delta > 1$, then at T=0 the consumer overestimates the amount of ink that she will buy at T=1: $\hat{i}^*(p_i) > i^*(p_i)$.

If the consumer buys a printer, she will enjoy a surplus of $w(p_p, p_i) = v(i^*(p_i)) - p_i \cdot i^*(p_i) - p_p$. Therefore, the consumer *should* buy a printer at T=0 *if and only if* $w(p_p, p_i) > 0$. But rather than the true surplus, $w(p_p, p_i)$, the consumer perceives $\hat{w}(p_p, p_i) = \hat{v}(\hat{i}^*(p_i)) - p_i \cdot \hat{i}^*(p_i) - p_p$. Therefore, the consumer *will* buy a printer at T=0 *if and only if* $\hat{w}(p_p, p_i) > 0$.

I begin by deriving the equilibrium behavior and welfare in the no-bundling case. I then compare this no-bundling benchmark to the case where bundling is feasible.

B. The No-Bundling Case

Absent bundling in a competitive market sellers will adopt per-product marginal cost pricing, namely $p_i = c_i$ and $p_p = c_p$. At T=1 the

consumer will buy an amount of ink $i^*(c_i)$ that satisfies $v'\left(i^*(c_i)\right) = c_i$. At T=0, the consumer surplus is $w(p_p, p_i) = v\left(i^*(c_i)\right) - c_i \cdot i^*(c_i) - c_p$, and the consumer *should* buy a printer *if and only if* $w(c_p, c_i) > 0$. But the consumer perceives a surplus of $\hat{w}(c_p, c_i) = \hat{v}\left(\hat{i}^*(c_i)\right) - c_i \cdot \hat{i}^*(c_i) - c_p$, and *will* buy a printer at T=0 if $\hat{w}(c_p, c_i) > 0$. This leads to the following observation.

Observation 2:

(1) If the consumer underestimates the value of printing, i.e., if $\delta < 1$, then $\hat{w}(c_p, c_i) < w(c_p, c_i)$ and the consumer might not buy a value-increasing printer.

(2) If the consumer overestimates the value of printing, i.e., if $\delta > 1$, then $\hat{w}(c_p, c_i) > w(c_p, c_i)$ and the consumer might buy a value-reducing printer.

Proof:

Rewriting the perceived welfare as a function of δ, $\hat{w}(p_p, p_i; \delta) = \delta \cdot v\left(\hat{i}^*(p_i)\right) - p_i \cdot \hat{i}^*(p_i) - p_p$, and applying the Envelope Theorem, I obtain: $\dfrac{\partial \hat{w}(p_p, p_i; \delta)}{\partial \delta} = v\left(\hat{i}^*(p_i)\right) > 0$. Noting that $w(p_p, p_i; \delta) = \hat{w}(p_p, p_i; \delta = 1)$, the results stated in Observation 2 follow.

QED

C. The Bundling Case

Bundling eliminates the $p_p = c_p$ and $p_i = c_i$ constraint, and replaces it with the more lax zero profit constraint, $\pi(p_p, p_i) = 0$, where $\pi(p_p, p_i) = p_p + p_i \cdot i^*(p_i) - \left[c_p + c_i \cdot i^*(p_i)\right]$ is the seller's profit. The bundling seller chooses a pricing scheme (p_p, p_i) that solves: $\max_{p_p, p_i} \left\langle \hat{w}(p_p, p_i) \right\rangle$ *s.t.* $\pi(p_p, p_i) = 0$.[51] FOCs reduce to:

$$(1) \quad \frac{\partial \hat{w}(p_p, p_i) / \partial p_p}{\partial \pi(p_p, p_i) / \partial p_p} = \frac{\partial \hat{w}(p_p, p_i) / \partial p_i}{\partial \pi(p_p, p_i) / \partial p_i}$$

[51] This formulation of the seller's maximization problem follows from an assumption of Bertrand competition among sellers. First, the seller's profit must equal zero; otherwise a competitor will offer a slightly lower price (on either dimension, p_p or p_i) and attract all of the consumers. Second, given zero profit, the seller must choose a price combination that maximizes consumer welfare; otherwise a competitor will offer a welfare-increasing price combination and attract all of the consumers.

The (absolute value of the) fraction on the left hand side of Equation (1) is the consumer welfare "bang" for each firm profit "buck" when p_p is lowered. The (absolute value of the) fraction on the right hand side of Equation (1) is the consumer welfare "bang" for each firm profit "buck" when p_i is lowered. These two ratios must be the same. Otherwise there is an "arbitrage" that can increase consumer welfare without reducing firm profits.

The fraction on the left hand side of Equation (1) simplifies (using the Envelope Theorem) to: $\frac{\partial \hat{w}(p_p, p_i)/\partial p_p}{\partial \pi(p_p, p_i)/\partial p_p} = -1$. The consumer welfare bang for firm profit buck is always -1, since p_p is just a zero sum transfer. The fraction on the right hand side of Equation (1) simplifies (using the Envelope Theorem) to: $\frac{\partial \hat{w}(p_p, p_i)/\partial p_i}{\partial \pi(p_p, p_i)/\partial p_i} = \frac{-\hat{i}^*(p_i)}{i^*(p_i) + (p_i - c_i)\cdot \partial i^*(p_i)/\partial p_i}$. A \$1 increase in ink price reduces perceived consumer welfare by $\hat{i}^*(p_i)$, i.e., by the number of ink cartridges that the consumer expects to buy. The same \$1 increase in ink price raises the seller's profit on infra-marginal cartridges by $\hat{i}^*(p_i)$, but costs the seller a profit of $p_i - c_i$ on the marginal cartridges that will no longer be sold given the higher price.

Substituting into Equation (1), I obtain:

$$(2) \quad \frac{-\hat{i}^*(p_i)}{i^*(p_i) + (p_i - c_i)\cdot \partial i^*(p_i)/\partial p_i} = -1$$

When $\delta = 1$, $\hat{i}^*(p_i) = i^*(p_i)$ and Equation (2) implies $p_i = c_i$. When $\delta < 1$, $\hat{i}^*(p_i) < i^*(p_i)$ and Equation (2) implies $p_i > c_i$ (given $di^*(p_i)/dp_i < 0$). And when $\delta > 1$, $\hat{i}^*(p_i) > i^*(p_i)$ and Equation (2) implies $p_i < c_i$ (given $di^*(p_i)/dp_i < 0$). Intuitively, a larger δ increases the number of ink cartridges that the consumer expects to purchase, and thus increases the consumer welfare bang for every firm profit buck attained by lowering the price of ink. Therefore, the price of ink is inversely related to δ.[52]

---

[52] The analysis thus far derives from the FOCs of the seller's optimization problem that were reduced to Equation (1). Interestingly, the same FOCs, and thus the same results, obtain when competition is replaced by monopoly, and the monopolist solves the problem: $\max_{p_p, p_i} \langle \pi(p_p, p_i) \rangle$ *s.t.* $\hat{w}(p_p, p_i) = 0$. Of course, while the ink price will remain unchanged, the printer price will necessarily increase.

The results obtained for ink price, when plugged into the zero profit constraint, produce the parallel results for the printer price. Rearranging the $\pi(p_p, p_i) = 0$ constraint, I obtain: $p_p - c_p = -(p_i - c_i) \cdot i^*(p_i)$. Therefore, when $\delta = 1$ implies $p_i = c_i$, it also implies $p_p = c_p$. When $\delta < 1$ implies $p_i > c_i$, it also implies $p_p < c_p$. And, when $\delta > 1$ implies $p_i < c_i$, it also implies $p_p > c_p$.

Proposition 1 summarizes:

Proposition 1:

(1) When consumers underestimate the value of printing, i.e., when $\delta < 1$, a bundling seller will set $p_i > c_i$ and $p_p < c_p$.

(2) When consumers overestimate the value of printing, i.e., when $\delta > 1$, a bundling seller will set $p_i < c_i$ and $p_p > c_p$.

The welfare implications of bundling are summarized in the following proposition.

Proposition 2:

(1) When consumers underestimate the value of printing, i.e., when $\delta < 1$,

a. If a printer is purchased both with and without bundling, i.e., if $0 < \hat{w}(c_p, c_i) < \hat{w}(p_p, p_i)$, then bundling reduces welfare.

b. If a printer is *not* purchased both with and without bundling, i.e., if $\hat{w}(c_p, c_i) < \hat{w}(p_p, p_i) \leq 0$, then bundling has no effect on welfare.

c. If a printer is not purchased absent bundling but is purchased with bundling, i.e., if $\hat{w}(c_p, c_i) \leq 0 < \hat{w}(p_p, p_i)$, then bundling increases welfare.

(2) When consumers overestimate the value of printing, i.e., when $\delta > 1$, bundling reduces welfare:

a. If a printer is purchased both with and without bundling, i.e., if $0 < \hat{w}(c_p, c_i) < \hat{w}(p_p, p_i)$, then bundling reduces welfare.

b. If a printer is *not* purchased both with and without bundling, i.e., if $\hat{w}(c_p, c_i) < \hat{w}(p_p, p_i) \leq 0$, then bundling has no effect on welfare.

c. If a printer is not purchased absent bundling but is purchased with bundling, i.e., if $\hat{w}(c_p, c_i) \leq 0 < \hat{w}(p_p, p_i)$, then bundling reduces welfare.

Proof:

When bundling is feasible prices are set to maximize $\hat{w}(p_p, p_i)$. Therefore, $\hat{w}(p_p, p_i) > \hat{w}(c_p, c_i)$.

(1) There are three possible scenarios:

Scenario 1—$0 < \hat{w}(c_p, c_i) < \hat{w}(p_p, p_i)$: A printer is purchased both with and without bundling. In this scenario no bundling is better, since it optimizes the ink purchase decision. Formally, as shown in the analysis preceding Proposition 1, when $\delta = 1$ $\hat{w}(\cdot,\cdot)$, which then equals $w(\cdot,\cdot)$, is maximized by $p_i = c_i$ and $p_p = c_p$. Therefore, for any set of prices $p_i \neq c_i$ and $p_p \neq c_p$: $w(p_p, p_i) < w(c_p, c_i)$. If a printer is purchased both with and without bundling, welfare is greater absent bundling.

Scenario 2—$\hat{w}(c_p, c_i) < \hat{w}(p_p, p_i) \leq 0$: A printer is *not* purchased both with and without bundling. In this scenario bundling does not affect welfare.

Scenario 3—$\hat{w}(c_p, c_i) \leq 0 < \hat{w}(p_p, p_i)$: A printer is not purchased without bundling, but a printer is purchased with bundling. In this scenario bundling increases welfare if $w(p_p, p_i) > 0$. I next show that $w(p_p, p_i) > \hat{w}(p_p, p_i)$ (and since in Scenario 3 $\hat{w}(p_p, p_i) > 0$ this implies $w(p_p, p_i) > 0$): Rewriting the perceived welfare as a function of δ, $\hat{w}(p_p, p_i; \delta) = \delta \cdot v(\hat{i}^*(p_i)) - p_i \cdot \hat{i}^*(p_i) - p_p$, and applying the Envelope Theorem, I obtain: $\frac{\partial \hat{w}(p_p, p_i; \delta)}{\partial \delta} = v(\hat{i}^*(p_i)) > 0$. It follows that $w(p_p, p_i; \delta) = \hat{w}(p_p, p_i; \delta = 1) > \hat{w}(p_p, p_i; \delta < 1)$.

(2) There are three possible scenarios:

Scenario 1—$0 < \hat{w}(c_p, c_i) < \hat{w}(p_p, p_i)$: A printer is purchased both with and without bundling. In this scenario no bundling is better, since it optimizes the ink purchase decision. (See part (1) of the proof for the formal derivation.)

Scenario 2—$\hat{w}(c_p, c_i) < \hat{w}(p_p, p_i) \leq 0$: A printer is *not* purchased both with and without bundling. In this scenario bundling does not affect welfare.

Scenario 3—$\hat{w}(c_p, c_i) \leq 0 < \hat{w}(p_p, p_i)$: A printer is not purchased without bundling, but a printer is purchased with bundling. In this scenario bundling reduces welfare, since this is an inefficient purchase: $w(p_p, p_i) < \hat{w}(c_p, c_i)$ (and in Scenario 3 $\hat{w}(c_p, c_i) < 0$). To see

that $w(p_p, p_i) < \hat{w}(c_p, c_i)$, recall that $w(c_p, c_i) < \hat{w}(c_p, c_i)$ (from Observation 2) and $w(p_p, p_i) < w(c_p, c_i)$ (proved in part (1), scenario 1).

QED

Paying Not to Go to the Gym*

Stefano DellaVigna
UC Berkeley and NBER
sdellavi@econ.berkeley.edu

Ulrike Malmendier
Stanford University and NBER
ulrikem@stanford.edu

April 15, 2005

Abstract

How do consumers choose from a menu of contracts? We analyze a novel data set from three US health clubs with information on both the contractual choice and the day-to-day attendance decisions of 7,978 members over three years. The observed consumer behavior is difficult to reconcile with standard preferences and beliefs. First, members who choose a contract with a flat monthly fee of over $70 attend on average 4.8 times per month. They pay a price per expected visit of more than $17, even though they could pay $10 per visit using a 10–visit pass. On average, these users forgo savings of $700 during their membership. Second, consumers who choose a monthly contract are 18 percent more likely to stay enrolled beyond one year than users committing for a year. This is surprising because monthly members pay higher fees for the option to cancel each month. We also document cancellation delays and attendance expectations, among other findings. Leading explanations for our findings are overconfidence about future efficiency or about future self-control. Overconfident agents overestimate attendance as well as the cancellation probability of automatically-renewed contracts. Our results suggest that making inferences from observed contract choice under the rational expectation hypothesis can lead to biases in the estimation of consumer preferences.

*Earlier versions of this paper were distributed under the titles "Self-Control in the Market: Evidence from the Health Club Industry" and "Overestimating Self-Control: Evidence from the Health Club Industry". We thank Doug Bernheim (the editor), two anonymous referees, Marios Angeletos, George Baker, Rachel Croson, Rajeev Dehejia, Richard Gilbert, Oliver Hart, Caroline Hoxby, Daniele Paserman, Antonio Rangel, Emmanuel Saez, Andrei Shleifer, Jeremy Tobacman, Klaus Wertenbroch, Justin Wolfers, and in particular Edward Glaeser, Lawrence Katz, and David Laibson for their comments. We also thank the participants of seminars in CREST, Dartmouth College, Harvard, Michigan (Ann Arbor), Northwestern, NYU, UC Berkeley, UCLA, the Ente Einaudi, the SITE 2001 and 2002, and at the Eastern Economic Association 2002 for their comments. We thank Bryan S. Graham for providing Mathematica code and Tobias Adrian, Augustin Landier, and Sendhil Mullainathan for inspiring conversations at the beginning of this project. Tricia Glynn, Burak Guner, Camelia Kuhnen, Scott McLinn, Boris Nenchev, and Nikita Piankov provided excellent research assistance. For financial support, DellaVigna thanks Bank of Italy, Malmendier thanks the German Academic Exchange Service (DAAD).

"Saturday 31 December. New Year's Resolutions. I WILL [...] go to the gym three times a week not merely to buy sandwich." (Fielding, 1999. *Bridget Jones' Diary: A Novel*)

A few months later: "Monday 28 April. [...] Gym visits 0, no. of gym visits so far this year 1, cost of gym membership per year £370; cost of single gym visit £123 (v. bad economy)." (Fielding, 2001. *Bridget Jones: The Edge of Reason*)

1 Introduction

Many firms offer consumers a menu of contracts. Cellular phone users choose combinations of monthly airtime minutes and prices. Credit card users choose between teaser rate offers and contracts with a constant interest rate. A large literature in industrial organization analyzes the profit-maximizing contract design (Tirole, 1990). A standard assumption in this literature is that consumers have rational expectations about their future consumption frequency and choose the utility-maximizing contract.

In this paper, we provide evidence that this may not always be the case. We present a novel data set from three US health clubs that allows us to analyze the contractual choices of consumers in light of their actual consumption behavior. The data set contains information both on the type of membership and the day-to-day attendance decisions of 7,978 health club members over three years. We find that consumers choose a contract that appears sub-optimal given the attendance frequency. In addition, low-attendance consumers delay cancelling this contract despite small transaction costs.

Our empirical analysis exploits the presence of a contractual menu. Consumers can choose between two flat-rate contracts—a monthly contract and an annual contract—and a pay-per-visit option. The monthly contract is automatically renewed from month to month until the consumer cancels. The annual contract, instead, expires after twelve months unless the consumer explicitly renews it. The variation in the per-usage pricing and in the renewal procedures allows us to identify several puzzling features of consumer behavior.

First, consumers who choose a monthly membership of over $70 per month pay on average seventy percent more than they would under the pay-as-you-go contract for the same number of visits. Eighty percent of the monthly members would have been better off had they paid per visit for the same number of visits.

Second, consumers who choose the monthly contract are 18 percent more likely to stay enrolled beyond one year than users choosing the annual contract. This is surprising because monthly members pay higher fees for the option to cancel each month. This result occurs even though high-attendance users sort into the annual contract at enrollment.

These and additional empirical findings (summarized in Table 1) are hard to reconcile with standard preferences and beliefs. We explore potential explanations including high transaction costs of payment per usage, risk aversion, underestimation of costs of attendance and of can-

cellation, time inconsistency, naiveté about the time inconsistency, and persuasion by health club employees.

In our view, the most parsimonious explanations are those allowing for overconfidence (naiveté). Consumers overestimate, for example, their future efficiency in pursuing costly activites or their future self-control. This leads to overestimation of attendance and of cancellation in automatically renewed contracts. Persuasion by health club employees can explain most findings.

In a simple yet economically significant decision, enrollment and attendance in a health club, consumers deviate systematically from the optimal contractual choice. In the health clubs of our sample, the average non-subsidized user chooses the monthly contract, and by doing so forgoes savings of about $700 per membership, out of a total amount of about $1,500 paid to the health club. The results of this study are likely to generalize to the 32.8m Americans who exercise in one of the 16,983 US health clubs. Therefore, both in terms of monetary magnitude and in terms of population involved, the non-standard behavior has a significant economic impact in the health club industry. Our findings are also consistent with findings on consumer behavior in the credit card industry (Shui and Ausubel, 2004) and employee choice of 401(k) plans (Madrian and Shea, 2001).

The analysis of consumer behavior is just the first step towards a better understanding of industries where consumers display non-standard preferences or beliefs. Profit-maximizing firms should respond to the non-standard features of consumer behavior in their contract design. This is the central theme of the growing literature on behavioral industrial organization (DellaVigna and Malmendier, 2004; Eliaz and Spiegler, 2004; Gabaix and Laibson, 2005; Heidhues and Koszegi, 2004). The large effect of small cancellation costs on renewal rates may explain the high frequency of contracts with automatic renewal in the newspaper, credit card, and mail order industry. The findings have implications also for the design of flat-rate pricing (Miravete, 2003). In DellaVigna and Malmendier (2004) we explore the implications for firm pricing of a leading explanation of our results, overconfidence about time inconsistency.

Our findings suggest caution in making inferences about consumer preferences from observed choices of products (Hendel and Nevo, 2004) or contracts (Miravete and Roeller, 2003) when actual consumption is unobserved. Inferences made under the assumption of rational expectations can lead to significant bias. For example, we would have concluded that monthly members attend on average at least twice a week. This erroneous conclusion would have overstated the impact of health club enrollment on health outcomes.

Finally, our findings have implications for the policy debate on obesity (Cutler, Glaeser, and Shapiro, 2003). Subsidizing enrollment in health clubs is likely to have only small effects on obesity rates, given the low average attendance of members.

The remainder of the paper is organized as follows. In Section 2 we introduce the main features of the health club data set. In Section 3 we develop predictions about the contractual

choice at enrollment and test the predictions empirically. In Section 4, we present a similar analysis of the contractual choice and consumption behavior over time. Section 5 discusses possible explanations for the empirical findings. Section 6 concludes.

2 Health club data set

Health club industry. As of January 2001, 16,983 clubs were operating in the US. The industry revenues for the year 2000 totalled $11.6bn. The memberships in the same period summed to 32.8m, up from 17.4m in 1987. Fifty-one percent of the users were members in commercial health clubs, while thirty-four percent were members in non-profit facilities. Only the market leader Bally Total Fitness with $1,007m revenues and 4m members is publicly traded. Few companies operate in more than 10 states. Ownership concentration is in the 10th percentile of US industries.

Data set. We collected a new panel data set from three health clubs located in New England, which we label clubs 1, 2, and 3. The data set contains information on the contractual choices and the day-to-day attendance of users that enrolled after April 1, 1997. The sample period lasts until August 24, 2000 for club 1 and until March 13, 2001 for clubs 2 and 3. The day-to-day record of usage is made available by the technology regulating the access to these health clubs, described below. The panel of contractual choices comes from the billing records. Each entry in the accounting data specifies the price paid for the transaction and a 4-letter code. This code allows us to track the membership type—standard, student, family, corporate—as well as details like the subsidizing company (if any).

Several companies located near the clubs subsidize their employees' attendance. For these corporate members, the health club receives part of the membership payments directly from the firms, with the remainder being paid by the members. The health club informs the companies periodically about the number of employees enrolled and their attendance. This creates incentives for the health club to record attendances accurately or, possibly, to overrecord them.

Contractual menu. We conducted a survey of the 97 health clubs in the metropolitan area of Boston to document the contract design in the industry.[1] Health clubs offer up to three options. 87 clubs offer a monthly contract. A monthly fee is automatically debited each month to a credit card or bank account until the user cancels. 90 clubs offer an annual contract. Both monthly and annual contracts have an initiation fee but no fee per visit. Finally, 82 clubs offer a pay-per-visit option, often in the form of a 10-visit pass. Health clubs 1 and 2 in our sample offer the three types of contract with the following additional features.[2]

[1] For details on the survey, see DellaVigna and Malmendier (2004).

[2] Contracts for one to six months with automatic expiration are also available. We do not include them in our analysis, since they are typically targeted towards occasional summer users. We also remove from the sample free limited-time memberships that are occasionally given to employees of the subsidizing companies.

1. The monthly contract has a monthly fee ranging between \$70 (discounted level) and \$85 (standard level). Non-corporate users also pay an initiation fee ranging from \$0 (in promotional periods) to \$150. Corporate users generally pay an out-of-pocket monthly fee between \$19 and \$65, depending on the subsidy paid by their company, and no initiation fee. Cancellation can be done in person at the club or by sending a written note.[3] If cancellation takes place before the 10th of the month, no further fees are due, and the users can attend until the end of the month. Members who cancel after the 10th have to pay the fee for the next month and can attend until the end of the following month.

2. The annual contract charges up-front 10 times the applicable monthly fee, e.g. \$850 for a standard membership[4]. Users thus get a discount of 2 months out of 12 in exchange for a yearly commitment. The initiation fee is the same as under the corresponding monthly contract. At the end of the year, the contract expires and members who wish to stay enrolled have to sign up again, either for an annual or for a monthly contract. In order to encourage renewal, the club sends out a reminder card one month before the contract expires.

3. The pay-per-visit system offers two options, either to pay \$12 per visit or to purchase a 10-visit pass for \$100. Transaction costs for the 10-visit pass are small. Users provide basic demographic information and receive a card for ten visits. Unfortunately, attendance is not tracked for the pay-per-visit users.

Users of club 3 face the same menu of contracts with lower prices and slightly different services. The monthly fee ranges from \$13 to \$52, and the initiation fee is at most \$50. The annual fee in the annual contract equals 10 times the corresponding monthly fee. The pay-per-visit options are a \$10 fee per visit, and a \$80 pass for 10 visits.

Under all types of membership, users receive cards which they have to deposit in a basket at the front desk when they enter. While they are exercising, a health club employee swipes them (marks the visit for the 10-visit passes), and users pick them up when they exit. This method guarantees a high recording precision even during peak hours. The three contracts give right to the same services, i.e., a temporary locker, towels[5], and access to the equipment. Also, both the monthly and the annual contract allow members to "freeze" (suspend) their membership for three months per year[6]. Users with a monthly contract do not have to pay their monthly fee during the freezing period. Annual members get additional usage time after the original twelve months.

[3] Some users cancel by discontinuing the payments to the health club.

[4] The annual fee can be paid in three installments due in the first 6 months.

[5] Towels are not included in memberships in Club 3.

[6] Monthly users can also quit for up to three additional months without repaying the initiation fee.

Sample construction. We match the information on attendance and on contract choice in the three clubs to form a longitudinal data set with monthly observations, covering the period from April 1997 to August 2000 (club 1) and to March 2001 (clubs 2 and 3). Our analysis focuses on *enrollment spells*. A spell starts whenever an individual enrolls (or re-enrolls) in a club and ends whenever the individual quits. We define spells to be censored if either the enrollment is ongoing at the end of the sample period, or the individual switches to a short-term contract or receives a promotional membership. Accordingly, spells are completed if the individual cancels the membership (under a monthly contract) or if the membership expires (under an annual contract) within the sample period. Individuals have multiple spells if they quit the club and re-enroll at some later date.

The initial sample includes 11,605 individuals. We drop individuals who were never enrolled in either a monthly or an annual contract (2,978 individuals). We eliminate spells with serious inconsistency in the billing data (132 spells). We also exclude users with a family membership to avoid issues regarding the joint consumption of the services (295 spells). Finally, in order to limit the sample to first-time users of these clubs, we drop users who had a free or a seasonal membership before they chose a monthly or an annual contract (293 individuals).

This leaves us with a sample of 7,978 individuals and 8,615 enrollment spells. In the paper, we use the sample 'First spell', which includes only the first enrollment spell for each individual. As Row 1 of Table 2 shows, club 1 has 19 percent more members than club 2, and more than twice as many members as club 3. The percentage of completed spells is similar across the clubs, above 60 percent. Of the 7,978 individuals enrolled in any club, 87 percent choose a monthly membership as their first contract. Health club members rarely change the type of contract they initially enroll in. We also use the sample 'First spell and no subsidy', a restriction of the sample 'First spell' to unsubsidized memberships. We consider a membership to be unsubsidized if, over the whole spell, the average out-of-pocket fee exceeds \$70 per month for enrollment in a monthly membership and \$700 per year (\$58 per month) for enrollment in an annual membership. This smaller sample includes 1,120 spells (14.03 percent of the full sample).

Descriptive statistics. In clubs 1 and 2, the average amount spent per spell is about \$550, and the average fee per month ranges between \$43 and \$53. For corporate users, these are the out-of-pocket payments and do not include the subsidies paid by the sponsoring firms. In club 3 the amounts are substantially lower, since the contracts are cheaper. In the sample 'First spell and no subsidy' (Columns 7 and 8), these amounts are 20 to 60 percent higher. The initiation fee averages \$4 in the sample 'First spell' since 86 percent of users do not pay it. In the 'First spell' sample, individuals with a monthly contract attend on average 4 times per month, and individuals with an annual contract attend on average 4.3 times per month. Attendance in club 1 (Column 1) is somewhat higher than in the other clubs. Freezing of a contract is rare in all the clubs. The bottom part of Table 2 displays the available demographic

controls. Users are somewhat more likely to be male than female and are on average in the early thirties. Corporate memberships account for 50 percent of the sample, while student memberships account for only 2 percent. Additional information on the data set construction is available in the Data Appendix.

3 Contract choice at enrollment

3.1 Predictions of the standard model

We set up a model of contract choice and health club attendance. We assume that health club attendance involves immediate effort costs and delayed health benefits, and that the effort costs are uncertain ex ante. In particular, costs can be high ($c = \bar{c}$) or low ($c = \underline{c}$) and individuals differ in the ex-ante probability that costs will be high. A contract (L', p', T') gives customers the right to exercize for a fee p' and for T' periods (days), once the flat fee L' is paid. We assume that consumers can choose between flat-fee contracts (like the monthly and annual contract) with $p' = 0$ and pay-per-visit contracts with $L' = 0$. We summarize here the results on contract choice for the case of standard preferences and rational beliefs. The derivation is in the working-paper version (DellaVigna and Malmendier, 2002).

Flat-rate vs. pay-per-usage. We consider first the choice at enrollment between a flat-rate contract $(L, 0, T)$ and a pay-per-visit contract $(0, p, T)$. Denote by δ the daily discount factor and by $E_F[v]$ the expected number of visits (over T days) under the flat-rate contract.

Prediction 1. (Price per expected attendance at enrollment) *For agents that choose a flat-rate contract,*

$$\frac{L}{E_F[v]} a(T) \leq p. \tag{1}$$

The factor $a(T) \equiv (1-\delta) T/(1-\delta^T)$ is a time-adjustment coefficient due to the fact that the flat fee L is paid up-front and the per-visit fee p is paid every period between 1 and T. For small T, such as $T = 30$ under the monthly contract, $a(T)$ is approximately 1. Equation (1) says that payment per expected visit under the flat-rate contract should be smaller than the per-visit-fee p. Intuitively, only consumers that attend frequently should choose the flat-rate contract. Suppose instead that a consumer under the flat-rate contract attends infrequently enough that the price per expected visit $L/E_F[v]$ is greater than the per-visit fee p. If this consumer switched to the pay-per-visit contract without changing state-contingent attendance, she would have higher utility. Reoptimizing the attendance choices, she must be even better off.

Annual vs. monthly contract. The annual contract A requires a yearly commitment. The monthly contract M offers the option to cancel in any period but charges a higher fee per month. Consumers that anticipate a high enough probability of being low-cost types ($c = \underline{c}$)

prefer the monthly contract for its flexibility. Users who believe that they will be high-cost types prefer the annual contract. The users that select the annual contract, therefore, are more likely to be frequent users. In Prediction 2, we use attendance in the initial months $E[v]$ (before the selective exit) as a measure of the likelihood to be a frequent user.

Prediction 2. (Attendance of monthly and annual members) *The average initial attendance of annual members is higher than the average initial attendance of monthly members:*

$$E_A[v] > E_M[v].$$

A third test for the standard model is whether consumers have rational expectations about their attendance.

Prediction 3. (Forecast of attendance) *The average forecast of attendance equals the average actual attendance.*

3.2 Empirical analysis

We test Prediction 1 using the sample of users enrolled in an unsubsidized flat-rate membership in clubs 1 and 2. We analyze separately users in club 3 given the lower fee per visit. As benchmark measure of price per visit, we use the price per visit under the 10-visit pass, \$10, rather than the \$12 visit-by-visit fee: the 10-visit pass is cheaper for users with a monthly or annual contract, given their attendance frequency.[7]

Monthly contract. For users initially enrolled in a monthly contract, we compute the price per expected attendance for each month. We limit the analysis to the first 6 months of tenure to target inexperienced users. We use the sample 'First spell and no subsidy' (912 spells) to ensure comparability to standard health clubs with no corporate subsidy.

The first Column in Table 3 reports the average monthly fees in months 1 through 6, with standards errors in parentheses. The sample for month t consists of users who initially enrolled in a monthly contract and have had a continuous history of membership up to month t under either a monthly or an annual contract. Consumers drop out of the sample when they cancel or are censored. For users who switch to an annual contract, the monthly fee is the monthly share of the annual fee. The average monthly fee exceeds \$80 in all months, except in the joining month which is typically pro-rated, and in month 3, a promotional free month for 18.6 percent of the sample. The average number of visits for users in the tth month of tenure (Column 2) declines from 5.45 in month 2 to 4.32 in month 6. (Month 1 covers only part of a month.)

The third Column in Table 3 presents the ratio of the average fee in month t (Column 1) and the average attendance in month t (Column 2). This ratio is the estimated price per

[7] The (hypothetical) average price per average attendance from using the 10-visit pass, given the distribution of attendance for users enrolled with the monthly and the annual contract, is \$10.86. The benefits of a lower price relative to the \$12 per-visit fee outweigh the losses from unused coupons for these users. The single-visit fee of \$12 is targeted towards one-time users such as travelling businessmen.

expected attendance for month t, $(L/E_F[v])\,a(T)$ in Prediction 1. In each of the six months we reject the hypothesis that the price per expected attendance is smaller than \$10 (or than \$12). The estimate ranges between \$14 and \$16 in the first three months and is higher than \$17 in the subsequent three months. As a summary measure, we compute the ratio of average monthly payment (Column 1) and average monthly attendance (Column 2) in the first six months across all individuals.[8] The resulting price per average attendance in the first six months of enrollment equals \$17.13, well above \$10 (or \$12).

In addition to averages, we consider also the distribution of these measures in the first six months (Table 4). We measure the price per attendance as the ratio of total attendance over total payment in the first six months of membership in a monthly contract (Column 2). Only 20 percent of the individuals pays less than \$10 per visit. The remaining 80 percent would have saved money choosing the pay-per-visit contract, holding constant the number of visits.

Annual contract. We also test Prediction 1 on the users who chose an annual contract at enrollment. We use the sample 'First spell and no subsidy' further restricted to users who joined the club at least 14 month before the end of the sample period. This ensures that we observe the annual contract in its entirety.[9] The final sample consists of 145 spells.

The bottom row of Table 3 presents the estimation results. The average monthly share of the annual fee for the first year (Column 1), adjusted for discounting, is \$71.02.[10] The average number of monthly visits in the first year (Column 2) is 4.69. The resulting price per average attendance (Column 3), \$15.15, is substantially higher than \$10 (or than \$12). The estimate is somewhat lower than for the monthly contract, consistent with selection of users with higher expected attendance into the annual contract (Prediction 2). Table 4 shows the distribution across users of attendance (Column 3) and of the price per attendance (Column 4) in the first year of an annual membership. Only 24 percent *pays* less than \$10 per visit.

Finding 1. (Price per expected attendance at enrollment) *Users who choose an unsubsidized flat-rate contract pay a price per average attendance of over \$17 in the monthly contract and over \$15 in the annual contract. The share of users who pay ex post less than \$10 per visit is 20 percent in the monthly contract and 24 percent in the annual contract.*

Size of the Effect. As a monetary measure of the deviation from the standard model, for monthly and annual memberships we compute the difference between actual expenses over the whole enrollment spell and imputed expenses for the same number of attendances with 10-visit passes[11]. This measure understates the savings from paying per visit since the agents could

[8] For each individual, we compute the average over all available months until the sixth, with the exception of miscoded months and months with freezing. When averaging across individuals, we weigh all individuals equally, independent of tenure.

[9] We exclude 3 annual contracts that are terminated before the 12th month. Health clubs are required to accept cancellations for medical reasons or for relocation more than 25 miles away from the clubs.

[10] We use a daily discount factor of .9998, implying an adjustment factor $T(1-\delta)/\left(1-\delta^T\right)$ equal to 1.037.

[11] This measure takes into account the potential loss associated with not using fully a 10-visit pass.

re-optimize their attendance. The 'average loss' measure is positive if the user would have saved money purchasing 10-visit passes, and negative otherwise. We use the sample 'First spell and no subsidy' for spells that start before October 1997.

The average loss per spell is \$698 for agents initially enrolled in a monthly contract. This amount is 47.87 percent of the overall \$1,517 spent on the health club membership. For agents initially enrolled in an annual contract, there is a small but insignificant gain of \$61.

The observed deviation from the standard model has large monetary consequences for users in the monthly contract. For users in the annual contract, the automatic expiration moderates the possible losses.

Robustness. We now check the robustness of Finding 1.

1. *Sample.* So far we have restricted attention to the unsubsidized sample and pooled the results across clubs. We now include all users in the sample 'First Spell' who initially chose a monthly contract and disaggregate the results by club. Separately for each club, we regress health club attendance on the monthly fee using an Epanechnikov kernel. The measure of attendance is the average attendance per month in the first 6 months. We cross-validate club-by-club with a grid search to compute the optimal bandwidth for the price.[12] In club 1 (Figure 1a), the average monthly attendance from the kernel regression lies between 3 and 5 and is increasing in price, although the estimates are not very smooth given the small bandwidth suggested by the cross-validation. We use the average attendance from the kernel regression to compute the ratio of price and average attendance for each level of price (Figure 1b). The price per average attendance is significantly higher than \$10 for users paying a monthly fee in excess of \$53. The estimates for club 2 are comparable (Figures 1c and 1d) and somewhat smoother given the larger optimal bandwidth. In club 3 the price per average attendance is higher than the per-visit fee of \$8 for users paying a fee in excess of \$46 (Figure 1f).

2. *Underrecording of attendance.* The high price per attendance could result from under-recording of attendance due to a faulty computer system or moral hazard problems with the staff. Health club employees may also seek to avoid queues of users waiting to swipe. The three health clubs in our sample had incentives to address these problems, since they provide reports of attendance to the corporations subsidizing employee memberships. They therefore put in place one of the most advanced and reliable systems to track attendance in the industry. Unlike in most clubs, a front-desk employee collects the cards from the members and swipes them while the member is exercising. Therefore, card swiping does not generate queues. We also witnessed the procedure if a member has forgotten the card: the employee looks the name up in the computer and records the attendance. Thus, while errors may occur in both directions—failure to swipe and double swiping—, the health club data used in our analysis is unusually accurate.

[12] Pagan and Ullah (1999), pp. 110–120.

As supporting evidence, we can test whether random events such as computer crashes or idiosyncratic laziness of employees affect substantially the accuracy of the attendance record. We calculate the fraction of members attending on each day in the sample and regress it on a set of controls: 6 day-of-the-week dummies, 11 month dummies, 3 year dummies, and 15 holiday dummies. If recording precision is highly variable, the R^2 of this regression should be low. The R^2 of the regression for club 1 instead is as high as .8785, with the day-of-the-week dummies explaining most of the variance. The regression for clubs 2 and 3 yield an even higher R^2 of .8915.[13] The high explanatory power of these regressions suggests that daily variation in recording precision is limited.

3. *Ex-post subsidies.* Some HMOs reimburse members partially for health club expenses. To the extent that these reimbursements make the annual and the monthly contract cheaper relative to the pay-per-visit contract, they induce users to choose flat-rate contracts. However, the HMOs in the State where the three clubs operate[14] offer discounts either only on the initiation fee, or both to flat-rate and pay-per-usage contracts.

4. *Membership benefits.* Consumers' choice of the monthly or annual contract could be due to contract-specific membership benefits. The only benefit not available under the per-visit payment, though, is the option to rent an overnight locker at an extra fee, and only 5.52 percent of the users ever rent a locker. If we exclude these users, the results on price per average attendance for the monthly contract do not vary.

Overall, we observe a robust deviation from Prediction 1. Non-subsidized users enrolled in contracts with flat fees pay a price per average attendance that is significantly higher than the per-visit price available as an alternative contract. The result is robust to the type of contract (monthly or annual), the sample (the amount of subsidy), and the club considered. The results do not appear to depend on measurement error, ex-post subsidies, or unobserved benefits. The deviation from Prediction 1 is large: unsubsidized members of a monthly contract pay 70 percent in excess of the $10 fee.

To test Prediction 2 on the initial sorting between the monthly and the annual contract, we compare the average number of visits in months 2, 3 and 4 of tenure for individuals initially enrolled in the monthly and in the annual contract.[15] Given that the price per visit p is zero for both contracts, differences in attendance should reflect differences in the expected future attendance cost. Column 1 of Table 5 reports the results for the sample 'First Spell.' In each

[13] Detailed results are available in DellaVigna and Malmendier (2002), Appendix Table 1.

[14] We report the results in Appendix Table 3 in DellaVigna and Malmendier (2002). We thank Nancy Beaulieu for providing the list of HMOs.

[15] We exclude the first month because attendance is pro-rated over the number of effective days of membership, and the pro-rating procedure is slightly different for the annual and the monthly contract. We do not extend the comparison to months after the fourth since users who experience a high cost can quit under the monthly contract but not under the annual contract.

month, expected attendance is higher under the annual than under the monthly contract, and significantly so in months 3 and 4. Overall, average attendance in months 2 to 4 is 9.5 percent higher under the annual contract. The magnitude of this difference is comparable to variation in average attendance by age groups and by gender. When we break down the sample into 24 age-gender-month cells, average attendance is higher under the annual contract in 20 cells out of 24. Even after controlling for some heterogeneity, individuals with higher attendance are more likely to choose the annual contract at enrollment.

Finding 2. (Attendance of monthly and annual members) *Average attendance in months 2-4 is 9.5 percent higher under the annual contract than under the monthly contract.*

While consumers' choice between flat-rate contracts and a per-visit fee is hard to explain in the standard framework (Finding 1), their choice between the monthly and annual contract (Finding 2) is consistent with standard preferences and beliefs. Consumers sort according to the expected attendance.

Finally, we elicit the expectations of health club members about their future attendance using a survey of 48 randomly chosen respondents interviewed in a mall[16]. The mall is not near a health club, so the respondents are not selected on health club attendance. We ask the ones that report to be members or to attend a health club how often they expect to visit their health club in the next month, September.[17] This question attempts to measure directly whether health club users have rational expectations. Although we do not observe actual attendance among these 48 survey respondents, it is unlikely to differ substantially from attendance in our data set, which is very robust across demographic subgroups. Across 24 (gender)*(club)*(age) subgroups, the average monthly attendance over the membership is lower than 4.75 visits for 23 out of 24 groups, with an overall average of 4.17 monthly visits.

Finding 3 (Forecasts of attendance). *The average forecasted number of monthly visits,* 9.50 *(s.e. 0.66), is more than twice as large as average attendance, 4.17.*

The overestimation displayed by the subjects matches Finding 1. If health club consumers expect to attend 9.5 times per month, they should indeed choose a flat-rate contract, rather than paying per visit.

We also present the subjects with the following scenario: 'Suppose that, based on your previous experience you expect to attend on average 5 times per month (about once a week), if you enroll in a monthly membership. You plan to attend the health club throughout the next year. Would you choose a monthly contract with a monthly fee of $70 per month or 10-visit passes for $100 (each visit costs $10)?' This question attempts to measure whether users endowed with realistic expectations about attendance would still overwhelmingly choose flat-rate

[16]The interviews were done in August 2002 in Walnut Creek, California.

[17]In our sample, average attendance in September is five percent lower than over the rest of the year.

contracts. In the hypothetical scenario, 18 consumers out of 48 prefer the monthly contract, and 30 prefer the 10-visit pass. With realistic expectations about attendance, therefore, the majority of people prefers to pay per visit.

These findings suggest that health club members have unrealistic expectations about their future attendance. However, one should take responses to hypothetical questions with caution, particularly because the survey sample differs from the health club sample.

4 Contract choice over time

4.1 Predictions of the standard model

In the previous Section, we have analyzed consumers' initial choice of membership contract. In this Section, we compare the renewal decisions of monthly and annual members. We take advantage of two differences in the renewal procedure between the two flat-rate contracts. First, the renewal default differs. The monthly contract is automatically renewed and requires a (small) effort—sending a letter or cancelling in person—in order to discontinue the membership. The annual contract automatically expires after 12 months, and cancellation requires no effort. Second, members with a monthly contract can cancel at any month, while members with an annual contract are committed for a year. We evaluate the impact of these differences on cancellation lag, survival probabilities, and average attendance over time in a simple setup with standard preferences and beliefs. (Details are in DellaVigna and Malmendier, 2002).

Calibration. We illustrate the effect of the renewal default on cancellation with the following calibration. Consider two agents with identical preferences and identical effort costs of attendance. One is enrolled in the monthly contract, the other in the annual contract. At the end of the contractual period, each consumer can either renew with a monthly or an annual contract, or switch to the pay-per-visit contract (which is equivalent to dropping out). Denote with s the (possibly negative) daily savings from switching to the pay-per-visit contract, which we assume to be deterministic[18]. The savings s are decreasing in the future health club attendance. For example, a member with a monthly fee of \$70 who expects not to attend any more has $s = \$70/30 = \$2.33$. Denote by δ the daily discount factor and by k the one-time effort cost of cancellation.

Under the annual contract, this cost is zero, and the agent drops out if $\delta s/(1-\delta) > 0$, that is, for $s > 0$. Under the monthly contract, the cost k is stochastic, with i.i.d. draws each period (day) from the c.d.f. F. In each period, the agent can switch to payment per visit at the

[18] For simplicity, we are neglecting the learning over time about the savings s. In a model with learning, agents may wait to cancel for two reasons. First, as we capture in the calibrations, they may wait for a lower realization of k. Second, they may wait for a lower realization of s. Our calibrations show that the predictions are robust to the first option value argument. Adding a second option value regarding s is unlikely to substantially change the predictions.

realized cost k or postpone switching. The benefit of postponement is the option value of a lower future realization of k, while the cost is the foregone savings s. The value function V solves $V = E\left[\max\left(-k, -\delta s + \delta V\right)\right]$. The solution of the agent's dynamic programming problem is a threshold level k^*. The agent switches to payment per visit if the realized transaction cost is smaller than k^*. Without solving for k^*, we derive an upper bound on the expected number of periods (days) until cancellation, $E\left[T\right] = \left(1 - F\left(k^*\right)\right)/F\left(k^*\right)$, under the assumption $\delta = 1$. In Section 4.2 we then compare the predicted $E\left[T\right]$ with an empirical proxy. Denote by $k_{.2}$ the bottom quintile of the cost distribution, that is, $k_{.2} \equiv F^{-1}(.2)$, and denote by $\underline{k}$ the lower bound of the cost distribution. Then $E\left[T\right]$ must be smaller than $\max\left(4, \left[k_{.2} - \underline{k}\right]/s\right)$. The derivation is as follows. For a cost realization of $k_{.2}$, the agents either switch to payment per visit, or not. If they do switch for $k = k_{.2}$, the expected delay is at most $\left(1 - F\left(k_{.2}\right)\right)/F\left(k_{.2}\right) = 4$ days. If they do not switch for $k = k_{.2}$, revealed preferences imply that the benefit of delay—bounded above by $k_{.2} - \underline{k}$—must be higher than the cost of delay, $E\left[T\right]s$. This yields the bound.

In order to calibrate the upper bound for the expected delay $E\left[T\right]$, we make the conservative assumptions $k_{.2} = \$10$ (corresponding to the value of one hour of time on a calm day) and $\underline{k} = 0$. For these values, an individual who expects not to attend the health club any more ($s = \$70/30 = \$2.33$) delays on average no more than $\max\left(4, 10/2.33\right)$, that is, 4.3 days. An individual who expects to attend four times a month ($s \approx \left(70 - 40\right)/30 = \$1$) delays on average no more than 10 days. Under the standard model, therefore, monthly members with low expected attendance switch almost immediately to payment per visit. The switching behavior of monthly members is thus similar to the one of annual members. We summarize a first prediction on contract choice over time.

Prediction 4. (Cancellation lags under the monthly contact) *Low-attenders under the monthly contract delay cancellation for at most a few days.*

Survival probability. We now compare the renewal behavior for monthly and annual contracts when both contracts are up for renewal, i.e., after 12 or 24 months. The survival probability $S_{j,t}$ is the probability that a consumer initially enrolled in contract j (equal to Monthly or Annual) is still enrolled in one of the flat-rate contracts – either monthly or annual – after t months, with $t = 12, 24$. For example, $S_{M,12}$ is the probability that a monthly member has not switched to payment per visit by month 12. Similarly, $S_{A,12}$ is the probability that an annual member renews with an annual or a monthly contract after 12 months.

Sorting at enrollment (Prediction 2) implies that users who selected into the annual contract are ex-post more likely to be frequent users. These users are more likely to renew – either with a monthly or with an annual membership. This increases $S_{A,t}$ relative to $S_{M,t}$. Cancellation costs for the monthly contract, instead, act to increase $S_{M,t}$ relative to $S_{A,t}$. The calibrations above, however, suggest that in a standard model the effect of cancellation costs is very small. We therefore expect the sorting effect to dominate.

Prediction 5. (Survival probability) *The survival probability after one and after two years*

is higher for agents who initially chose the annual membership than for agents who initially chose the monthly membership: $S_{A,t} > S_{M,t}$, *for* $t = 12, 24$.

Attendance over time. Over time, monthly and annual members learn about their attendance patterns, and therefore about s. Learning induces selective exit of individuals with ex-post low attendance. Define as stayers individuals initially enrolled in a flat-rate contract who do not switch to a pay-per-visit contract after the first year. Attendance of stayers in later periods should be higher than attendance of the initial group, since the low-attenders have switched to paying per visit. In the standard model, this prediction holds in similar form[19] for both the annual and the monthly contract.

Prediction 6. (Expected attendance over time for annual contract) *Among users initially enrolled in an annual contract, the expected attendance in the second year among stayers is higher than the expected attendance in the first year for the initial group.*

Prediction 7. (Expected attendance over time for monthly contract) *Among users initially enrolled in a monthly contract, the expected attendance among stayers should increase from month to month.*

4.2 Empirical Analysis

Cancellation lags. To test Prediction 4, we adopt a conservative measure of cancellation delay $E[T]$ for low-attenders. We measure this lag as the number of full months between the last attendance and contract termination for users with a monthly contract at the time of termination. For example, if an agent attends the last time on March 10 and cancels on April 5, we count the 51 days between last attendance (March 10) and membership termination (April 30) as *one* full month. This is likely to understate the true cancellation lag for low-attenders on two grounds: (1) the measure does not include months with low, but positive, monthly attendance and (2) members may attend the club one last time in order to cancel after a long period of non-attendance. We restrict the sample to users who paid no initiation fee, to ensure minimal costs of re-joining.[20]

Finding 4. (Cancellation lags under the monthly contract) *On average, 2.29 full months elapse between the last attendance and contract termination for monthly members, with associated membership payments of \$185. This lag is at least 4 months for 20 percent of the users.*

Even though the transaction costs of cancellation are likely to be lower than \$15 (time cost of sending a cancellation letter or visiting the club), users spend on average \$185 in membership

[19] The main difference is that for the annual contract the comparison can be made only across years, since the selective exit is possible only every 12 months.

[20] We include users with an unsubsidized membership (monthly fee higher than \$70 or annual fee higher than \$700) who joined the club before the month of April 1998.

fees after their last attendance. This lengthy delay is at odds with the calibrations in Section 4.1, which imply an average delay of at most 5 to 10 days.

Survival probability. To test Prediction 5, ideally we would compute the percentage of monthly members and of annual members that are still enrolled one year after the initial enrollment. However, we need to take into account that (1) the first month in a contract is pro-rated, so every annual member is still enrolled in the 13th (calendar) month, and (2) 11.5 percent of annual contracts last one additional month due to promotions. We therefore define the survival probability as the share of members still enrolled in a flat-rate contract at the 15th calendar month. In order to estimate the survival probability, we set survival s_i to 1 if individual i is enrolled in the 15th month of active, paid (monthly or annual) membership, and 0 otherwise.[21] We use the following empirical specification to test Prediction 5:

$$s_i = 1 \text{ if } s_i^* = \alpha + \gamma M_i + BX_i + \varepsilon_i \geq 0, \tag{2}$$

where ε_i is normally distributed and M_i is a dummy variably that equals 1 if the first contract for individual i is a monthly contract, and 0 otherwise. The vector of controls X includes gender, a quadratic function of age, a dummy for corporate membership, a dummy for student membership, 11 dummies for the month and 4 dummies for the year of enrollment. We use the sample 'First spell' restricted to users who joined the club at least 15 active months before the end of the sample period. We also drop users with missing values of a control variable, as well as spells that are censored before the 15th active month.

The coefficient γ captures the difference in survival probability between users initially enrolled in a monthly contract and users initially enrolled in an annual contract. The coefficients in Table 6 are the marginal change in response to an infinitesimal change in the continuous independent variables, and a discrete change for the independent dummy variables. In the specification without controls (Column 1), $\hat{\gamma}$ is positive and significant. Enrollment in a monthly contract increases survival by 5.43 percentage points relative to the baseline rate of 39.83 percent survival with the annual contract. The introduction of the controls increases the coefficient $\hat{\gamma}$ from .0543 to .0719 (Column 2). Controlling for some of the unobserved heterogeneity reduces the downward bias on the coefficient due to the initial sorting (Prediction 2). For example, individuals enrolled with a monthly contract are significantly younger than users with an annual contract (Table 2), and young people are less likely to renew (Column 2 of Table 6). Failing to control for age biases the coefficient $\hat{\gamma}$ downward.

Finding 5 (Survival probability). *The survival probability after 14 months for the monthly contract is 18.1 percent higher than for the annual contract.*

It is worth reiterating that "survival" includes renewal with either of the two flat-rate contracts. We can thus rule out that liquidity concerns (i.e. the difficulty of making an annual

[21] We exclude from the count of active months in the monthly contract promotional periods, months in which the contract was frozen, and months (up to 3 in a row) in which the agent has temporarily quit the club.

payment all at once) and concerns about a second long-term commitment for one year induce annual members to quit.

Robustness. In Columns 3 through 12 of Table 6 we check the robustness of the findings. We measure enrollment at the 16th month after the joining date as an alternative measure of survival. With demographic controls, users initially enrolled in the monthly contract are 5.82 percentage points more likely to be enrolled in the 16th month (Column 4) than users initially enrolled in the annual contract. Alternatively, we measure enrollment at the 27th and 28th month after the joining date (Columns 5 through 8). In the specifications with controls, the estimate of γ is positive, although not significantly different from 0.

We also replicate the results of Columns 1 and 2 of Table 6 for the sample 'First spell and no subsidy' (Columns 9 and 10) and for the larger sample 'First spell' restricted to users who pay at least \$60 per month in the monthly contract or \$600 per year in the annual contract (Columns 11 and 12).[22] In the first, smaller sample the estimated $\hat{\gamma}$ has a similar magnitude as in the benchmark specification, but the estimates are imprecise. In the second, wider sample, the coefficient $\hat{\gamma}$ is positive and large (.0925 with controls), as well as precisely estimated. Overall, the results on survival probability are robust to the measure of past attendance, the measure of survival, and the sample.

Attendance over time. Finally, we test Predictions 6 and 7 on the dynamics of average attendance. We first consider spells starting with an annual contract in the sample 'First spell and no subsidy' and lasting at least two years.[23] We display the results in Columns 1 to 3 of the bottom part of Table 7.

Finding 6 (Average attendance over time in annual contract). *In the annual contract, average monthly attendance for the initial group in the first year, 4.69, is significantly lower than for stayers in the second year, 6.85.*

The difference in attendance between the two groups is large: the baseline group in the first year attends on average 46 percent less than stayers in the second year. Consequently, the price per average attendance in the first year, \$15.15, is significantly higher than in the second year, \$10.77. The results for the larger sample 'First spell' are comparable (Columns 4 to 6 of Table 7).

Figure 2a shows the within-year dynamics of the price per average attendance. The sample at month t is given by users in 'First spell and no subsidy' who have joined with an annual membership and are still enrolled with a flat-rate contract in the t-th month of tenure. Over the first 12 months the price per average attendance increases from 12.5 to 18, as negative shocks accumulate. At renewal (months 13 and 14), the price per attendance is halved.

[22] In both cases, we drop individuals who have missing values for a control or who joined the club later than 14 active months before the end of the sample period.

[23] The results remain unchanged if we restrict the sample further to users who renew with an annual contract after 12 months.

For spells starting with a monthly contract, the sample for average attendance at month t is given by the users in 'First spell and no subsidy' who have joined with a monthly membership and are still enrolled with a flat-rate contract in the t-th month of tenure. Columns 1 to 3 of the top part of Table 7 show the results by six-month groups.

Finding 7 (Average attendance over time in monthly contract). *Average monthly attendance in the first six months of a monthly contract, 4.85, is 26 percent higher than in the next six months and is significantly higher than in any of the later six-month periods among stayers.*

The price per average attendance in the first six months, \$17.13, is significantly lower than in any of the later six-month periods.[24] As Figure 2b shows, the price per average attendance increases over the first 10 months from about \$15 to about \$20, and remains constant thereafter. The results extend to the monthly spells in the sample 'First spell' (Columns 4 to 6).

Summary. Unsubsidized monthly members spends on average \$185 for periods with no attendance before cancellation (Finding 4), despite small transaction costs of cancellation. In addition, after one year, more monthly members are still enrolled in a flat-rate contract than annual members (Finding 5). Surprisingly, members who pay higher fees for the option to cancel each month are more likely to renew past a year. This result does not arise because of sorting but despite sorting (Finding 2). The result is economically and statistically significant and robust across specifications. Finally, average attendance decreases by 26 percent between the first six months and the next six months in the monthly contract (Finding 7), a pattern opposite to the one found for annual contracts (Finding 6).

5 Interpretations

We now consider which assumptions about consumer preferences and beliefs can explain the seven empirical findings, summarized in Table 1. Two Findings are consistent with standard economic models. Health clubs members use information on expected future attendance to sort into the monthly and annual contract (Finding 2) and to sort out of the annual contract (Finding 6). The other findings, instead, are hard to reconcile with the standard framework. Consumers pay \$17 per expected attendance under the monthly contract (Finding 1) and appear to overestimate future attendance (Finding 3). In addition, monthly members with low attendance accumulate delays in cancellation (Finding 4), leading to a higher renewal probability after one year relative to the annual contract (Finding 5). Finally, average attendance among survivers decreases over time for the monthly contract (Finding 7). This finding is puzzling since we observe the opposite pattern for the annual contract (Finding 6).

[24] The results remain unchanged if we restrict the sample further to users who have had a monthly contract at all times until month t.

We first consider if enriched versions of the standard model (Interpretations 1 and 2) can explain the additional findings. We then discuss non-standard preferences and beliefs (Interpretations 3 to 9) as possible explanations. In the end, we summarize which explanations rationalize all the empirical findings.

1. *Risk aversion.* Users who are risk averse in income may prefer a flat-rate contract to the pay-per-visit contract (Finding 1) because the former contract minimizes the variance of the payments.[25] Over the small amounts of money required for a monthly contract, however, health club members should be locally risk neutral (Rabin, 2000).

2. *Transaction costs.* Users may choose a flat-rate contract even though they attend little (Finding 1) if paying per visit entails large transaction costs. For the same reason, they may also postpone the cancellation of a monthly contract (Finding 4). However, the actual transaction costs are small. Users can purchase a ten-visit pass by filling out a simple form, and can then enter the club for ten visits with the same procedure as users with a monthly or annual contract. A transaction-cost-based explanation requires a time cost of over $70 for the few minutes necessary to fill out the form. A related explanation involves psychological transaction costs, such as distaste for payment per visit (Loewenstein and Prelec, 1998). These costs would also need to be high. Moreover, these explanations do not rationalize the overestimation of future attendance (Finding 3) or the differential renewal behavior for annual and monthly contract (Findings 5, 6, and 7).

3. *Membership benefits.* Findings 1 and 4 could arise from psychological benefits of the monthly and annual memberships. These contracts may make the member feel "virtuous" or provide the opportunity to impress others. Arguably, these psychological benefit should apply also to 10-visit passes, since in both cases consumers complete an initial registration procedure and receive a card, which can be shown to friends. However, even if consumers treat monthly and annual memberships as special, it is hard to explain the differential renewal patterns for monthly and annual contracts (Findings 5-7). If anything, the annual contract provides more membership utility, given that it signals a stronger commitment. This would imply a higher survival probability for the annual contract, against Finding 5.[26]

4. *Time-variation in preferences for exercise.* If people enroll whenever they are most enthusiastic about exercise, a rational (but slow) updating process with mean reversion can explain the delay in cancellation (Finding 4) and the decrease in attendance among surviving monthly members (Finding 7). Mean reversion, however, explains neither the initial overpay-

[25] This result requires a utility function that is additively separable in income and health club net benefits. Under the assumption that the utility function is a concave function of the sum of income and health club net benefits, the predictions are reversed: more risk-averse agents are more likely to choose the pay-per-visit contract.

[26] Taste for membership likely implies that high-attendance users switch from the monthly to the annual contract to signal commitment. This switch instead happens for only 1.5 percent of the 7,079 spells initiated with a monthly contract.

ment (Finding 1), nor the difference between renewal patterns of monthly and annual members (Findings 5 and 6).

5. *Limited memory.* Rational agents with limited memory may fail to cancel their monthly membership promptly after they stop attending (Finding 4) because they forget. Distraction can also explain Findings 6 and 7: non-attenders fail to cancel in time, but they get automatically disenrolled under the annual contract. Rational consumers, however, should anticipate their future limited memory and be wary of the monthly contract. Instead, over 90 percent of customers with flat-rate contracts choose the monthly contract (Table 2). In addition, even if we allow for overestimation of future memory, this interpretation does not explain Findings 1 and 3.

6. *Time inconsistency with sophistication.* Flat-rate contracts are attractive to sophisticated agents with (β, δ) preferences (Strotz, 1956; Phelps and Pollak, 1968; Laibson, 1997; O'Donoghue and Rabin, 1999). These agents have, in addition to the usual discount factor δ, a discount factor $\beta < 1$ between present and future payoffs. Their discount function is $1, \beta\delta, \beta\delta^2, ...$ Given that health club attendance involves immediate costs and delayed benefits, such present-biased agents attend the health club less often than they wish at the time of enrollment. They may purchase a flat-fee membership as a commitment device that increases future attendance (Finding 1).

These agents also delay one-time activities with immediate costs, such as contract cancellation. However, the cancellation delays of these agents are too short to account for Findings 4 through 7, as we show with an extension of the calibrations in Section 4.1. Using the same revealed-preference argument, we obtain a bound on cancellation delay[27] for sophisticates given by $E[T] \leq \max(4, [k_{.2}/\beta - \underline{k}]/s)$. Under the calibrated magnitudes[28] $k_{.2} = \$10$, $\underline{k} = 0$, and $\beta = .8$, non-attenders[29] ($s = \$2.33$) delay at most 5.33 days on average. Under the same assumptions, low-attenders ($s = \$1$) delay at most 12.5 days. These bounds do not depend on the assumption $\delta = 1$. We solve the dynamic programming problem as a function of β, assuming a discount factor $\delta = .9995$ (corresponding to a yearly discount factor of .83). We consider the low-attendance case ($s = \$1$) and assume $k \sim N(15, 4)$.[30] The resulting expected cancellation delay $E[T]$ (Figure 3a) is 5 days for $\beta = .8$ and is less than 15 days even for a β as low as .5. This calibrated delay is substantially smaller than the observed delay of over 60 days. Figure 3b shows the corresponding probability of a delay T of over 120 days (4 months). This probability is essentially zero for all β above .4, contrary to the empirical finding that 20

[27] The uniqueness of the equilibrium level of k^* can be proved along similar lines of Proposition 1 in Choi et al. (2004).

[28] Laibson et al. (2004), Paserman (2004) and Shui and Ausubel (2004) estimate the hyperbolic model on field data and find values of β between .5 and .8.

[29] The savings s for sophisticated agents include the benefits of commitment to a higher future attendance under the flat-rate contract (see DellaVigna and Malmendier, 2002).

[30] The results are essentially insensitive to any choice of $\mu \in [10, 30]$ and $\sigma^2 \in [1, 49]$.

percent of users delay for over 4 months. Time inconsistency with sophistication, therefore, cannot generate the delays observed in the data.

7. *Time inconsistency with partial naiveté.* Agents with (β, δ)-preferences may be overconfident about their future self-control and expect to have a discount parameter $\hat{\beta}$, with $\beta < \hat{\beta} \leq 1$ (Akerlof, 1991; O'Donoghue and Rabin, 2001). These (partially) naive agents may pay more than \$10 per expected visit (Finding 1) because they overestimate their future attendance (Finding 3). (This is in addition to the commitment device reason.) We now extend the calibrations in Section 4.1 to show that naive (β, δ) agents may accumulate substantial delays in the cancellation of an automatically-renewed contract. Figure 3a plots the expected cancellation delay for a naive agent with low attendance ($s = \$1$), $\delta = .9995$, and costs $k \sim N(15, 4)$. For $\beta = 0.7$, the cancellation delay of the naive agent matches the delay of over 60 days observed in the data. Moreover, the same level of β matches also the probability of delays lasting over 120 days (Figure 3b), .2. Differently from time-consistent and time-inconsistent sophisticated agents, the predicted delay for naive agents matches the empirical estimates. A model of naive (β, δ) agents, therefore, can explain all the Findings in the paper.[31]

8. *Overestimation of net benefits.* Users may choose flat-rate contracts (Finding 1) because they overestimate the future benefits of attendance or underestimate the expected future costs. Projection bias (Loewenstein, O'Donoghue, and Rabin, 2002) may reinforce the effect if health club consumers have high attendance expectations at sign-up. This interpretation is consistent with Findings 3 and 4, but it does not explain Findings 5, 6, and 8 on higher survival for the monthly than for the annual contract. In order for overestimation to explain all of the empirical findings, consumers need to have unrealistic expectations about both the costs of attendance and the costs of cancellation. This is the case if consumers overestimate their future efficiency, that is, their ability to perform desirable tasks such as health club attendance and contract switching.

9. *Persuasion.* Given that users attend on average less than eight times per month, flat-rate contracts are on average more profitable for the health clubs than pay-per-visit contracts. Health club employees, therefore, have incentives to persuade consumers to sign flat-rate contracts. They can do this either by not providing (sufficient) information about the pay-per-visit alternative or by urging people to take up the monthly or annual contract. We address the first concern, underprovision of information, by considering the contractual choices of a subgroup that is surely well-informed. In our data, members of a specific HMO can choose between a 20% discount on the flat-rate contracts and a \$6 payment per visit. Members claiming the discount must have obtained the information from the HMO itself, which explicitly lists both options. Nevertheless, the price per expected attendance over months 1 to 6 for HMO members

[31] The amount of delay predicted by the naive model is decreasing in the variance of the cost distribution. For σ substantially larger than 4, the calibrations of the naive model do not match the data.

enrolling with a monthly contract equals \$10.65 (s.e. 0.25), significantly higher than the \$6 price per visit. Thus, even informed members display the tendency to choose the more costly flat-rate contract.

Alternatively, health club employees may exert pressure on members to choose a flat-rate contract (Finding 1) (Bernheim, 1994). Employee persuasion may explain also the cancellation lag for the monthly contract (Finding 4), even though members can also cancel in writing. Persuasion is unlikely to explain the difference in renewal between montly and annual contract, though: health club employees can exert pressure to renew both on monthly members and on annual members. Persuasion does not explain the survey evidence of overestimation of attendance (Finding 3).

Out of the nine explanations above, the most successful ones, in our view, involve both overestimation of attendance and overestimation of cancellation. Overestimation of future attendance (Finding 3) leads consumers to choose flat-rate contracts (Finding 1). Overestimation of future cancellation leads consumers to delay cancellation in the monthly contract (Finding 4), but not in the annual contract which requires no cost to cancel (Findings 5 through 7). A model with these features is the partially naive (β, δ) model of O'Donoghue and Rabin (2001), which we calibrate to the data. A model of overestimation of future efficiency makes the same predictions, without reference to self-control. In addition, persuasion by health club employees is a plausible explanation for some of the findings.

Heterogeneity. The leading explanation suggests that one mechanism—overestimation of future self control or of future efficiency— is at the root of all findings. If this is the case, and there is heterogeneity in overestimation, we expect a correlation between the findings. In particular, monthly members that pay a high price per attendance should also be more likely to accumulate a long cancellation lag. This is not necessarily the case if the different findings are driven by different phenomena (such as, for example, risk aversion for Finding 1 and limited memory for Findings 4-7).

We test this prediction for users enrolled in the monthly contract. As a measure of cancellation lags, we use the number of consecutive full months between the last attendance and the expiration (as in Section 4.2). As a measure of price per attendance, we take the ratio of the payments to the health club over the attendance for the period between sign-up and n months before the last attendance, with n equal to 1, 2, 3 and 4. We limit the time frame in order to avoid a spurious correlation between the price per attendance and months of delay due to low attendance in the final months. Finally, we take the log of 1 plus the measures in order to reduce the skewness of both variables. The correlation between the cancellation lag and the price per attendance is positive and significant, with values between .213 ($n = 1$) and .204 ($n = 4$). Longer lags n between the two measures do not affect the estimate, suggesting that the correlation is not likely to be spurious.

Finding 8 (Correlations). *Users who pay a high price per attendance in the monthly contract subsequently display a longer gap between last attendance and contract termination.*

Similarly, we predict that individuals who accumulate a long delay in cancellation should be less likely to freeze a contract if they face a temporary period of non-attendance. We find a highly significant negative correlation of -.1035 between a freezing proxy[32] and the cancellation delay. These results are consistent with the idea that a unique explanation—such as overestimation of efficiency or self-control—drives both the results on the high price per attendance for flat-rate memberships (Section 3.2) and the results on renewal behavior (Section 4.2).

6 Conclusion

How do consumers choose from a menu of contracts? In this paper we consider contract choice in health clubs. Using a new panel data set from three US health clubs, we find that members who choose a contract with a flat monthly fee of over $70 attend on average 4.8 times per month. They pay a price per expected visit of more than $17, even though they could pay $10 per visit using a 10–visit pass. On average, these users forego savings of $700 during their membership. We also find that consumers who choose the monthly, automatically renewed contract are 18 percent more likely to stay enrolled beyond one year than users committing for a year. This is surprising because monthly members pay higher fees for the option to cancel each month. We present additional evidence, including results on cancellation delays and estimates of attendance expectations from a survey. These results are difficult to reconcile with a standard model. We present a number of explanations for the findings. The leading explanations involve overestimation of future efficiency or future self-control.

The analysis of consumer behavior is a first step. Rational, profit-maximizing health clubs can observe the features of consumer behavior using data sets like the one analyzed in this paper. In DellaVigna and Malmendier (2004), we characterize the profit-maximizing contract for goods with immediate costs and delayed benefits, such as health club attendance. For consumers that are overconfident about future self-control – one of the leading explanations in this paper – the profit-maximizing contract involves below marginal cost pricing of attendance and automatic renewal with a transaction cost of cancellation. The typical contract of health clubs in the Boston area indeed has these features. The evidence on contractual design is consistent with the findings on consumer behavior.

[32] The raw measure of freezing is a dummy variable which equals 1 if the individual ever froze before the last attendance. To correct for periods of non-attendance, we run a probit of this dummy on the longest consecutive number of months with no attendance before the last attendance, and the number of periods longer than 2 months with no attendance. The residual of this regression is the final measure of freezing.

7 Data Appendix

The data on consumer behavior come from the attendance panel and the billing records. A 7-digit identification number allows us to link multiple spells of the same individual.

Attendance panel. Each time a user with a flat-rate contract exercises, a staff member swipes the electronic card of the user, and therefore creates an attendance record. An observation of the attendance panel consists of the individual id, the date of the visit, basic demographic information (birthday, gender), a code for short-term memberships, the enrollment and the expiration date (for members that terminated the membership). All information other than the date of visit is constant across the observations for a given individual.

Billing records. The health clubs keep an official record of the customer payments. The billing data provide detailed and accurate information about the category of users—retail (the default), student, family, corporate—as well as the type of transaction. Each line of the billing panel consists of the individual id, the date of the contractual transaction, the 4-digit code that identifies the transaction, and the price paid (if any). For example, line "1234567 1/1/98 R564 55" indicates that user 1234567 paid an out-of-pocket monthly fee of $55 on January 1, 1998. This monthly fee applies to employees of the company linked to code R564. For the monthly contract, typical transactions are the payment of the initiation fee, the monthly fee, and items such as an overnight locker or a personal trainer. Other codes involve monthly freezes of memberships, bounced payments, and termination of a membership for delinquency in the payments. For the annual contract, typical transactions are the payment of the initiation fee and of the annual fee.

We use the price stated in the records as a measure of the monetary payments to the clubs. We could alternatively use the 4-digit code and a conversion table (based on the prices as of August 2000) to recover an imputed price. The correlation between the two measures of price is .9668. None of the results changes if we use the imputed price instead of the actual price.

Monthly panel. We merge the attendance and the billing panel into a unique data set, and we then transform the data into a balanced panel with monthly observations. Each observation consists of a variable defining the membership (not enrolled/enrolled in a monthly contract/enrolled in an annual contract/in a freeze), the number of attendances in the month, and the price paid for the month. For an annual contract, the monthly price is 1/12th of the original price. We pro-rate the fees in the first month of monthly and annual contracts that start in the middle of a month. We also pro-rate the fees in the final month of an annual contract. Monthly contracts always terminate on the last day of the month, so no pro-rating is needed for the last month.

Enrollment spells. We define an *enrollment spell* as the time period of continuous monthly and/or annual membership, including possible freezes of the membership. If no more than one full calendar month of non-enrollment separates two contracts of an individual, we still include them in one spell. For example, this is the case if an annual contract expiring on 1/15/98 is renewed on 3/17/98. The missing monthly payment may be due to an (unrecorded) one-month promotional offer, a delay in payment, or missing data for a monthly payment.

We consider an enrollment spell *censored* if it is either ongoing at the end of the panel or if it is followed by a short-term contract or a promotional membership. Else the spell is completed. Short-term contracts are one-month, two-month, three-month, and four-month memberships with automatic expiration. These are uncommon contracts designed for summer users. We identify promotional contracts as a sequence of months with no contract and attendance in at least half of the months. We assume that in these periods health club members are using a free temporary membership, which the clubs grant in various promotional or charitable initiatives.

References

[1] **Akerlof, George A.** "Procrastination and Obedience," *American Economic Review*, May 1991, *81*(2), pp. 1–19.

[2] **Bernheim, Douglas.** "A Theory of Conformity," *Journal of Political Economy,* October 1994, *102*(5), pp. 841–877.

[3] **Choi, James J., Laibson, David, Madrian, Bridgette, and Metrick, Andrew.** "Optimal Defaults and Active Decisions," mimeographed, December 2004.

[4] **Cutler David M., Glaeser Edward L., and Shapiro, Jesse M.** "Why Have Americans Become More Obese?" *The Journal of Economic Perspectives,* August 2003, *17*(3), pp. 93–118.

[5] **DellaVigna, Stefano and Malmendier, Ulrike. "**Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics,* May 2004, *119*, pp. 353–402.

[6] **DellaVigna, Stefano and Malmendier, Ulrike. "**Overestimating Self-Control: Evidence from the Health Club Industry," *Stanford GSB Research Paper* 1880, October 2002.

[7] **Eliaz, Kfir and Spiegler, Ran.** "Contracting with Diversely Naive Agents," mimeographed, 2004.

[8] **Fielding, Helen.** *Bridget Jones' Diary: A Novel.* Penguin USA, 1999.

[9] **Fielding, Helen.** *Bridget Jones: The Edge of Reason.* Penguin USA, 2001.

[10] **Gabaix, Xavier and Laibson, David.** "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," mimeographed, Harvard University, April 2005.

[11] **Gourville, John T. and Soman, Dilip.** "Payment Depreciation: The Behavioral Effect of Temporally Separating Payments from Consumption," *Journal of Consumer Research,* September 1998, *25*(2), pp. 160–174.

[12] **Heidhues, Paul and Koszegi, Botond.** "The Impact of Consumer Loss Aversion on Pricing," mimeographed, UC Berkeley, 2004.

[13] **Hendel, Igal and Nevo, Aviv**. "Inter-temporal Substitution and Storable Products," *Journal of the European Economic Association*, April 2004, *2*(2), 536-547.

[14] **Laibson, David I.** "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, May 1997, *112* (2), pp. 443–77.

[15] **Laibson, David I., Repetto, Andrea, and Tobacman, Jeremy.** "Estimating Discount Functions from Lifecycle Consumption Choices," mimeographed, Harvard University, January 2004.

[16] **Loewenstein, George and Prelec, Drazen**. "The Red and the Black: Mental Accounting of Savings and Debt," *Marketing Science,* 1998, *17,* pp. 4–28.

[17] **Loewenstein, George, O'Donoghue, Ted D. and Rabin, Matthew**. "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics,* July 2002.

[18] **Madrian, Brigitte C. and Shea, Dennis.** "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics,* November 2001, *116*(4), pp. 1149–1187.

[19] **Miravete, Eugenio J.** "Choosing the Wrong Calling Plan? Ignorance and Learning," *American Economics Review*, March 2003, *93*, pp. 297–310.

[20] **Miravete, Eugenio J. and Roeller, Lars-Hendrik.** "Competitive Non-Linear Pricing in Duopoly Equilibrium: The Early US Cellular Telephone Industry," *CEPR Discussion Paper* 4069, September 2003.

[21] **O'Donoghue, Ted D. and Rabin, Matthew**. " Doing It Now or Later," *American Economics Review*, March 1999, *89*(1), pp. 103–124.

[22] ——. "Choice and Procrastination," *Quarterly Journal of Economics*, February 2001, *116*(1), pp. 121–160.

[23] **Pagan, Adrian and Ullah, Aman.** *Nonparametric Econometrics.* Cambridge University Press: Cambridge, 1999.

[24] **Paserman, M. Daniele**. "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation," mimeographed, Hebrew University of Jerusalem, August 2001.

[25] **Phelps, Edmund S. and Pollak R.A**. "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies,* April 1968, *35*(2), pp. 85–199.

[26] **Rabin, Matthew.** "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica,* September 2000, *68*(5), pp. 1281-1292.

[27] **Shui, Haiyan and Ausubel, Lawrence M.** "Time Inconsistency in the Credit Card Market," mimeographed, University of Maryland, May 2004.

[28] **Strotz, Robert H**. "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies,* 1956, *23*(3), pp. 165–180.

[29] **Tirole, Jean**. *The Theory of Industrial Organization.* MIT Press: Cambridge, 1988.

Figure 1. Average attendance and price per average attendance (Kernel regressions)

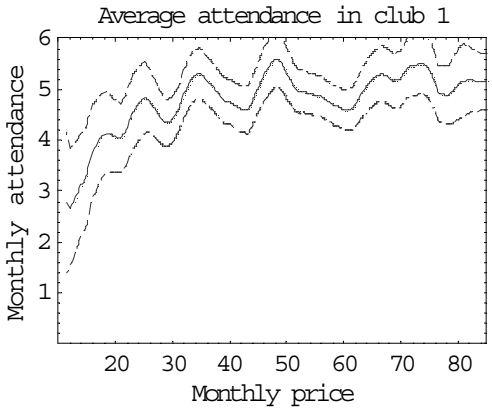


Figure 1a. Kernel regression of attendance on price (club 1, bandwidth 4).

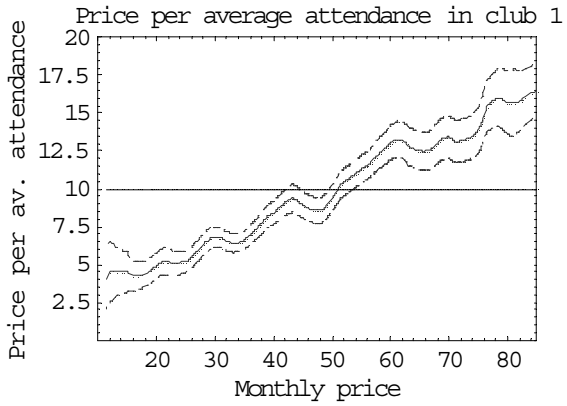


Figure 1b. Price per average attendance as a function of the monthly price (club 1, bandwidth 4).

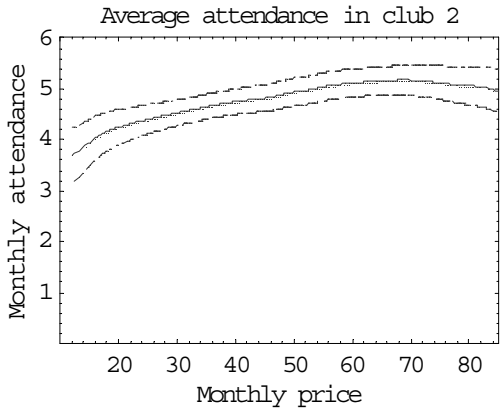


Figure 1c. Kernel regression of attendance on monthly price (club 2, bandwidth 16).

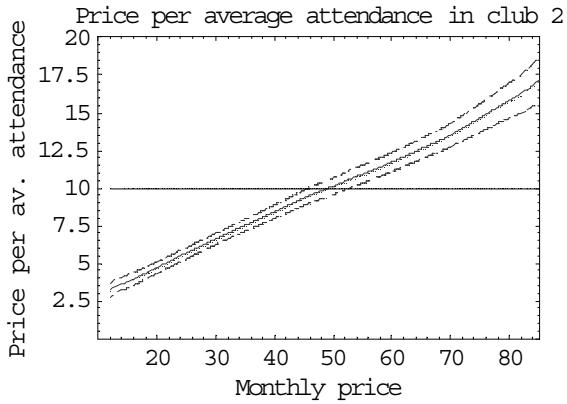


Figure 1d. Price per average attendance as a function of the monthly price (club 2, bandwidth 16).

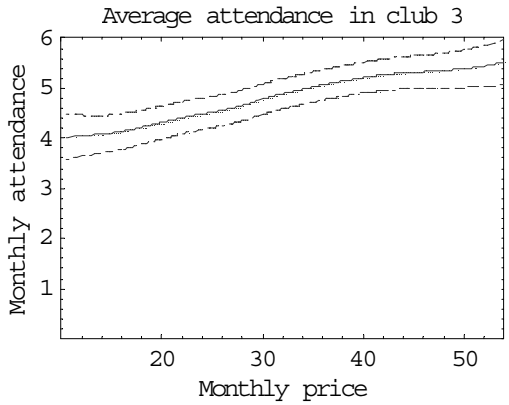


Figure 1e. Kernel regression of attendance on price (club 3, bandwidth 16)

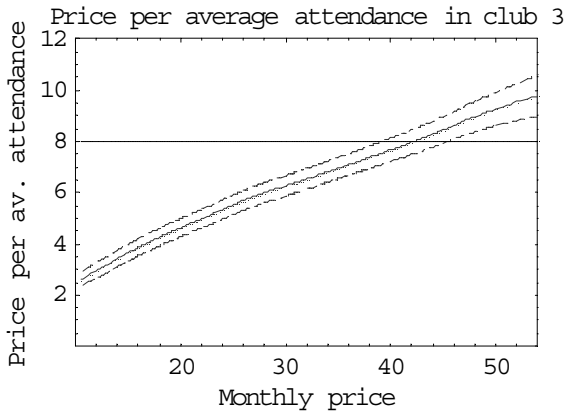


Figure 1f. Price per average attendance as a function of the monthly price (club 3, bandwidth 16).

Notes: Point estimates and 95 percent confidence intervals plotted. The sample is 'First spell' for individuals initially enrolled with a monthly contract. The individual price variable is the average price over the first six months. The individual attendance variable is the average attendance over the first six months. Figures 1a, 1c, and 1e show a kernel regression of attendance on price using an Epanechnikov kernel. The bandwidth is determined by cross-validation with a grid search separately for each club. Figures 1b, 1d, and 1f show the ratio of the price and the expected attendance predicted for that price using the kernel regression. Confidence intervals are derived using the Delta method.

Figure 2. Price per average attendance over time

Figure 2a. Price per average attendance
Annual contracts with annual fee >=$700

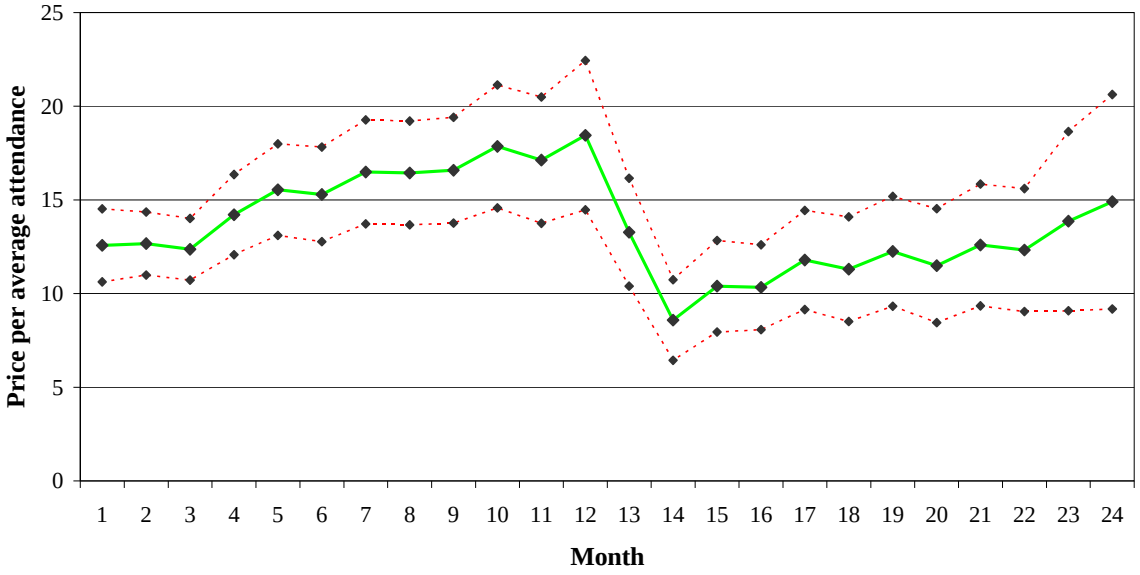
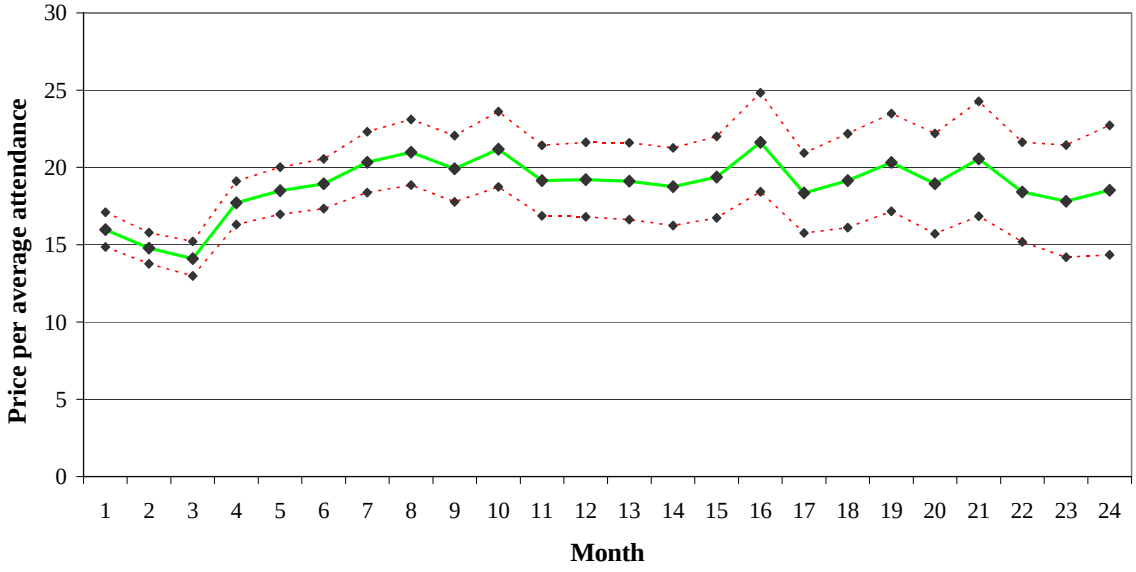


Figure 2b. Price per average attendance
Monthly contracts with monthly fee>=$70



Notes: Point estimates and 95 percent confidence intervals plotted. Figure 2a plots the ratio of average price and average attendance at month *n* of tenure. The sample is 'First spell and no subsidy, all clubs' for individuals initially enrolled in the annual contract and still enrolled at month *n* of tenure. Figure 2b plots the ratio of average price and average attendance at month *n* of tenure. The sample is 'First spell and no subsidy, all clubs' for individuals initially enrolled in the monthly contract and still enrolled at month *n* of tenure. Standard errors for the ratio of average price and average attendance computed using the bivariate Delta method.

Figure 3. Calibration of expected delay in cancellation

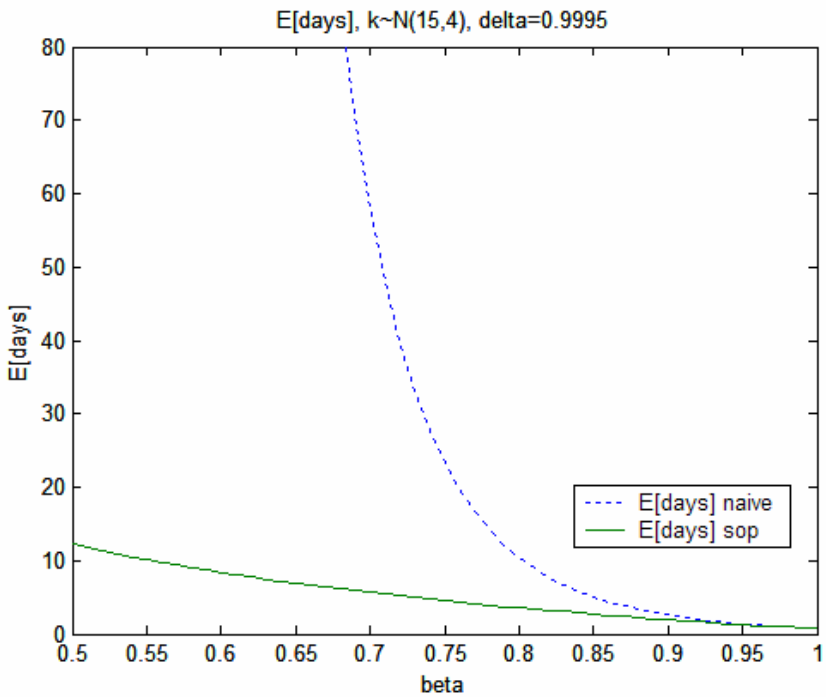


Figure 3a. Simulated expected number of days before a monthly member switches to payment per visit. Assumptions: Cost *k~N(15,4)*, daily savings *s=1* and daily discount factor delta=.9995. The observed average delay is 2.3 months (70 days) (Finding 4).

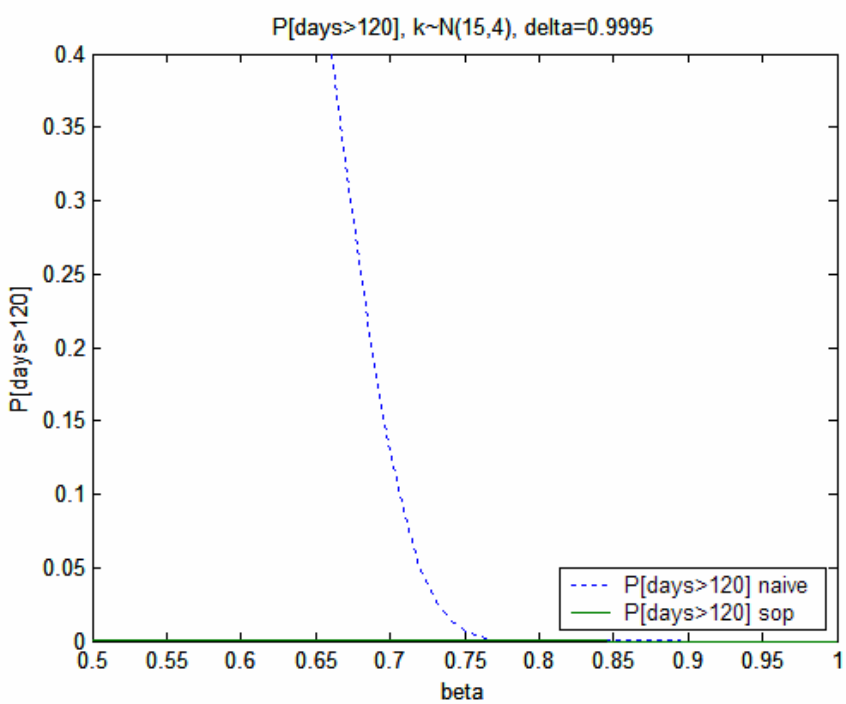


Figure 3b. Simulated probability that cancellation delays last more than 120 days. Assumptions as in Figure 3a. The observed share of agents with delay over 120 days is 20 percent (Finding 4).

Table 1: Empirical Findings and Possible Explanations

	Standard model	Trans. costs of payment per usage	Membership benefits per usage	Limited memory	Time inconsist. with sophistication	Time inconsist. with naivete'	Overestimation of future efficiency	Persuasion
Finding 1. Price per average attendance = $17		distaste of pay per usage	membership benefits		commitment	commitment, overestimation of attendance	overestimation of attendance	pressure of salesman
Finding 2. Average attendance in first 4 months higher in annual than monthly contract	sorting at enrollment	sorting at enrollment	sorting at enrollment	sorting at enrollment	sorting at enrollment	sorting at enrollment	sorting at enrollment	sorting at enrollment
Finding 3. Users predict 9.5 monthly visits; actual monthly visits are 4.2						overestimation of attendance	overestimation of attendance	
Finding 4. Interval between last attendance and termination 2.3 full months		distaste of pay per usage	membership benefits	forget to cancel		overestimation of cancellation	overestimation of cancellation	pressure of salesman
Finding 5. Survival probability at 14th month 18.1 percent higher for monthly than for annual contract				forget to cancel		overestimation of cancellation	overestimation of cancellation	pressure of salesman
Finding 6. Average attendance 46 percent higher in second year for annual contract	learning, sorting out	learning, sorting out	learning, sorting out	learning, sorting out	learning, sorting out	learning, sorting out	learning, sorting out	learning, sorting out
Finding 7. Decreasing average attendance over time in monthly contract				forget to cancel		overestimation of cancellation	overestimation of cancellation	pressure of salesman
Finding 8. Positive correlation of price per average attendance and interval between last attendance and termination						heterogeneity in naiveté	heterogeneity in overconf.	

Table 2: Descriptive Statistics

	Sample: First spell				**Sample:** First spell		**Sample:** First spell and no subsidy	
	Club 1	Club 2	Club 3	All clubs	All clubs		All clubs	
	All Contr. (1)	All Contr. (2)	All Contr. (3)	All Contr. (4)	First Contract Monthly (5)	First Contract Annual (6)	First Contract Monthly (7)	First Contract Annual (8)
Number of spells								
total	3548	2984	1446	7978	7079	899	912	208
completed spells	2440	1850	994	5284	4775	509	588	111
Total Amount in $	569.05	548.97	312.24	515.00	501.28	622.97	920.51	1041.80
	(509.94)	(559.85)	(307.50)	(509.07)	(513.41)	(459.66)	(713.68)	(543.94)
	N = 3548	*N* = 2984	*N* = 1446	*N* = 7978	*N* = 7079	*N* = 899	*N* = 912	*N* = 208
Initiation fee	6.32	1.99	2.85	4.07	3.87	5.62	14.46	17.07
	(26.64)	(12.16)	(12.88)	(20.13)	(19.45)	(24.82)	(41.66)	(45.15)
	N = 3548	*N* = 2984	*N* = 1446	*N* = 7978	*N* = 7079	*N* = 899	*N* = 912	*N* = 208
Average fee per month								
monthly contract	52.40	49.37	31.36	47.42	47.31	56.50	78.52	74.82
	(18.34)	(18.93)	(10.91)	(19.10)	(19.05)	(20.51)	(5.07)	(15.33)
	N = 3185	*N* = 2663	*N* = 1314	*N* = 7162	*N* = 7079	*N* = 83	*N* = 912	*N* = 22
annual contract	48.33	43.76	24.15	42.83	45.94	42.48	69.89	66.26
	(18.35)	(17.34)	(8.62)	(17.50)	(15.77)	(17.66)	(4.19)	(4.21)
	N = 445	*N* = 405	*N* = 151	*N* = 1001	*N* = 102	*N* = 899	*N* = 7	*N* = 208
Average attendance per month								
monthly contract	4.06	4.02	3.75	3.99	3.98	4.62	3.96	5.45
	(3.85)	(3.82)	(3.67)	(3.81)	(3.81)	(3.79)	(3.77)	(4.15)
	N = 3185	*N* = 2663	*N* = 1314	*N* = 7162	*N* = 7079	*N* = 83	*N* = 912	*N* = 22
annual contract	4.45	4.22	4.16	4.31	5.76	4.15	6.07	4.26
	(3.90)	(4.06)	(3.98)	(3.97)	(4.20)	(3.92)	(4.04)	(3.87)
	N = 445	*N* = 405	*N* = 151	*N* = 1001	*N* = 102	*N* = 899	*N* = 7	*N* = 208
Contract choice per spell								
months with monthly contract	9.18	8.91	8.86	9.02	10.11	0.44	11.70	0.55
	(8.42)	(9.14)	(8.91)	(8.78)	(8.70)	(2.12)	(9.06)	(2.35)
	N = 3548	*N* = 2984	*N* = 1446	*N* = 7978	*N* = 7079	*N* = 899	*N* = 912	*N* = 208
months with annual contract	1.58	1.95	1.41	1.69	0.15	13.76	0.08	15.16
	(4.75)	(5.79)	(4.84)	(5.18)	(1.52)	(7.47)	(1.08)	(7.97)
	N = 3548	*N* = 2984	*N* = 1446	*N* = 7978	*N* = 7079	*N* = 899	*N* = 912	*N* = 208
freezing	0.28	0.30	0.18	0.27	0.30	0.05	0.37	0.04
	(0.97)	(1.12)	(0.73)	(1.00)	(1.05)	(0.37)	(1.22)	(0.32)
	N = 3548	*N* = 2984	*N* = 1446	*N* = 7978	*N* = 7079	*N* = 899	*N* = 912	*N* = 208
Female	0.44	0.49	0.47	0.46	0.48	0.34	0.38	0.35
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.47)	(0.49)	(0.48)
	N = 3539	*N* = 2984	*N* = 1446	*N* = 7969	*N* = 7071	*N* = 898	*N* = 912	*N* = 208
Age at sign-up	30.71	31.54	35.05	31.81	31.52	34.04	33.14	34.40
	(8.43)	(8.94)	(9.27)	(8.92)	(8.78)	(9.65)	(9.70)	(10.78)
	N = 3343	*N* = 2855	*N* = 1363	*N* = 7561	*N* = 6710	*N* = 851	*N* = 855	*N* = 197
Corporate member	0.43	0.61	0.43	0.50	0.50	0.53	0.16	0.16
	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.37)	(0.37)
	N = 3548	*N* = 2984	*N* = 1446	*N* = 7978	*N* = 7079	*N* = 899	*N* = 912	*N* = 208
Student	0.05	0.00	0.00	0.02	0.02	0.01	0.00	0.00
	(0.22)	(0.04)	(0.06)	(0.15)	(0.15)	(0.12)	(0.06)	(0.07)
	N = 3548	*N* = 2984	*N* = 1446	*N* = 7978	*N* = 7079	*N* = 899	*N* = 912	*N* = 208

Notes: Standard deviation in parentheses. An enrollment spell starts whenever an individual enrolls (or re-enrolls) in the club and ends whenever the individual quits or is censored. The sample "First spell" consists of the first enrollment spell. The sample "First spell and no subsidy" restricts the sample "First spell" to those spells in which the average adjusted monthly fee is at least $70 if the spell starts with a monthly contract and at least $58 if the spell starts with an annual contract. The spells in column "First Contract Monthly" start with a monthly contract. The spells in column "First Contract Annual" start with an annual contract. "Average price per month" refers to the out-of-pocket fee in the case of corporate users.

Table 3: Price per Average Attendance at Enrollment

	Sample: First spell and no subsidy, all clubs		
	Average price per month	Average attendance per month	Average price per average attendance
	(1)	(2)	(3)
	Users initially enrolled with a monthly contract		
Month 1	55.09	3.45	15.98
	(0.78)	(0.13)	(0.57)
	N = 873	*N* = 873	*N* = 873
Month 2	80.53	5.45	14.78
	(0.44)	(0.18)	(0.51)
	N = 797	*N* = 797	*N* = 797
Month 3	70.02	4.97	14.09
	(1.04)	(0.18)	(0.57)
	N = 780	*N* = 780	*N* = 780
Month 4	81.72	4.61	17.71
	(0.26)	(0.19)	(0.72)
	N = 766	*N* = 766	*N* = 766
Month 5	81.87	4.43	18.50
	(0.25)	(0.18)	(0.78)
	N = 701	*N* = 701	*N* = 701
Month 6	81.88	4.32	18.94
	(0.28)	(0.19)	(0.82)
	N = 639	*N* = 639	*N* = 639
Months 1 to 6	83.00	4.85	17.13
	(0.40)	(0.14)	(0.52)
	N = 912	*N* = 912	*N* = 912
	Users initially enrolled with an annual contract, join 14 month before the end of sample period.		
Year 1	71.02	4.69	15.15
	0.50	0.38	1.24
	N = 145	*N* = 145	*N* = 145

Notes: Standard errors in parentheses. Standard errors for "Average price per average attendance" measure computed using the bivariate Delta method. The number of observations is denoted by N. An enrollment spell starts whenever an individual enrolls (or re-enrolls) in the club and ends whenever the individual quits or is censored. The sample "First spell" consists of the first enrollment spell. The sample "First spell and no subsidy" restricts the sample "First spell" to those spells in which the average adjusted monthly fee is at least $70 if the spell starts with a monthly contract and at least $58 if the spell starts with an annual contract. The sample for the *t*-th month includes spells that are ongoing, not frozen, and not miscoded at month *t*. For the 6-month period, the sample includes spells that are ongoing, not frozen, and not miscoded in at least one month in the period. For the 1-year period in the annual contract, the sample includes only spells that started at least 14 months before the end of the sample period, and that were not prematurely terminated because of medical reasons or relocation.

The "Average price" in period t is the average fee across people enrolled in period t. The "Average attendance" in period t is the average number of visits across people enrolled in period t. The measure in Column (3) is the ratio of the measure in Column (1) and the measure in Column (2).

Table 4: Distribution of Attendance and Price per Attendance at Enrollment

	Sample: First spell and no subsidy, all clubs			
	First contract monthly, Months 1-6 (monthly fee >= $70)		First contract annual, Year 1 (annual fee >= $700)	
	Average attendance per month (1)	Price per attendance (2)	Average attendance per month (3)	Price per attendance (4)
Distribution of measures				
10th percentile	0.33	7.59	0.23	6.05
20th percentile	1.00	10.17	0.86	8.67
25th percentile	1.50	11.39	1.23	10.65
Median	3.91	20.89	3.58	20.34
75th percentile	7.00	58.39	6.58	59.82
90th percentile	10.75	107.50	11.00	119.64
95th percentile	12.83	170.00	13.25	239.28
	N = 912	N = 912	N = 145	N = 145

Notes: The number of observations is denoted by N. An enrollment spell starts whenever an individual enrolls (or re-enrolls) in the club and ends whenever the individual quits or is censored. The sample "First spell" consists of the first enrollment spell. The sample "First spell and no subsidy" restricts the sample "First spell" to those spells in which the average adjusted monthly fee is at least $70 if the spell starts with a monthly contract and at least $58 if the spell starts with an annual contract. The spells in column "First Contract Monthly, months 1-6" start with a monthly contract. The spells in column "First Contract Annual, year 1" start with an annual contract. The variable "Price per attendance" is defined as the ratio of the average price over the average attendance over the firsy period (6 months for the monthly contract, one year for the annual contract).

Table 5: Average Attendance in Monthly and Annual Contracts (Sorting)

	Average attendance during the n-th month since enrollment **Sample:** First spell, all clubs		
	Month 2	Month 3	Month 4
Monthly contract	5.500	4.998	4.592
	(0.0658)	0.069	0.070
	N = 6380	*N* = 5783	*N* = 5390
Annual contract	5.797	5.583	5.151
	(0.1865)	0.191	0.188
	N = 874	*N* = 858	*N* = 839

Notes: Standard errors in parentheses. The number of observations is denoted by N. An enrollment spell starts whenever an individual enrolls (or re-enrolls) in the club and ends whenever the individual quits or is censored. The sample "First spell" consists of the first enrollment spell. The spells in row "Monthly Contract" start with a monthly contract. The spells in row "Annual Contract" start with an annual contract. The sample in month n includes spells that are ongoing, not frozen, and not miscoded.

Table 6: Probit of Renewal Decision.

Sample:	First spell with non-missing controls, all clubs								No subsidy I, all clubs		No subsidy II, all clubs	
Dependent variable:	Enrollment at the 15th month		Enrollment at the 16th month		Enrollment at the 27th month		Enrollment at the 28th month		Enrollment at 14th active month		Enrollment at 14th active month	
Controls:	No Controls	Controls + Time Dummies	No Controls	Controls + Time Dummies	No Controls	Controls + Time Dummies	No Controls	Controls + Time Dummies	No Controls	Controls + Time Dummies	No Controls	Controls + Time Dummies
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dummy for enrollment with monthly contract	0.0543	0.0719	0.0376	0.0582	-0.0009	0.0262	-0.0016	0.0294	0.0527	0.0465	0.0812	0.0925
	(0.0217)	(0.0219)	(0.0221)	(0.0222)	(0.0261)	(0.0252)	(0.0264)	(0.0252)	(0.0479)	(0.0501)	(0.0370)	(0.0378)
Female		-0.0405		-0.0405		-0.0759		-0.0811		-0.0443		-0.0306
		(0.0143)		(0.0144)		(0.0166)		(0.0167)		(0.0398)		(0.0280)
Age		0.0145		0.0164		0.0239		0.0265		0.0312		0.0253
		(0.0046)		(0.0047)		(0.0054)		(0.0054)		(0.0115)		(0.0083)
Age square		-0.0001		-0.0002		-0.0002		-0.0003		-0.0004		-0.0003
		(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)
Corporate member		0.0747		0.0700		0.0713		0.0705		0.2071		0.0043
		(0.0144)		(0.0145)		(0.0168)		(0.0169)		(0.0484)		(0.0319)
Student member		-0.1151		-0.0953		-0.0916		-0.0767		0.1527		-0.1616
		(0.0503)		(0.0516)		(0.0591)		(0.0604)		(0.2722)		(0.0669)
Month and year of enrollment		X		X		X		X		X		X
Renewal probability for monthly=0	0.39826	0.41621	0.39251	0.40771	0.26769	0.28527	0.2589	0.2730	0.47015	0.54261	0.43662	0.43735
Number of observations	*N* =4990	*N* =4990	*N* =4860	*N* =4860	*N* =2874	*N* =2874	*N* =2777	*N* =2777	*N* =704	*N* =704	*N* =1362	*N* =1362

Notes: Standard errors in parentheses. The number of observations is denoted by N. Entries in the Table represent the marginal coefficients of the probit in response to an infinitesimal change in the continuous variables, and a discrete change for the dummy variables. An enrollment spell starts whenever an individual enrolls (or re-enrolls) in the club and ends whenever the individual quits or is censored. The sample “First spell with non-missing controls” consists of the first enrollment spell for individuals for whom the demographic controls “age” and “female” are available. The sample is further restricted to individuals who join at least 14 active months before the end of the sample period. The sample "No Subsidy I" is a restriction of the sample “First spell with non-missing controls” to individuals paying on average a per-month fee of at least $70. The sample "No Subsidy I" is a restriction of the sample “First spell with non-missing controls” to individuals paying on average a per-month fee of at least $60. See the text for a definition of Enrollment after 14 active months. The controls "Month and year of enrollment" indicate that the probit contains 11 dummies for the month of enrollment and 4 dummies for year of enrollment.

Table 7: Attendance and Price per Average Attendance Over Time

	Sample: First spell and no subsidy, all clubs			**Sample:** First spell, all clubs		
	Average price per month	Average attendance per month	Average price per average attendance	Average price per month	Average attendance per month	Average price per average attendance
	(1)	(2)	(3)	(4)	(5)	(6)
	Users initially enrolled with a monthly contract					
Months 1-6	83.00	4.85	17.13	50.44	4.87	10.36
	(0.40)	(0.14)	(0.52)	(0.25)	(0.05)	(0.12)
	N = 912	*N* = 912	*N* = 912	*N* = 7079	*N* = 7079	*N* = 7079
Months 7-12	82.04	3.59	22.87	53.03	3.91	13.56
	(0.24)	(0.16)	(1.05)	(0.31)	(0.07)	(0.25)
	N = 606	*N* = 606	*N* = 606	*N* = 3961	*N* = 3961	*N* = 3961
Months 13-18	81.47	3.93	20.74	53.03	4.39	12.07
	(0.35)	(0.23)	(1.24)	(0.41)	(0.10)	(0.29)
	N = 339	*N* = 339	*N* = 339	*N* = 2192	*N* = 2192	*N* = 2192
Months 19-24	81.67	3.87	21.10	54.18	4.39	12.35
	(0.37)	(0.29)	(1.61)	(0.58)	(0.13)	(0.39)
	N = 200	*N* = 200	*N* = 200	*N* = 1181	*N* = 1181	*N* = 1181
	Users initially enrolled with an annual contract					
Year 1	71.02	4.69	15.15	47.57	4.48	10.62
	(0.50)	(0.38)	(1.24)	(0.75)	(0.17)	(0.44)
	N = 145	*N* = 145	*N* = 145	*N* = 598	*N* = 598	*N* = 598
Year 2	73.78	6.85	10.77	50.09	6.59	7.60
	(1.06)	(1.00)	(1.57)	(1.81)	(0.49)	(0.60)
	N = 36	*N* = 36	*N* = 36	*N* = 112	*N* = 112	*N* = 112

Notes: Standard errors in parentheses. Standard errors for "Average price per average attendance" measure computed using the bivariate Delta method. The number of observations is denoted by N. An enrollment spell starts whenever an individual enrolls (or re-enrolls) in the club and ends whenever the individual quits or is censored. The sample "First spell" consists of the first enrollment spell. The sample "First spell and no subsidy" restricts the sample "First spell" to those spells in which the average adjusted monthly fee is at least $70 if the spell starts with a monthly contract and at least $58 if the spell starts with an annual contract.

For the 6-month periods, the sample includes spells that are ongoing, not frozen, and not miscoded in at least one month in the period. For year 1 in the annual contract, the sample includes only spells that started at least 14 months before the end of the sample period, and that were not prematurely terminated because of medical reasons or relocation. For year 2, the sample includes only spells that started with an annual contract at least 26 months before the end of the sample period, and that lasted at least 25 months. The spells in row "First contract monthly" start with a monthly contract. The spells in row "First contract annual" start with an annual contract. The "Average price" in period t is the average fee across people enrolled in period t. The "Average attendance" in period t is the average number of visits across people enrolled in period t. The measure in Column (3) is the ratio of the measure in Column (1) and the measure in Column (2).

Technological Advances, Consumer Transaction Costs, and Consumer Welfare

Rajeev K. Tyagi*
Graduate School of Management
University of California, Irvine

January 2002

*Mailing Address: Graduate School of Management, University of California, Irvine, CA - 92697; Phone: 949-824-5584; Fax: 949-725-2850; Email: rktyagi@uci.edu

Technological Advances, Consumer Transaction Costs, and Consumer Welfare

Abstract

Consumers incur many transaction costs in purchasing and using most products. This paper examines the effects of a reduction in such consumer transaction costs caused by market-level technological advances. Using a model in which consumers are simultaneously heterogenous in their transaction costs and in their marginal valuation of product quality, we highlight two mechanisms that can cause such reductions in consumer transaction costs to lower consumer surplus and reduce consumer share of the total social surplus. We also show that when a seller can invest to reduce consumer transaction costs, consumers may be better off with higher transaction costs. Finally, the paper shows how such reductions in consumer transaction costs can either lower or raise consumer prices.

(**Keywords**: *Technological Advances; Consumer Transaction Costs; Product Features; Product Pricing; Economic Model*)

"Everything you buy costs you twice, once for the good itself, and once for the transaction.[1] *"*

1 Introduction

Transaction costs are an inevitable part of consumer purchase and consumption experiences. For almost all of the products we consume, we have to incur the monetary, time, and hassle costs of going to a store, waiting in line, making a payment, sometimes customizing the product to own requirements before using it, sometimes learning how-to-use the product properly, and finally using the product itself.[2] These consumer transaction costs have the following two key features: First, their magnitudes can be large enough to "matter" and affect consumer choice behavior — for example, consumers with "non-technical" backgrounds may not use a software product if its interface is not easy-to-use, consumers with high valuations of time may not visit an Internet business that has slow access speed, and so forth.[3] Second, these consumer transaction costs can reduce due to market-level technological developments — for example, developments in graphical-interface systems have reduced consumer learning and memorization costs involved in using computer operating systems and program-

[1] "The Economics of Waiting," (*Forbes* 1996).

[2] Although we can include search costs also as part of transaction costs, we do not examine the *search* aspect of consumer choice behavior. What we have in mind are costs that *even* perfectly-informed consumers pay while purchasing and consuming products.

[3] Formal surveys and empirical research also show that consumer transaction costs affect consumer choice behavior. For example, Bell, Ho, and Tang (1998) show that consumers' distances to stores affect which store they visit; Crafton (1979) shows that the amount of time needed to shop, once inside a shop, affects consumer choice of store; Kim and Park (1997) show that household opportunity cost of time, based on members' income and employment status, affects household frequency and pattern of grocery shopping trips; Bellante and Foster (1984) find that opportunity cost of time, based on wives' employment status, affect whether and how much outside services household use; Currim, Weinberg, and Wittink (1981) find that consumer distances to performing-art centers affect their demand for performing-art events; a *Greenfield Online* survey (2000) shows that consumers choose on-line shopping to off-line shopping to "save time" (60%) and "avoid crowd" (47%); and a *Forrester Research* survey (1998) shows that consumers choose some web-sites over others based on their "ease of use" (66%) and "speed of downloading" (58%).

ming languages. This paper focuses on reductions in consumer transaction costs caused by such *market-level advances*, and asks the following two main questions: (i) Can a consumer become worse off from such a reduction in its transaction cost? (ii) Can such reductions in consumer transaction costs lower the consumer share of the total social surplus? As secondary questions, we also enquire about the effect of this reduction in consumer transaction cost on seller's profitability, about a seller's incentive to invest in such market-level technological advances, and about the direction of change in consumer prices.

To examine these questions, we construct a simple model in which a single seller sells a single product to consumers that must incur certain transaction costs if they want to purchase and use the seller's product. Consumers are heterogenous not only in their transaction costs, but also in their marginal valuations of product quality. In this framework, we show that reductions in consumer transaction costs caused by market-level technological advances can lower the surplus of some consumers and also lower the consumer share of the total social surplus. Further, such reductions in consumer transaction costs are shown to make the seller better off and, depending on the nature of the sellers' quality-production process, either lower or raise consumer prices. Finally, we show that if the seller can invest to reduce consumer transaction costs, it may have an incentive to reduce these consumer costs by more than what consumers, in aggregate, prefer. That is, a reduction in consumer transaction costs caused by seller's investment may reduce the total consumer surplus.

Although these results are derived in a simple setting (one seller, one product, seller choosing one product feature and price), the intuition is more general and relies on the following three features of markets:

(1) *Market-level* technological advances generally reduce *individual* consumers' transaction costs by different amounts;[4]

(2) Consumers are usually heterogenous in their preferences for product features; and

(3) Fixed costs or increasing returns to scale generally prevent sellers from offering *each* consumer in the market its "perfect" product.

Given that the sellers can not offer products that perfectly suit each and every consumer's preference, the sellers choose product features that suit some consumers' preferences better than others'; and, naturally, the sellers' decisions on which consumers' preferences to satisfy better is influenced by which consumers are more profitable to them. Since market-level technological advances generally reduce transaction costs of different consumers by different amounts, reservation prices of different consumers increase by different amounts, changing the relative profitability of different consumers to the sellers. This may make the sellers alter their product features such that the products now suit better the preferences of a different set of consumers. This change then hurts the set of consumers whose preferences are now less satisfied by the offered product features. Thus, although this set of consumers benefits directly from technological advances reducing their transaction costs, their surplus

[4]For example, development of software graphical interface has reduced total transaction cost (e.g., usage cost) of tech-savvy consumers by lesser amounts than of non-tech-savvy consumers; improvement in road conditions reduce total transaction cost (e.g., time cost of travel) of consumers living closer to a retail outlet by lesser amount than of consumers living farther from the retail outlet, and so forth. In general, consumers' transaction costs reduce by different amounts because of their different opportunity costs of time, different degrees of familiarity with the product, different levels of expertise, etc.

may overall be lowered. And if the loss to this set of consumers exceeds the net gain to the set of consumers whose preferences are now better satisfied by the altered product offerings, the total consumer surplus may also reduce.

Such market-level technological advances can affect consumers adversely through *yet another* mechanism: they can reduce the heterogeneity in the total transaction costs incurred by different consumers in a market,[5] making consumers more *homogeneous* in their "effective" reservation prices (valuation of product minus total transaction cost). And, as is well known, the more homogenous the consumers are in their reservation prices, the higher the proportion of consumer surplus that a seller can extract using its marketing instruments (e.g., Adams and Yellen [1976]). Therefore, market-level technological advances can homogenize consumers' effective reservation prices, leading to consumers able to retain only a smaller share of the total social surplus.[6]

To see these ideas in some context, consider the computer operating system market. In the days of the command-line operating systems such as DOS (where one had to remember and type the commands), tech-savvy consumers had lower transaction costs (learning cost, memorization cost, hassle cost) of using the system and less tech-savvy consumers had higher such transaction costs. Advances in the graphical-interface-systems have clearly reduced the total transaction costs of less tech-savvy consumers by larger amount than those of tech-savvy consumers. This has increased the effective reservation price of the less tech-savvy consumers by more than that of the tech-savvy consumers, making the less tech-savvy

[5] For example, advances in graphical-interface-systems have reduced the difference in transaction costs (of use of computer software) incurred by non tech-savvy consumers and by tech-savvy consumers; improvements in local road conditions reduce the difference in travel costs incurred by consumers living closer to the store and consumers living farther from the store.

[6] Of course, a seller capable of perfect price discrimination can always extract all of consumer surplus. Thus, our story holds for the more realistic case where the seller can not indulge in first-degree price discrimination.

consumers more profitable to operating system sellers (as compared to before). This has made the sellers design their product features keeping more in mind the preferences of these less tech-savvy consumers. Thus, although the tech-savvy consumers have benefitted by ease of use of graphical interface systems, they may have lost overall because the software features are now farther away from their preferences and closer to those of the less tech-savvy consumers. Similarly, in the computer programming-language market, even though the tech-savvy consumers have also benefitted from the introduction of graphical interfaces (e.g., Visual Basic vs DOS-based Basic), they may have become worse-off overall because of the increased price and the changed feature-set of the product (perhaps skewed towards the tastes of less tech-savvy consumers).

In both the above examples of computer operating system software and programming language software, advances in graphical interface systems have played another important role by diminishing the difference in the total transaction costs (of use) borne by tech-savvy and non-tech-savvy users — non-tech-savvy consumers can now use these software *about as easily* as can tech-savvy consumers. This *homogenization* on the dimension of total transaction costs, in turn, implies a homogenization in the effective reservation prices of these consumers, allowing sellers to better extract consumer surplus.

The intuition for why consumer price can either rise or fall is simple. To the extent that a reduction in consumer transaction cost increases the consumers' effective reservation prices, the seller can increase its price. However, as described earlier, the seller may change its product design keeping the preferences of a different set of consumers more in mind now. If this change involves lowering product quality, then the seller can only command a lower price. The relative magnitudes of these two forces – increase in consumers' effective reservation price and change in seller's product quality – determine whether consumers face

a lower or a higher price.[7]

■ **Relationship to Literature.** The results in this paper are related to ideas in a number of streams of literature. The literature on congestion points out that consumers exert negative externality on other consumers by making common facilities such as roads more congested (e.g., Venables 1999); in our model, some consumers exert *negative preference-externality* on other consumers through their "disproportionate" influence on the sellers' product design and pricing policies.

Spence (1976) points out that markets may not offer some consumers their preferred products because the sellers face fixed costs of introducing product variants. Our model has the related idea that, in the presence of seller's fixed costs, heterogeneity in consumer preferences for product features may cause some consumers to exert negative externality on other consumers through product markets. Our model further applies this idea to examine how market-level technological advances reducing individual consumer transaction costs may adversely affect consumer welfare.

The literature on bundling shows that a firm may offer bundled products to homogenize consumer reservation prices, making it easier for it to extract higher consumer surplus (e.g., Adams and Yellen 1976). In our model, exogenous technological advances, by reducing heterogeneity in individual consumer transaction costs, serve the role of homogenizing consumers' effective reservation prices. Thus, even without bundling, a firm in our framework faces more homogenous consumer reservation prices, allowing it to extract higher consumer

[7] To the extent that one can associate Internet-based markets with lower consumer transaction costs, our model implies that we should not expect consumer prices to always decrease. This seems consistent with the mixed findings on the direction of change in consumer prices on Internet-based markets. For example, while some studies find higher consumer prices in some Internet markets (e.g., Lee [1998], Bailey [1998]), other studies report lower consumer prices in such markets (e.g., Brynjolfsson and Smith [2000]). Lal and Sarvary (1999) provide an alternate reason for why the Internet may, in some settings, lead to increased consumer prices.

surplus and increase its share of the pie.

Finally, our paper is also related to the retailing literature that shows that consumer travel costs can influence a retailer's decision about the breadth and depth of the product assortment it carries (e.g., Betancourt and Gautschi [1990], Ratchford and Stoops [1992], and Messinger and Narasimhan [1997]). In our model, consumer transaction costs influence the seller's product design and pricing decisions.

The plan for the remainder of this paper is as follows. In Section 2, we consider the case in which consumer transaction costs reduce due to exogenous technological advances. Section 3 allows the seller to invest and influence the level of consumer transaction costs. Section 4 summarizes the results.

2 The Model

To focus on the key mechanisms, we consider a simple setting in which a single seller offers a single product to consumers. The product is characterized by its price and its quality, both of which are chosen by the seller.

■ **Consumer Description.** Consumers differ on two dimensions. First, each consumer has a certain marginal valuation of product quality, say θ, with $\theta \in [0, 1]$. Second, each consumer incurs a certain transaction cost, say tx, with $x \in [0, 1]$, to purchase and consume the product.

Parameter $t > 0$ is a market-level technological parameter affecting the individual transaction costs incurred by *each* consumer in the market, and parameter x is a consumer-specific parameter. Thus, the higher the market-level technological parameter t, the higher the transaction costs incurred by all consumers in the market; and the higher a given consumer's x,

the higher the transaction cost of this particular consumer. One can think of the following examples to fix ideas about this formulation of total transaction costs: (i) in the case of use of a computer operating-system software, even though everyone in the market may face the same "ease-of-use" features in the operating system (market-level parameter t), differences in consumers' technical expertise and familiarity with the software (x) may lead to consumers incurring different total transaction costs; (ii) in the case of purchase from a retail store, even though all shoppers may face the same number of payment counters and the same service level (market-level parameter t), differences in consumers' time-costs (x) may lead to consumers incurring different total transaction costs.

Consumers, as represented by the parameters (θ, x), are distributed uniformly over the unit square, with a total mass equal to one. A consumer with parameter (θ_i, x_i) gets a normalized utility of 0 if it buys the outside good (i.e., not buy in this product category);[8] on the other hand, if it buys a product with quality q at price p, it gets the following utility:[9]

$$\underbrace{q\theta_i}_{\text{benefit}} - \underbrace{p}_{\text{seller's price}} - \underbrace{tx_i}_{\text{total transaction costs}} .$$

Thus the same level of product quality, q, provides different benefits ($q\theta_i$) to different consumers because of differences in their marginal valuations of quality (θ_i). These differences in θ could arise, for example, from differences in consumers' tastes or incomes. Similarly, the same market-level technological parameter t leads to different total transaction costs (tx_i) to different consumers because of differences in their x. As mentioned earlier, these differences in x could arise from differences in consumers' opportunity cost of time, technical skills,

[8] We ignore any income effect.

[9] Neven and Thisse (1990) use a similar formulation to combine the features of horizontal- and vertical-differentiation frameworks. However, they do not allow consumers the option of buying the outside good, and also assume that quality improvement is costless for the seller.

familiarity with the product, etc.

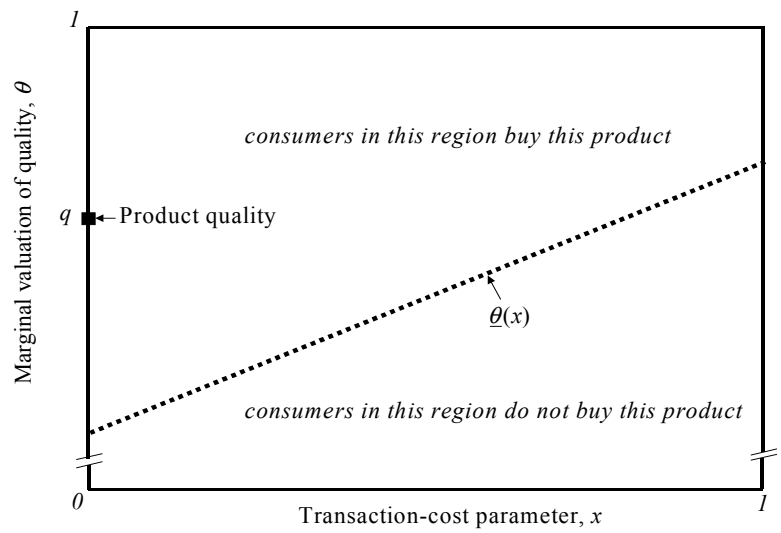


Figure 1: *Consumer-mix facing the seller.*

■ **Demand Structure.** The consumer with transaction-cost parameter x that is indifferent between buying and not buying has $\theta = \underline{\theta}(x)$ where $\underline{\theta}q - p - tx = 0$, or

$$\underline{\theta}(x) = \frac{p}{q} + \frac{t}{q}x. \tag{1}$$

Thus, the seller's demand from consumers with transaction-cost parameter x is $D(x) = 1 - \underline{\theta}(x)$ and its total demand is

$$D(p, q, t) = \int_0^1 D(x)dx = 1 - \frac{p}{q} - \frac{t}{2q}. \tag{2}$$

Figure 1 uses equation (1) to show the demand structure.[10] The line representing the mar-

[10] We confine the technological parameter t to the range $(0, 0.16)$ to rule out the uninteresting case where the total transaction costs are so high that at $x = 1$, even the consumer with the highest marginal valuation of quality is unable to afford to buy the product.

ginal consumers, $\underline{\theta}(x)$, slopes upwards in this figure because if a consumer with high value of parameter x (i.e., with high transaction cost of buying the product) finds it optimal to buy the product, then it must be that he values the product quality highly (i.e., he must have a high θ).

■ **Cost Structure.** On the cost-side, there are the following two main possibilities:

- the burden of quality production falls on the seller's *variable* costs; and
- the burden of quality production falls on the seller's *fixed* costs.

There are several industry examples of each of these two possibilities. For example, producing a higher-quality computer (more advanced microprocessor, more memory, larger hard-disk, etc.) or a higher-quality car (higher-quality engine, leather seats, etc.) requires higher variable cost of production. On the other hand, producing a higher-quality software (more features, smaller code-size, etc.) requires higher upfront fixed costs in terms of manpower, etc., but once the higher-quality software is designed, making copies of it does not cost extra; in many "image-oriented" consumer-product markets, higher advertising expenditures (fixed costs) may enable a seller to increase its "perceived" quality; in retail settings, higher quality may mean more fixed-cost expenditures on ambience, number of counters, and parking spaces, etc.

We next examine how, under each of these two cost structures, market-level advances reducing consumer transaction costs affect consumer prices and consumer surplus.

2.1 Case where burden of quality production falls on seller's variable costs

Here the seller's profit is:

$$\pi = \left(p - q^2\right) D(p, q, t),$$

where the expression for $D(p, q, t)$ is given in (2) and where we assume a quadratic quality-production function.

■ **Consumer Price.** We first examine the effect of a reduction in market-level parameter t on consumer price. Solving the seller's profit-maximizing first-order condition w.r.t. its price p, we get

$$\widehat{p}(q, t) = \underbrace{-\frac{t}{4}}_{\text{reservation-price effect}} + \underbrace{\frac{q(t) + q(t)^2}{2}}_{\text{quality-effect}}. \tag{3}$$

Thus, a market-level advance that reduces consumer transaction costs affects consumer price through the following two routes: (i) *Direct reservation-price route.* Consumers' effective reservation prices increase, which increases the seller's ability to charge a higher price; (ii) *Indirect product-quality route.* The seller may find it optimal to change product quality, which affects its ability to charge a higher or a lower price (because buyers pay higher price for higher quality).

Overall, if the seller reacts to market-level advances reducing consumer transaction costs by increasing its product quality, then both the above-mentioned mechanisms raise the consumer price. However, if the seller reacts to this decrease in consumer transaction costs by decreasing its product quality, then the two mechanisms affect seller's price in opposite directions and the net result could be either a decrease or an increase in consumer price.

So, we next examine how a reduction in parameter t affects the seller's quality choice.

Inserting the value of $\widehat{p}(q,t)$ from (3) in the seller's profit function gives

$$\pi = \left[\frac{4q(q-q^3-t)+t^2}{16q}\right] - \left[\frac{q(2q-2q^2-t)}{4}\right]$$

where the expression in the first square bracket is the seller's revenue and that in the second square bracket is the seller's cost. The choice of optimal quality q is determined by the first-order condition $\partial\pi/\partial q = 0$, or:

$$\underbrace{\frac{1}{16}\left[4-12q^2-\frac{t^2}{q^2}\right]}_{\text{marginal revenue}} = \underbrace{\frac{1}{4}\left[4q-6q^2-t\right]}_{\text{marginal cost}}. \tag{4}$$

One can then see that a decrease in market-level parameter t increases both the seller's marginal revenue and the seller's marginal cost. Intuitively, a reduction in t (i) increases the seller's marginal revenue because now all consumers are richer and can pay more for any given quality; (ii) increases the seller's marginal cost because more consumers can now afford to buy the seller's product and the seller has to incur quality-production cost for *each* unit it sells (since quality production is a variable cost here). Overall, the net effect of a reduction in t on seller's optimal quality depends on which of these two effects – the increase in marginal revenue or the increase in marginal cost – is larger. Solving equation (4), we get

$$q^*(t) = \frac{1}{6}\left[1+\sqrt{1+6t}\right], \tag{5}$$

which clearly decreases as t reduces. Thus the seller reacts to market-level advances reducing consumer transaction costs by decreasing its product quality.

Inserting the expression of $q^*(t)$ from (5) in that of $\widehat{p}(q,t)$ in (3) gives the following

expression for optimal price:[11]

$$p^*(t) = \frac{1}{18}\left[2 - 3t + 2\sqrt{1+6t}\right], \tag{6}$$

which also decreases as t decreases (in the relevant range of t mentioned in footnote 10). Thus, although a reduction in consumer transaction costs implies an increase in consumer effective reservation prices – which alone would imply that the seller can charge a higher price – the effect of decreased product quality overwhelms this reservation-price effect in our setup, and hence the seller charges a lower consumer price.

■ **Consumer Mix and Consumer Surplus**. We next examine how a reduction in market-level parameter t affects the marginal consumers and consumer surplus. Using (5) and (6) in (1), we find that

$$\underline{\theta}(t,x) = \frac{2 - 3t + 2\sqrt{1+6t} + 18tx}{3 + 3\sqrt{1+6t}}.$$

Since

$$\frac{\partial \underline{\theta}(t,x)}{\partial t} = \frac{-1+6x}{2\sqrt{1+6t}} \lesseqgtr 0 \text{ for } x \lesseqgtr \frac{1}{6},$$

we can see that a decrease in t brings in some new consumers to this market (with $x > 1/6$, shown in Area D in Figure 2), but excludes some existing consumers from this market (with $x < 1/6$, shown in Area A in Figure 2).

We next examine the effect of a reduction in t on total consumer surplus, which is

$$CS = \int_0^1 \int_{\underline{\theta}}^1 (\theta q - p - tx) d\theta dx.$$

[11] We confirm that the second-order conditions are satisfied in the relevant range of t (mentioned in footnote 10) by the optimal roots given in (5) and (6), and hence that these roots indeed maximize the seller's profit.

Calculated at p^* and q^*, this yields

$$CS^*(t) = \frac{63t^2 + 2\left(1+\sqrt{1+6t}\right) - 6t\left(1+2\sqrt{1+6t}\right)}{108\left[1+\sqrt{1+6t}\right]}, \text{ and} \tag{7}$$

$$\frac{\partial CS^*(t)}{\partial t} = \frac{(-2+6t\sqrt{1+6t}) + (-2\sqrt{1+6t}+63t^2)}{24\sqrt{1+6t}\left[1+3t+\sqrt{1+6t}\right]} < 0.$$

Thus technological advances reducing consumer transaction costs increase the overall consumer surplus. To see the effect on the surplus of an *individual* consumer with parameter (x, θ), consider the following expression for this consumer's surplus:

$$CS(\theta, x) = \theta q^* - p^* - tx = \frac{1}{18}\left[\left(1+\sqrt{1+6t}\right)(3\theta - 2) + t(3-18x)\right].$$

Since

$$\frac{\partial CS(\theta, x)}{\partial t} = \frac{1}{6}\left[1 - 6x + \frac{3\theta - 2}{\sqrt{1+6t}}\right]$$

is not uniformly negative in the whole parametric space describing the existing consumers (i.e., in the $\theta \in [\underline{\theta}, 1]$ and $x \in [0, 1]$ space), we know that some existing consumers become worse off with a reduction in their transaction cost tx. We further analyze this expression for $\partial CS(\theta, x)/\partial t$ and show in Figure 2 that the following two types of existing consumers become worse-off: (i) those who could afford the product earlier (mainly because of their low transaction costs), but can no longer do so [Area A in Figure 2] ; and (ii) those who continue to buy the product, but lose because of the changed quality-price offering [Area B in Figure 2].

Overall, as shown in Figure 2, consumers who lose are the ones with low transaction-cost parameter (x) and high marginal valuations of quality (θ). This is intuitive since these consumers have low transaction costs to start with (i.e., low tx), and hence have the least

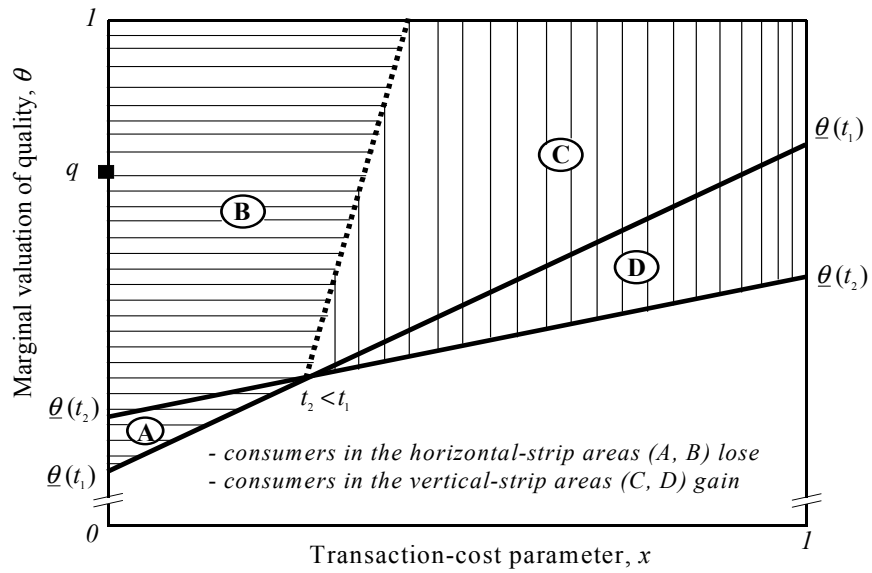


Figure 2: *Effects of reduced consumer transaction costs on consumer surplus (burden of quality production on seller's variable cost).*

to gain from technological changes reducing these costs; and since they value the quality highly (high θ), they have the most to lose from the reduction in product quality caused by a reduction in consumer transaction cost.

■ **Seller's Profitability.** Using the equilibrium expressions of p^* and q^* in the expression for seller's profit, we get

$$\pi^*(t) = \frac{\left[1 - 6t + \sqrt{1+6t}\right]^2}{54\left[1 + \sqrt{1+6t}\right]}, \text{ and} \tag{8}$$
$$\frac{\partial \pi^*(t)}{\partial t} = \frac{-(1 + 3t + \sqrt{1+6t}) + 18t^2}{3\sqrt{1+6t}\left[1 + \sqrt{1+6t}\right]^2} < 0.$$

Thus the seller benefits from a reduction in consumer transaction costs. This is intuitive because a reduction in consumer transaction cost not only increases the effective reservation prices of all the existing consumers, but can also bring in some new consumers to the seller's

market. Thus, the seller can not be worse off.

■ **Consumer Share of the Total Social Surplus.** Using (7) and (8), we find that a reduction in market-level parameter t decreases the consumer share of the total social surplus. The intuition for this result relies on the well-known fact that the less heterogenous the consumers are in their reservation prices for a product, the higher the proportion of surplus that sellers can extract using their marketing instruments (Adams and Yellen 1979). Since the total consumer transaction costs in our framework are distributed uniformly in the interval $[0, t]$, a reduction in t reduces the heterogeneity in total consumer transaction costs and hence makes consumers more homogenous in their "effective" reservation prices for the product (as also evidenced in Figure 2 by a flattening of the line $\underline{\theta}(x)$ representing marginal consumers). This enables the seller to extract more surplus from consumers.

We summarize all the findings of this Section in the following Proposition.

Proposition 1 (burden of quality production on seller's variable cost). *A market-level advance that reduces consumer transaction costs:*

i) reduces the surplus of some consumers (specifically, those with low transaction costs and high valuations of quality);

ii) reduces the consumer share of the total social surplus;

iii) increases the seller's profit; and

iv) lowers the consumer price.

2.2 Case where burden of quality production falls on seller's fixed costs

Here the seller's profit is:

$$\pi = pD(p, q, t) - q^2,$$

where the expression for $D(p, q, t)$ is given in (2) and where we assume a quadratic quality-production function. Since the analysis follows the same steps as in the case where quality production affects seller's variable costs, we relegate the analysis to the Appendix and present the results here.

Proposition 2 (burden of quality production on seller's fixed cost). *A market-level advance that reduces consumer transaction costs:*

i) reduces the surplus of some consumers (specifically, those with low transaction costs and low valuations of quality);

ii) reduces the consumer share of the total social surplus;

iii) increases the seller's profit; and

iv) raises the consumer price.

The results in the fixed-cost case differ from those in the variable-cost case on the following two counts: first, the seller here finds it optimal to *increase* the consumer price, and second, it is now the consumers with *low* marginal valuations of quality (and low transaction costs) who become worse off.

Intuitively, a reduction in consumer transaction cost in the fixed-cost case leads to an increase in consumer price because it increases consumers' effective reservation price and

causes the seller to increase its product quality – both outcomes favoring a price increase. And the intuition for why the seller increases its product quality is that a reduction in market-level parameter t (i) increases the seller's marginal revenue – all consumers are now richer and able to pay more for any given quality, (ii) *does not* impact its marginal cost – although more consumers can now afford to buy the seller's product, this incremental sale does not mean any additional costs for the seller (quality production being a fixed cost here).

The intuition for why it is the consumers with low marginal valuations of quality (and low transaction costs) who become worse off here is that consumers with low marginal valuation of quality (θ_i) are the ones who have the most to lose from the increase in product quality initiated by the seller.

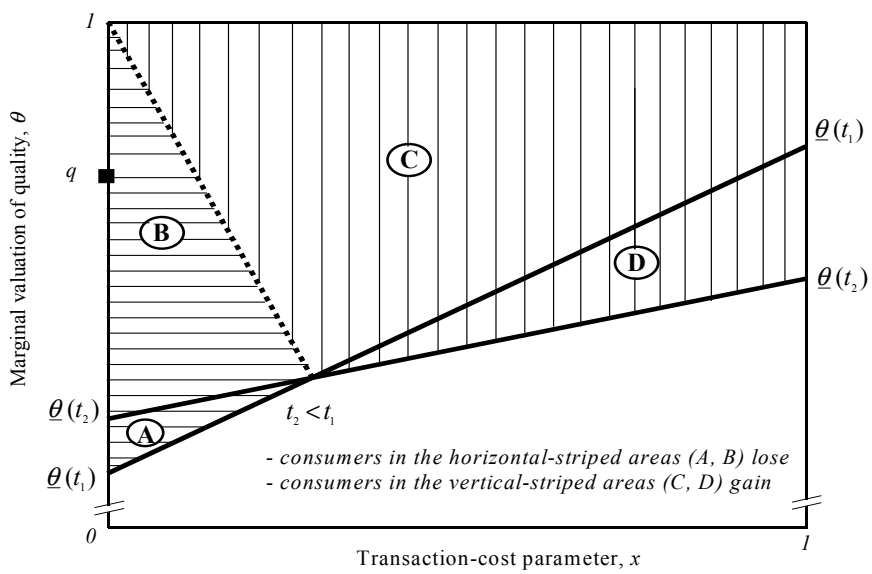


Figure 3: *Effects of reduced consumer transaction costs on consumer surplus (burden of quality production on seller's fixed cost).*

■ **Summary.** We have shown that a market-level advance that reduces consumer transaction costs:

- increases the seller's profit, lowers the surplus of some consumers, and reduces the consumer share of the total social surplus — *irrespective* of whether the burden of quality production falls on the seller's variable costs or fixed costs.

- lowers or raises consumer price — depending on whether the burden of quality production falls on the seller's variable costs or fixed costs. Specifically, if the burden of quality production falls on the seller's fixed costs, then the seller increases product quality which, together with increased consumer reservation price, leads to increased consumer price. On the other hand, if the burden of quality production falls on the seller's variable costs, then the seller decreases product quality which, together with increased consumer reservation price, can lead to either an increase or a decrease in consumer price (in our setup, the net effect is a decrease in consumer price).

3 Seller's Incentive to Invest in Reducing Consumer Transaction Costs

The analyses in the previous Sections considered the case in which consumer transaction costs reduce due to *exogenous* market-level advances. However sellers in most markets have the ability to invest in lowering consumer transaction costs. For example, a seller can make the use of its product easier (e.g., through toll-free helpline, on-line help features in software packages), make transportation easier (e.g., through cheaper or free home deliveries), make its product more customizable, increase the scope of use or general-purposeness of its product,

and so forth. In this Section, we allow for such an investment by the seller and examine whether the seller may have an incentive to reduce consumer transaction costs by more than what consumers, in aggregate, prefer.

Intuitively, from the consumer perspective, while the advantage of a reduction in technological parameter t is the lowering of their transaction costs, the disadvantage is that they may become more homogenized in their effective reservation prices. From the seller's perspective, a reduction in technological parameter t is beneficial because it can (i) increase consumers' effective reservation price, (ii) bring in new consumers to its market, and (iii) homogenize consumers' effective reservation prices. Therefore, depending on the cost the seller faces in reducing technological parameter t, it is possible that the seller finds it optimal to reduce consumer transaction costs by more than what consumers, in aggregate, prefer.

To illustrate this trade-off formally, we modify our model by allowing the seller to invest in reducing market-level technological parameter t to αt, with $\alpha \in [0, 1]$. The consumer utility function can then be rewritten as $q\theta_i - p - (\alpha t)x_i$, and the seller's profit function (in the case where the burden of quality production falls on variable costs)[12] becomes

$$\pi = (p - q^2 - c(t, \alpha))D(.),$$

where $D(.)$ is the consumer demand and $c(t, \alpha)$ is the variable cost to the seller of reducing market-level technological parameter t to αt. As mentioned earlier, this cost could represent the cost of adding "ease-of-use" features in a product, modifying the product to make it more transportable, and offering more toll-free minutes for technical help, etc. Choosing $c(t, \alpha) = (1 - \alpha)^2 t$ as an example, we show in the Appendix and in Figure 4 the following

[12] Similar results hold when the cost of producing quality is a fixed cost.

result.

Proposition 3. *A seller may invest and reduce consumer transaction costs by more than what consumers, in aggregate, prefer. In other words, the total consumer surplus may reduce with a reduction in their transaction costs caused by seller's investment.*

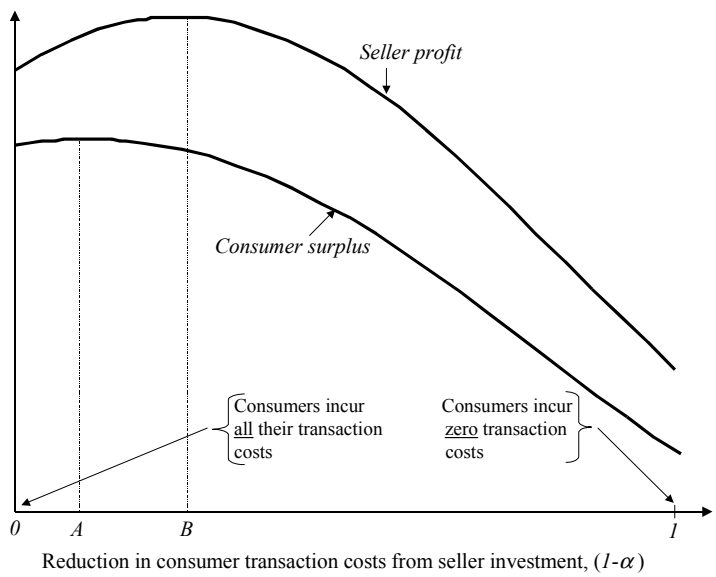


Figure 4: *Effects of seller investment in reducing consumer transaction costs.*

One can see from Figure 4 that as the consumer transaction costs decrease, the consumer surplus increases initially (in the region 0-A) but then decreases later (in the region A-1) — implying that consumers, in aggregate, may be worse-off when they incur lower transaction costs.

Intuitively, when the seller invests to reduce market-level technological parameter t, while consumers do gain from a reduction in their transaction costs, they also face the following two adverse effects: first, consumers become more homogenous in their transaction costs and hence in their effective reservation prices (in other words, in Figures 2 and 3, the line $\underline{\theta}(x)$

representing marginal consumers becomes flatter), allowing the seller to extract consumer surplus better; second, the seller passes through some of its cost of reducing consumer transaction costs to the consumers. These two adverse effects imply that consumer surplus may be higher when they incur higher transaction costs.

Overall, the analysis in this Section highlights some mechanisms through which seemingly pro-consumer firm practices that lower consumer transaction costs may also lower total consumer surplus.

4 Concluding Remarks

This paper used a simple model to show the following four main results. First, some consumers can become worse-off with a reduction in their transaction cost if this reduction arises from a market-level technological advance. Second, such market-level technological advances may homogenize consumers' effective reservation prices, leading to a reduction in the consumer share of the total social surplus. Third, when a seller can invest to reduce consumer transaction costs, consumers may be better off with higher transaction costs. And finally, whether consumers face higher or lower prices with a reduction in their transaction costs may partly depend on the nature of quality-production process in that market.

The paper discussed how these results are intuitively driven by the following three market characteristics: (i) market-level technological advances reduce individual consumers' transaction costs by different amounts; (ii) consumers are heterogenous in their preferences for product features; and (iii) fixed costs or increasing returns to scale prevent sellers from offering each consumer in the market its "perfect" product.

To make transparent the connection between these market characteristics and the key results, we kept the model simple. One can add more features to the model to make it more

"realistic", but that would only generate other effects to *complement* the effects generated by our model. For example, we derived the results controlling for the number of firms and the number of products; if a reduction in consumer transaction cost induces entry of new firms and/or new products, then that effect should be taken into account to complement the results of this paper.[13] Overall, we see the contribution of this paper as identifying some mechanisms through which reduction in consumer transaction costs – caused by either market-level technological advances or firm actions – can affect consumers adversely.

[13] The theoretical and empirical effects of entry on prices and consumer welfare are already well-understood in the existing literature (e.g., see Carlton and Perloff 1994).

APPENDIX

■ *Proof that the seller increases product quality and product price in response to a reduction in market-level parameter t, if the burden of quality production falls on its fixed costs.*

Here the seller's profit is:

$$\pi = pD(p, q, t) - q^2,$$

where the expression for $D(p, q, t)$ is given in (2) and where we assume a quadratic quality-production function. Solving the seller's profit-maximizing first-order condition w.r.t. its price p, we get:[14]

$$\widehat{p}(q, t) = -\frac{t}{4} + \frac{q(t)}{2}.$$

Thus, as before, a reduction in consumer transaction costs (via reduction in market-level parameter t) affects the seller's pricing policy directly by increasing the consumers' effective reservation prices (*reservation-price effect*), and indirectly through its effect on seller's quality choice (*quality-effect*).

Inserting the above value of $\widehat{p}(q, t)$ in the seller's profit function, we get

$$\pi = \left[\frac{(2q - t)^2}{16q}\right] - \left[q^2\right]$$

where the expression in the first square bracket is the seller's revenue and that in the second square bracket is the seller's cost. The choice of optimal quality q is determined by the first-order condition $\partial\pi/\partial q = 0$, or:

$$\underbrace{\left[\frac{1}{4} - \frac{t^2}{16q^2}\right]}_{\text{marginal revenue}} = \underbrace{[2q]}_{\text{marginal cost}}.$$

One can then see that while a reduction in market-level parameter t increases the seller's marginal revenue, it *does not* impact its marginal cost – implying that the seller's optimal reaction is to increase its product quality. This increase in product quality also implies that a reduction in consumer transaction cost increases the consumer price *unambiguously* in the fixed-cost case.

■ *Deriving the expressions used to draw Figure 4 (i.e., the case in which the seller can invest to reduce market-level technological parameter, t).*

The consumer utility function is:

$$U_i = q\theta_i - p - \alpha t x_i.$$

[14] $q > t/2$ is a required condition for positive consumer price, $\widehat{p}(q, t)$.

The consumer with transaction-cost parameter x that is indifferent between buying and not buying has $\theta = \underline{\theta}(x)$ where $U(\underline{\theta}) = 0$, or

$$\underline{\theta}(x) = \frac{p}{q} + \frac{\alpha t}{q} x.$$

Thus, the seller's demand from consumers with transaction-cost parameter x is $D(x) = 1 - \underline{\theta}(x)$ and its total demand is

$$D(p, q, \alpha; t) = \int_0^1 D(x)dx = 1 - \frac{p}{q} - \frac{\alpha t}{2q}.$$

Therefore, the seller's profit-function is:

$$\pi = \left[p - q^2 - (1-\alpha)^2 t\right] D(p, q, \alpha; t), \tag{A.1}$$

where the burden of quality production falls on the seller's variable cost ($= q^2$), and the burden of reduction in market-level technological parameter also falls on the seller's variable cost ($= (1-\alpha)^2 t$). Solving the first-order profit-maximizing conditions, we get the following expressions for optimal price and quality:

$$p^*(\alpha, t) = \frac{1}{18}\left[3\left(4 - 9\alpha + 4\alpha^2\right)t + 2(1 + Z(\alpha, t))\right],$$
$$q^*(\alpha, t) = \frac{1}{6}\left[1 + Z(\alpha, t)\right].$$

where $Z(\alpha, t) = \sqrt{1 + 6\left(2 - 3\alpha + 2\alpha^2\right)t}$. Inserting these values in the seller's profit function (given in (A.1)) and in the consumer surplus function ($CS = \int_0^1 \int_{\underline{\theta}}^1 (\theta q - p - \alpha t x) d\theta dx$), we get

$$\pi^*(\alpha, t) = \frac{\left[1 - 6(2 - 3\alpha + 2\alpha^2)t + Z(\alpha, t)\right]^2}{54\left[1 + Z(\alpha, t)\right]}, \text{ and}$$
$$CS^*(\alpha, t) = \frac{9(16 - 48\alpha + 71\alpha^2 - 48\alpha^3 + 16\alpha^4)t^2 + 2\left[1 + Z(\alpha, t)\right] - 6t(2 - 3\alpha + 2\alpha^2)\left[1 + 2Z(\alpha, t)\right]}{108\left[1 + Z(\alpha, t)\right]},$$

which are used to draw Figure 4.

References

Adams, W.J., and J.L. Yellen (1976), "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, August, 475-498.

Bailey, J. P. (1998), "Electronic Commerce: Prices and Consumer Issues for Three Products: Books, Compact Discs, and Software," working paper, R.H. Smith School of Business, University of Maryland.

Bell, D.R., T. Ho, and C.S. Tang (1998), "Determining Where to Shop: Fixed and Variable Costs of Shopping," *Journal of Marketing Research*, 35(3), 352-369.

Bellante, D., and A.C. Foster, "Working Wives and Expenditure on Services," *Journal of Consumer Research*, 11(2), 700-707.

Betancourt, R., and D. Gautschi (1990), "Demand Complementarities, Household Production, and Retail Assortments," *Marketing Science*, 9(2), 146-161.

Brynjolfsson, E., and M.D. Smith (2000), "Frictionless commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 46(4), 563-585.

Carlton D.W., and J.M. Perloff (1994), *Modern Industrial Organization* (second edition), Harper Collins, New York.

Crafton, S.M. (1979), "Convenience Store Pricing and the Value of Time: A Note on the Becker-Devany Full Price Model," *Southern Economic Journal*, 45(4), 1254-1260.

Currim, I.S., C.B. Weinberg, and D.R. Wittink (1981), "Design of Subscription Programs for a Performing Arts Series," *Journal of Consumer Research*, 8(2), 67-75.

D'Aspremont C.J., J. Gabszewicz and J. Thisse (1979), "On Hotelling's Stability in Competition," *Econometrica*, 47, 1145-1150.

Forrester Research, Inc. (1998), *Transactive Evolution*, research report, Forrester Research.

Greenfield On-line Survey (2000), in *Internet World,* January, 5-7.

Kim, B-D., and K. Park (1997), "Studying Patterns of Consumer's Grocery Shopping Trip," *Journal of Retailing*, 73(4), 501-517.

Lal, R., and M. Sarvary (1999), "When and How is the Internet Likely to Decrease Price Competition?" *Marketing Science*, 18(4), 485-503.

Lee, H.G. (1998), "Do Electronic Marketplaces Lower the Price of Goods," *Communications of the ACM*, 41(1), 73-80.

Messinger, P. R., and C. Narasimhan (1997), "A Model of Retail Formats Based on Consumers' Economizing on Shopping Time," *Marketing Science*, 16(1), 1-23.

Ratchford, B.T., and G.T. Stoops (1992), "An Econometric Model of a Retail Firm," *Managerial and Decision Economics*, 13 (May-June), 223-231.

Neven, D., and J.F. Thisse (1990), "On Quality and Variety Competition," in J.J. Gabszewicz, J.F. Richard, and L.A. Wolsey (Eds.), *Economic Decision Making: Games, Econometrics and Optimization*, Amsterdam: North-Holland, 175-199.

Spence, M. (1976), "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies*, 43(2), 217-35.

Venables, A.J. (1999), "Road Transport Improvements and Network Congestion," *Journal of Transport Economics & Policy*, 33(3), 319-328.