System Design, User Cost and Electronic Usage of Journals

Robert S. Gazzale * Jeffrey K. MacKie-Mason †

August 17, 2000

Abstract

Dramatic increases in the capabilities and decreases in the costs of computers and communication networks have fomented revolutionary thoughts in the scholarly publishing community. In one dimension, traditional pricing schemes and product packages are being modified or replaced. We designed and undertook a large-scale field experiment in pricing and bundling for electronic access to scholarly journals: PEAK. We provided Internet-based delivery of content from 1200 Elsevier Science journals to users at multiple campuses and commercial facilities. Our primary research objective was to generate rich empirical evidence on user behavior when faced with various bundling schemes and price structures. In this article we explain the different types and levels of cost that users faced when accessing individual articles, and report on the effect of these costs on usage. We found that both monetary and non-monetary user costs have a significant impact on the demand for electronic access. We also estimate how taking user costs into account would change the "optimal" (least cost) bundle of access options that an institution should purchase.

*Robert S. Gazzale is a Ph.D. student in the Department of Economics, University of Michigan. Mr. Gazzale is a research assistant on the PEAK project.

†Jeffrey K. MacKie-Mason is a Professor of Economics, Information and Public Policy at the University of Michigan, Ann Arbor, Michigan, USA. Professor MacKie-Mason is the Research Director of the PEAK project.

1 Introduction

Electronic access to scholarly journals has become an important and commonly accepted tool for researchers. The user community has become more familiar with the medium over time and has started to actively bid for alternative forms of access. Technological improvements in the communication networks paired with the decreasing costs of hardware create greater incentives for innovation. Consequently, although publishers and libraries face a number of challenges, they also have promising new opportunities.[1] Publishers are creating many new electronic-only journals on the Internet, while also developing and deploying electronic access to literature traditionally distributed on paper. They are modifying traditional pricing schemes and content bundles, and creating new schemes to take advantage of the characteristics of digital duplication and distribution.

The University of Michigan operated a field trial in electronic access pricing and bundling called "Pricing Electronic Access to Knowledge" (PEAK). We provided a host service consisting of roughly four and a half years of content (January 1995 – August 1999) of all approximately 1200 Elsevier Science scholarly journals. Participating institutions had access to this content for over 18 months. Michigan provided Internet-based delivery to over 340,000 authorized users at twelve campuses and commercial research facilities across the U.S. The full content of the 1200 journals was received, catalogued and indexed, and delivered in real time. At the end of the project the database contained 849,371 articles, and of these 111,983 had been accessed at least once. Over $500,000 in electronic commerce was transacted during the experiment.

We have elsewhere described the design and goals of the PEAK research project (MacKie-Mason and Riveros (2000) [1]. In MacKie-Mason, Riveros and Gazzale (2000) [2], we detailed the pricing schemes offered to institutions and individual users. We also reported and analyzed usage statistics, including some data on the economic response of institutions and individuals to the different price and access options.

In this paper, we focus on an important economic behavior question: how much does usage respond to various differences in user cost? We pay careful attention to the effect of both pecuniary costs and non-pecuniary costs such

[1] See MacKie-Mason and Riveros (1999) [1] for a discussion of the economics of electronic publishing.

as time and inconvenience.

It is well known that the usage of information resources responds to the monetary cost users bear. We find that even modest per article fess drastically suppressed usage. It is also true, but perhaps less appreciated, that non-pecuniary costs are important for the design of digital information access systems. We find that the number of screens users must navigate, and the amount of external information they must recall and provide (such as passwords), has a substantial impact on usage. These non-pecuniary costs are also important for pricing and the funding of digital library collections: the higher the non-pecuniary costs, the less users will be willing to pay for access.

An interesting aspect of the PEAK project, and one that is applicable to a wide variety of economic problems, is the two-layered decision problem on the consumer side. At the first level, institutions make access product purchasing decisions. For example, first an institution must decide whether or not to purchase subscriptions to journals. If it is decided that subscriptions should be purchased, the institution must then decide how many subscriptions to purchase as well as which ones. These decisions then have a potentially large effect on the costs that users face to access electronic journal articles, whether it be the requirement that users obtain and use a password or pay a monetary cost. The consumer then decides whether she will pay these costs to access a given article.

Different users are going to have different valuations for electronic access to journal articles. Furthermore, even the same user will not value all requested articles the same. Standard economic theory tells us that a user will access that article if the marginal benefit that the user obtains from the article (i.e. the value) is greater than the marginal cost. Information regarding the users "price" elasticity is thus of import to the institutional decision-maker in attempting to maximize user welfare subject to a budget constraint that is increasingly binding. It is therefore vital that institutions have as much information as possible about the degree to which increased marginal costs decrease the demand for electronic access. Similarly, demand elasticity information is vital to firms designing pricing schemes as system design decisions will affect non-pecuniary costs faced by the users, and thus overall demand for access.

Our analysis of the PEAK data sheds some light on users' elasticity of demand. We find that the imposition of non-pecuniary user costs, such as password use and the need for take action to have monetary costs subsidized,

causes a severe reduction in the demand for electronic access to journal articles. Further, we find some preliminary evidence that users are more likely to bear these costs when they are expected. Further analysis on this is warranted. We estimate the amount of demand that is choked-off with successive increases in the marginal cost of access. Finally, we adjust our estimate of an institutions "optimal" (i.e. lowest cost) bundle to account for the fact that the imposition of user costs for access biases actual demand towards lower cost access.

2 Access Models Offered

Participating institutions in the PEAK experiment were offered packages containing two or more of the following three access products:

1. *Traditional Subscription:* Unlimited access to the material available in the corresponding print journal.

2. *Generalized Subscription:* Unlimited access to any 120 articles from the entire database of priced content, typically the two most current years. Articles are selected on demand, after they are published, as users request articles that are not otherwise paid for, until the subscription is exhausted.[2] Articles selected for generalized subscriptions may be accessed by all authorized users at that institution.

3. *Per Article:* Unlimited access for a single individual to a specific article. If an article is not available in a subscribed journal, nor a generalized subscription, nor are there unused generalized subscription tokens, then an individual may purchase access to the article.

The per article and generalized subscription options allow users to capture value from the entire corpus of articles, without having to subscribe to all journal titles. Once the content is created and added to the server database, the incremental cost of delivery is approximately zero. Therefore, to create maximal value from the content, it is important that as many users as possible have access. The design of the pricing and bundling schemes affect both how

[2] 120 is the approximate average number of articles in a traditional printed journal for a given year. We refer to this bundle of option to access articles as a set of tokens, with one token used for each article added to the generalized subscription during the year.

Institution ID	Group	Traditional	Generalized	Per Article
5, 6, 7, 8	Green		X	X
3, 9, 10, 11, 12	Red	X	X	X
13, 14, 15	Blue	X		X

Table 1: Access Models

much value is delivered from the content (the number of readers) and how that value is shared between the users and the publisher.

Generalized subscriptions may be thought of as a way to pre-pay (at a discount) for interlibrary loan requests. One advantage of generalized subscription purchases is that the "tokens" cost substantially less per article than to per article license price. Institutions do, however, need to purchase tokens at the beginning of a year and thus bear some risk. There is an additional benefit: unlike an interlibrary loan, all users in the community have ongoing unlimited access to the articles obtained via generalized subscription token. To the publisher, generalized subscriptions represent a committed flow of revenue at the beginning of each year, and thus shift some of the risk to the users. Another benefit to the publisher is that that they open up access to the entire body of content to all users, and by thus increasing user value from the content, provide and opportunity to obtain greater returns from the publication of that content.

Participating institutions were assigned randomly to one of three different experimental treatments, which we labeled as the Red, Green and Blue groups. Institutions in every group could purchase articles on a per article basis. Those is the Green group could purchase generalized subscriptions, while those in the Blue group could purchase traditional subscriptions. Institutions in the Red group could purchase all types of access. Twelve institutions participated in PEAK: large research universities, medium and small colleges and professional schools, and corporate libraries. Table 2 shows the distribution of access models and products offered to the participating institutions.

3 Summary of User Costs

The PEAK experiment was designed to assess response to various pricing and access schemes for digital collections. Since the content was traditional refereed scholarly literature, the experiment was designed in the context of

the larger environment and institutions that have developed for scholarly communication. In particular, PEAK implemented access through the traditional intermediary: the research library.

The role of the research library affected the design of the experiment and thus the research questions we could investigate. As previously noted, the research librarian, by means of the combination of access products he selects, determines the costs faced by individual users. The individual users then make article-level access decisions.

When confronted with the PEAK access options and prices, nearly all of the participating libraries purchased substantial access on behalf of their users. Once consequence of these institution-level choices is that there was little incentive for most individuals to purchase additional access. Although we measured over 200,000 unique individual uses of the system, we observe relatively few individual decisions to pay for access not previously purchased by the libraries. As a consequence, it is important that we try to learn from user response to non-pecuniary as well as pecuniary costs.

Substantial amounts of PEAK content were available at zero user cost, merely by an IP authentication that was automatically performed. That is, when a user requested content of this type, it is displayed if her workstation's IP address is from an authorized institution with no additional hindrance to the user.[3] Content available at zero user cost included:

- all "unmetered" content, which included articles published at least two calendar years prior as well as all non-full-length articles;
- articles in journals to which the institution purchased an electronic Traditional Subscription; and
- articles which had previously been purchased by a user at the institution with a Generalized Subscription token.

The next level of user cost was the need to enter a password. The transactions cost of password entry would have ranged from small to substantial. In the worst case, the user needed to navigate elsewhere in the system to fill out a form requesting a password, and then wait to receive it. Once received, the user had to enter the password, If the user already had a password, then

[3]This is also true if a user accessed PEAK from an IP address not associated with her institution. The only difference is that she would have had to already have entered a password to access PEAK.

the only cost to him is to find or recall his password and enter it. Content accessible via password entry included:

- articles purchased via Generalized Subscription (i.e. articles in journals to which the institution does not have a Traditional Subscription, assuming not all Generalized Tokens have not been used);
- subsequent access to an article which an individual has already purchased on a per-article basis.

If the institution did not have any unused Generalized Subscription tokens, then content not available at zero cost could be access by payment of a $7 per-article fee. The user who wished to pay the per-article fee would also bear two non-pecuniary costs: (1) password recall and entry, as above for the use of a Generalized subscription token, and (2) credit card recall and entry. [4]

Institutions 13 and 14 were exceptions to many of these rules. At both, per-article access for all requests was paid (invisibly by the user) by the institution, so users never faced a pecuniary cost. At institution 14, a user still faced the non-pecuniary cost of finding her password and entering it for appropriate "paid"[5] content. However, all users at institution 13 accessing from associated IP addresses were considered password authenticated. Thus users at institution 13 could access all PEAK content at zero total (pecuniary *and* non-pecuniary) cost. We use these differenced in user cost between these two institutions and the others in the analysis below.

4 Effects of User Cost on Access

As noted above, costs required to access PEAK content depended on a variety of factors. One factor is the type of content requested ("metered" versus "unmetered"). Looking only at metered content, the costs associated with access depended in large part on the access products purchased by a user's institution. Further, even looking at costs faced by users associated with

[4] In the first eight months of the experiment, users paid with a First Virtual VPIN account, rather than with a credit card. Because a VPIN was an unfamiliar product, the non-pecuniary costs were probably higher than for credit card usage, although formally the user needed to undertake the same steps.

[5] Paid content is metered content not including article in journals to which an institution purchased a Traditional Subscription.

the same institution, access costs depended not only on the specific products selected by an institution (i.e. the specific journals to which an institution holds a Traditional Subscription; the number of Generalized Tokens purchased), but also on the actions of other users at the institution (whether a token was previously used to purchase a requested article; how many tokens are remaining). In the following sections, we estimate the effects of these marginal costs on the quantity and composition of metered access.

4.1 Effects of User Cost on Access by Experimental Group

To test the impact of user cost on usage on aggregate institutional access, we compared the access patterns of institutions in the Red group with those in the Blue group. Red institutions had both generalized and traditional subscriptions available; Blue had only traditional. In particular, we looked at the number of "paid" accesses to individual articles (paid by generalized tokens or per article fee) per 100 unmetered accesses, normalized to account for the number of traditional subscriptions.[6]

Table 2 presents our statistic for relative demand for paid access: calculated Normalized Paid Access per 100 Unmetered Accesses. One thing that is clear that demand differed among institutions with the same access products even after controlling for the size of an institution's subscription base. This suggests that there are institution-specific attributes affecting demand for paid access. It is also possible that our controlling for subscription size is not complete. Users at an institution with a large Traditional Subscription base, such as institution 3, have a lower expected marginal cost for access than users at institutions with a smaller number of Traditional Subscriptions.

[6]Our statistic, which we shall refer to as Normalized Paid Access, is equal to $\frac{A_{paid}}{A_{free}} \cdot (Scale)$, where A_{paid} is the total number of paid access, A_{free} the total number of unmetered accesses, and *Scale* is equal total number of free accesses divided is the total number of accesses to free content in journals to which the instittion does not have a Traditional Subscription. We multiply by Scale due to the fact that the more that accesses that are covered by Traditional Subscriptions, the less likely a user is to require paid access. To control for different overall usage intensity (due to different numbers of active users, differences in the composition of users, differences in research orientation, differences in user education about PEAK, etc.) we scaled by accesses to unmetered content. We thus use unmetered accesses as a proxy for the number of user sessions, and therefore our statistic is an measurement of paid accesses per session.

Institution	Group	Normalized Paid Accesses Per 100 Unmetered Accesses
3	Red	13.50
9	Red	20.36
10	Red	31.69
11	Red	7.59
12	Red	26.44
Avg.	Red	15.14
13	Blue	50.97
14	Blue	15.06
15	Blue	4.72

Table 2: Normalized Paid Access per 100 Unmetered Accesses, by Institution

Thus the difference between expected marginal cost and actual marginal cost may be important. We shall return to this point later.

We can make some interesting comparisons between the Red group and the institutions in the Blue group. While institution number 13, as a member of the Blue group, only had traditional subscriptions and per article access available, users at this institution did not need to authenticate for any content, and thus faced no marginal cost in accessing paid content. Most users at Red institutions faced the cost of authenticating to spend a token.[7] We would therefore expect a higher rate of paid access at institution 13, and this is in fact the case.

Paid access at institution 14 was similarly subsidized by the institution. However, in contrast to institution 13, authentication was required. As the marginal cost of paid access for institution 14 is exactly the same as those institutions in the Red group, we would therefore expect that their demand for paid access would be similar. This is in fact the case. Finally, per article access for users at institution 15 was not automatically subsidized. Thus, users faced very high marginal costs for paid content.[8] In addition to the $7.00 per article fee, users needed to authenticate with a password and enter their credit card information. We would therefore expect the rate of paid

[7] Only 28% of unmetered accesses from Red group users were password authenticated. This suggests that a large majority of users attempting to access paid content would not already be password authenticated. For these users, the need to password authenticate would truly be a marginal cost.

[8] In addition to the user costs, if the institution does not reimburse individual purchases of articles, users will also face a significant monetary cost. We do not have complete information about the reimbursement process at institution 15, but we can assert that their users had at least higher transaction costs than the other institutions in the sample.

access to be lower than that of the Red group, as was the case.

	No month dummies	Month dummies
Dependent variable is weekly normalized paid access per 100 free access		
Constant	87.535*	108.615*
	10.394	14.643
Blue: Credit Card (Inst. 15)	-280.490*	-270.879*
	(37.627)	(35.508)
Red + Inst.14	-58.999*	-57.764*
	(7.900)	(7.186)
Out of Tokens	-25.070*	-25.665*
	(1.635)	(2.533)
Graduate Students/Faculty Ratio	43.821*	41.748*
	(7.301)	(6.912)
Percentage Engineering, Science and Medicine	-225.913*	-215.767*
	(7.535)	(36.553)
Sample Size	530	530
R^2	0.171	0..229

Standard errors are shown in parenthesis.
* Significant at the 99% level, ** Significant at the 95% level; *** Significant at the 90% level.

Table 3: Estimation Results on Effects of User Cost on Access

Table 3 summarizes the results from the estimation of the effects of user cost on access. We controlled for differences in the graduate students/faculty ratio and percentage of users in Engineering, Science and Medicine. The dependent variable, Paid Accesses per 100 Unmetered Accesses, controls for learning and seasonality effects. We thus see the extent to which paid access, starting from a baseline of access to paid content without marginal cost, falls as we increase marginal costs. Imposition of a password requirement reduces paid accesses by almost 60 accesses per 100 unmetered accesses, while the depletion of tokens at these institutions results in a further reduction of approximately 25 access.[9]

We gain further evidence of the extent to which the marginal costs associated with paid access throttles demand for metered access by comparing the composition of unmetered access with the composition of metered access. We calculated the ratio of free accesses to articles inside the institutions Traditional Subscription base to those outside. We would expect that, absent any additional user costs, this ratio would be the same for metered content.

[9]At certain of these institutions, the $7.00 was paid by the institution's inter-library loan department. In these cases, the additional marginal cost was not monetary, but rather in terms of time and effort.

Based on this ratio and the number of articles accessed under an institution's Traditional Subscriptions, we can predict the number of paid accesses that would have occurred if there were no additional marginal costs for paid access.

Institution	Year	Actual Per Predicted	Percent Free Access Psswd. Authent.	Credit Card Required	Password Entered When Prompted
3	1998	21.14%	11.10%	0	6.69%
10	1998	146.15%	45.35%	0	13.54%
11	1998	16.39%	8.81%	0	2.64%
12	1998	83.33%	51.71%	0	7.14%
13	1998	125.93%	98.75%	0	100.00%
14	1998	79.25%	54.50%	0	44.44%
15	1998	0.00%	22.19%	1	8.06%
3	1999	31.38%	19.06%	0	10.44%
10	1999	123.40%	43.88%	0	13.43%
11	1999	20.82%	18.50%	0	14.12%
13	1999	77.67%	100.00%	0	100.00%
14	1999	56.69%	63.22%	0	17.78%
15	1999	19.52%	12.21%	1	2.39%

Table 4: Paid Access as Percentage of Predicated Based on Free Access Composition

Table 4 presents actual paid access as a percentage of predicted for all institutions that have Traditional Subscriptions in a given year. The only true outlier is institution 10. We suspect that this might be partially due to the fact that they had the fewest Traditional Subscriptions, thus users expected to have to pay for metered content with password entry. As for institution 13, recall that users faced no marginal cost to access paid materials. We thus expect their paid access to be close to predicted. For all other institutions we generally see that the costs associated with paid access caused an appreciable reduction in the number of paid articles demanded. We also present in Table 4 factors which we believe help explain this shortfall, namely the percentage of free access that is password authenticated, whether or not a credit card is required for all paid access, and the rate at which passwords were entered for paid access when prompted.

Table 5 summarizes the results from the estimation of the effects of user cost on actual paid access as a percentage of prediction. Despite the small sample size, the results clearly demonstrate that as we increase the number of individuals who can access paid content without additional marginal costs (proxied by the percent of free access that is password authenticated), more paid access is demanded. The credit card required dummy is not significant,

Dependent variable is Actual paid access as a percentage of predicted.	
Percent Free Psswd. Auth.	2.122*
	(.446)
Prompted Login Percent	-1.051**
	(.544)
Credit Card Required	-.213
	(.249)
Sample Size	13
R^2	0.848

Standard errors are shown in parenthesis.
* Significant at the 99% level; ** Significant at the 95% level;
*** Significant at the 90% level

Table 5: Estimation Results on Effects of User Cost on Actual Paid as Percent of Predicted

although it must be noted that there are only two observations where credit cards are required, and both are the same institution. The coefficient for the percent of prompted users who login is of the wrong sign.

4.2 Effects of User Cost on Access: Per Article Fee

If an institution did not have any tokens, either through depletion or unavailability, a user wanting to view a paid article not previously accessed would have had 3 choices.[10] First, she could pay $7.00 in order to view the article, and also incur the non-pecuniary cost of entering credit card information and waiting for verification. If the institution subscribed to the print journal, she could substitute the print journal article for the electronic product. She could also request the article through a traditional interlibrary loan, which also involves higher non-price costs (from filling out the request form and waiting for the article to be delivered) than spending a token.[11]

Due to the system design, we are unable to determine the exact number of times that users were faced with the decision of whether or not to enter

[10]Recall that all users at an institution could access, without password authentication, any article previously purchased by that institution with a Generalized Token. For articles purchase on a Per Article basis, only the individual who purchased the article could view it without further monetary cost.

[11]The libraries at institutions 3 and 11 processed these requests electronically, through PEAK, while the library at institution 9 did not and thus incurred greater processing delays.

Institution	Estimated Credit Card Requests	Credit Card Payments	Percent
3	53	13	25.53%
6	260	194	74.62%
9	190	1	0.53%
11	562	61	10.85%
15	137	73	53.28%

Table 6: Actual Credit Card Payments as a Percent of Requests as Estimated from Transaction Logs

credit card information in order to access a requested article. We were able to identify in the transaction logs events consistent with the credit card decision. These events are, however, a noisy signal for the actual number of times users faced this decision.[12]

From this noisy signal, we were able to estimate the number of times credit card payment was requested. For every month, we determined the number of consistent events, and divided this number by the number of access requests handled by the system for that institution. For each institution that depleted its supply of tokens, we can thus calculate a weighted average level of noise per access request for months that they had tokens. We assume that anything over this average in months without tokens would represent requests for credit card payment. For institutions that never had tokens, we use as a baseline the weighted average of events per access request recorded by institutions with tokens.

In Table 6 we present the total number of payments as a percent of estimated requests for credit card payments. While the numbers of requests for credit card payments should be regarded as only an estimate, the relative percentages are consistent with our intuition. Institutions 6 and 15 never had any tokens. We would thus expect that users at these institutions had a relatively high expected cost of accessing articles and would thus pay fairly frequently.[13] Among the institutions where tokens were depleted, the payment rate is appreciably higher at institutions 3 and 11, which is consistent with the fact that interlibrary loan requests were handled through PEAK.

[12]This noisiness is demonstrated by the fact that the transaction logs has events consistent with a request for credit card for months in which institutions still had tokens.

[13]In addition, institution 6 is a corporate institution. It is possible that their users' budgetary constraints were not as binding as those associated with academic institutions.

Institution	Credit Card Requests	Credit Card Payments	Percent
3	128	13	10.16%
9	366	1	0.27%
11	1128	61	5.41%

Table 7: Actual Credit Card Payments as a Percent of Requests as Estimated from Token Expenditure Rate

We gain further understanding of the degree to which differences in user cost affects the demand for paid article access by looking at only those institutions that depleted their supply of tokens at various points throughout the project. There were three institutions in this category: institution 3 ran out of tokens in November 1998 and again in July 1999; institution 11 in May 1999; and institution 9 in June 1999.

For those institutions which had tokens available at certain times, we can estimate the number of credit card requests based on the number of tokens spent per free access. If we make the assumption that this rate of token expenditure would have remained constant were tokens still available, we can estimate the number of credit card requests to be equal to be equal to the estimated number of tokens that would have been spent were tokens available.

Table 7 presents the rate of credit card payments as estimated from the rate of token expenditure. The relative percentages are consistent with our previous estimates for these institutions. These estimates of requests for credit card payment suggest that our previous estimates were about 50% too low.

	Institution 3 1998	Institution 3 1999	Institution 9 1999	Institution 11 1999
30 days prior	13.56	18.43	20.2	16.03
30 days after	0.25	0.29	0.00	0.35
Percentage Decrease	-98.16%	-98.43%	-100.00%	-97.82%

Units: Normalized paid access per 100 unmetered accesses.

Table 8: Effect of Token Depletion on Demand for Paid Content

For each of the institutions that ran out of tokens, we present in Table 8 the normalized paid accesses per hundred free accesses for 60 days, the 30 days prior to running out of tokens and the 30 days after running out

of tokens. The results further demonstrate that when users are faced with increased user costs for electronic access, demand for these articles plummets.

User costs substantially diminish demand for paid content. We have further seen that not just price, but non-pecuniary costs such as password use have an appreciable effect on demand.

5 Effects of User Cost Expectation on Access

As demonstrated in Table 2, there is great variation in the demand for paid content, measured by Normalized Paid Access, even amongst institutions with similar user cost profiles. This statistic controls for number of Traditional Subscriptions, institution size, and seasonality. As previously noted, it is quite possible that we have not sufficiently controlled for the *effects* of the size of Traditional Subscription base. The larger this base, the more content that users at an institution can access at zero marginal cost. It is our hypothesis that this affects users' expectations. As users are habituated to receiving free content, they expect that they will receive the article they are requesting at zero marginal cost, and they are thus less likely to pay for content, in terms of entering a password when requested.

Institution	Normalized Paid Accesses Per 100 Unmetered	Percent of Unmetered in Subscription Base	Percent who Login when requested
3	13.50	83.62%	8.38%
10	31.69	6.94%	13.47%
11	7.59	74.21%	2.64 %
12	26.44	11.05%	7.14%
14	15.06	31.41%	29.63%

Correlation Coefficients:
Paid Access and Percent in Base: -0.87
Prompted Login and Percent in Base: -0.36

Table 9: Effect of Subscription Coverage on Paid Access

In Table 9 we present some evidence that users' expectations do matter. The institutions included are all of the institutions where password entry was required in order to spend a Generalized Subscription token, plus institution 14, where users faced similar costs. In the final column, we present the percent of users who login when prompted in order to spend a token. We use this as a proxy for willingness to pay. In the third column we present the percent of unmetered or free content that was selected from journals in an

institution's Traditional Subscription base. Assuming that there is no qualitative difference in the types of articles requested (metered vs. unmetered) from journals in or out of an institutions's subscription base, this should give a strong indication of the expected costs faced by users. Most of the articles requested by institution 3 are accessed at zero marginal cost, thus users do not expect to have to enter a password, whereas relatively few of the article requested by institution 10 are, thus users expect to have to enter a password to get the requested article.

It is our hypothesis that users who expect to get their requested article at no cost will be less likely to bear the non-pecuniary cost of password retrieval and entry than those whose expectation of getting the article for free is less strong.[14] The data appears to point in this direction. We see that the more that users get for free, the less their demand for free articles as summarized by Normalized Paid Access. Furthermore, we see a negative correlation, although not as strong, between getting articles at zero marginal cost and entering a password when requested.

6 Actual versus Optimal Choice

In determining to which scholarly print journals to subscribe, librarians are in an unenviable position.[15] They must determine which journals best match the needs and interests of their community subject to two important constraints. Their first constraint is budgetary. This constraint has become increasingly binding of late, as renewal costs have tended to rise faster than serial budgets [3]. The second constraint is that libraries have incomplete information in terms of community needs. At the heart of this problem is the fact that a traditional print subscription forces libraries to purchase publisher-selected bundles of information (the journal), while users are interested primarily in the articles therein. Users only read a small fraction of articles,[16] and the library generally lacks information about which articles

[14]Note that we are not addressing the question of whether expected costs effects a user's decision to access PEAK in the first place to search for an article.

[15]For an excellent discussion of the collection development officer's problem, see Haar [3]

[16]The percentage of articles read through June 1999 for academic institutions participating in PEAK ranged from .12% to 6.40%. An empirical study by King and Griffiths [4] found that about 43.6% of users who read a journal read five or fewer articles from the journal and 78% of the readers read 10 or fewer articles.

their community values. Further compounding their information problem is the fact that a library must make an ex ante (before publication) decision about the value of a bundle, while the actual value is realized ex post.

The electronic access products offered by PEAK enabled libraries to mitigate these constraints. First, users had access even to those articles included in the journals to which the institution does not subscribe. (At institutions which purchased traditional subscriptions, 37% of the most accessed articles in 1998 were outside the institution's traditional subscription base. This figure was 50% in 1999.) Second, the transaction logs that are feasible for electronic access allowed us to provide libraries with monthly reports not only on which journals their community valued, but also which articles. Detailed usage reporting should enable libraries to provide additional value to their communities. They can better allocate their serials budgets to the most valued journal titles or to other access products such as those offered by PEAK.

		Traditional		Generalized		Per Article	
Instid	Year	Actual	Optimal	Actual	Optimal	Actual	Optimal
3	1998	25,000	17,000	2,740	3,836	7	133
5	1998	N/A	0	15,344	6,576	0	169
6	1998	N/A	0	0	548	672	0
7	1998	N/A	0	24,660	12,604	0	0
8	1998	N/A	0	13,700	2,740	0	0
9	1998	0	556	13,700	6,576	0	56
10	1998	4,960	323	8,220	7,672	0	483
11	1998	70,056	5,217	2,192	13,700	0	84
12	1998	2,352	107	2,192	1,096	0	98
13	1998	28,504	139	N/A	0	952	1,120
14	1998	17,671	0	N/A	0	294	504
15	1998	18,476	0	N/A	0	0	1,176
3	1999	12,500	10,528	2,740	1096	84	0
5	1999	N/A	0	8,768	2,740	0	399
6	1999	N/A	0	0	548	686	0
7	1999	N/A	0	10,960	9864	0	511
8	1999	N/A	0	6,028	5480	0	462
9	1999	0	278	7,124	6,576	7	182
10	1999	2,480	1,401	8,768	6,576	0	210
11	1999	0	576	4,384	2,740	427	532
12	1999	0	0	1,644	548	0	539
13	1999	9,635	7,661	N/A	0	19964	7,175
14	1999	0	0	N/A	0	623	623
15	1999	8,992	1,058	N/A	0	511	1,694

Table 10: Actual versus. Optimal Expenditures on PEAK Access Products for 1998-1999

Instid	Year	Actual	Optimal	Savings	Percent
3	1998	27,747	20,969	6,778	24.43%
5	1998	15,344	6,745	8,599	56.04%
6	1998	672	548	124	18.45%
7	1998	24,660	12,604	12,056	48.89%
8	1998	13,700	2,740	10,960	80.00%
9	1998	13,700	7,188	6,512	47.53%
10	1998	13,180	8,478	4,701	35.67%
11	1998	72,248	19,001	53,247	73.70%
12	1998	4,544	1,301	3,243	71.37%
13	1998	29,456	1,259	28,197	95.73%
14	1998	17,965	504	17,461	97.19%
15	1998	18,476	1,176	17,300	93.63%
3	1999	15,324	11,624	3,699	24.14%
5	1999	8,708	3,139	8,708	63.96%
6	1999	686	548	138	20.12%
7	1999	10,960	10,375	585	5.34%
8	1999	6,028	5,942	86	1.43%
9	1999	7,131	7,036	94	1.33%
10	1999	11,247	8,187	3,060	27.21%
11	1999	4,559	3,848	711	15.60%
12	1999	1,644	1,087	557	33.88%
13	1999	29,599	14,836	14,763	49.88%
14	1999	623	623	0	0%
15	1999	9,502	2,751	6,751	71.04%

Table 11: Total Actual versus. Optimal Expenditures for 1998-1999

In order to estimate an upper bound on how much the libraries could benefit from better usage data, we analyzed each institution's accesses to the PEAK database in 1998 to determine what would have been their optimal bundle for the year if they had been able to *perfectly forecast* which articles would be accessed. We then calculated how much this bundle would have cost the institution, and compared this optimal cost with the institution's actual expenditures.

We present these results by access product in Table 10, and by total expenditures in Table 11. We found that actual expenditures were markedly higher than optimal purchases in 1998. In particular, institutions in the Red and Blue groups purchased far too many traditional subscriptions. Further, most institutions purchased too many generalized subscriptions. We believe that much of the over-budgeting can be explained by a few factors:

- First, institutions greatly overestimated demand for access, particularly with respect to journals for which they purchased traditional subscriptions. This difficulty in forecasting demands was compounded by delays some institutions faced in implementing the project and communicating with their users. In particular, none of the institutions in the Blue Group started the project until the third quarter of the year.

- Second, aspects of institutional behavior, such as "use it or lose it" budgeting and a preference for non-variable expenditures, might have factored into the decision making. A preference for non-variable expenditures, fixed in advance, would induce a library to rely more heavily on traditional and generalized subscriptions, and less on reimbursed individual article purchases or interlibrary loan.[17]

- Third, because they cost less per article, but allow ex post article selection just like per-article purchase, generalized subscriptions provide an insurance function to cover unanticipated demand. If libraries are risk-averse (in this case to the risk of large per-article purchases) they might be willing to pay an "insurance premium" to reduce the risk by buying more generalized subscription tokens than are expected to be used.

[17] With print publications and some electronic products libraries may be willing to spend more on full journal subscriptions to create complete archival collections. All access to PEAK materials ended in August 1999, however, so archival value should not have played a role in decision making.

	Traditional		Generalized	
Instid	Optimal Direction	Actual Direction	Optimal Direction	Actual Direction
3	-	=	+	+
5	N/A	N/A	-	-
6	N/A	N/A	+	=
7	N/A	N/A	-	-
8	N/A	N/A	-	-
9	+	=	-	-
10	-	=	-	+
11	-	-	+	+
12	-	-	-	+
13	-	=	N/A	N/A
14	-	=	N/A	N/A
15	-	+	N/A	N/A

Table 12: 1999 Expenditures: Actual Increase/Decrease versus. Predicted Optimal Cost

The PEAK project team provided the participating institutions with regular reports detailing user access to journals and articles. We hypothesized that librarian decisions about purchasing access products for 1999 might be consistent with a simple learning dynamic: increase expenditures on products they under-bought in 1998 and decreasing expenditures on products they over-bought in 1998. To see the extent to which institutions used this information in determining expenditures, we took for each institution the change in expenditure from 1998 to 1999 for each access product,[18] and compared this change with the change recommended by the learning dynamic. We present the results in Table 12.

Six of the nine institutions adjusted the number of generalized subscriptions in a manner consistent with what we predicted.[19] This adjustment of expenditures has not taken effect to the same degree for traditional subscriptions. Seven of the eight institutions bought more traditional subscriptions than optimal in 1998, yet only two of the seven responded by decreasing the number bought in 1999. Further, only three of the eight institutions made any changes at all to their traditional subscription lineup. This suggests an

[18] As 1999 PEAK access is for 8 months, the number of 1999 Generalized Subscriptions was multiplied by 1.5 for comparison with 1998.

[19] One of the institutions that increased token purchases despite over purchasing in 1998 was more foresightful than our simple learning model: its usage increased so much that it whose ran out of tokens less than six months into the final eight-month period of the experiment.

inertia that cannot be explained solely by direct costs to the institution, but perhaps can be partially explained by looking at total costs. As it is less costly for users to access articles included in traditional subscriptions (due to the fact that passwords are not required), perhaps libraries want to ensure access to certain journals at the least possible user cost. It may also be that the traditional emphasis on building complete archival collections for core journal titles carried over into electronic access decision making even though PEAK offered no long-term archival access.

Dependent variable is Forecast Error	
Year 1999	-35.691*
	(9.282)
Green	54.574*
	(9.975)
Red	53.331*
	(8.071)
Blue	85.757*
	(9.233)
Sample Size	24
R^2	0.850

No constant term is included in the regressions.
Standard errors are shown in parenthesis.
* Significant at the 99% level, ** Significant at the 95% level;

Table 13: Estimation Results on Forecast Error

In Table 13 we report the results from the estimation of the forecast error after controlling for differences in the experimental groups. From the Year 1999 dummy variable we can see that institutions did a better job of forecasting their optimal bundle in 1999. We also considered other control variables, such as the institution level of expenditures, fraction of the year participating in the experiment and number of potential users, but their contribution to explaining the forecast error was not statistically significant.

In addition to comparing the total number of subscriptions for an institution with the optimal number, we can also identify the optimality of individual subscriptions. In particular, we can assess how many of the optimal number of subscriptions an institution actually held. Further, we can identify how many of an institutions subscriptions were not optimal in the sense that access would have been less expensive with via other available access products.

In Table 14 we present an analysis of the subscriptions selected by in-

Instid	Year	Total Subscriptions	Percent Optimal	Percent of Optimal Not Subscribed To	Percent of Subs Accessed
3	1998	907	53.25%	3.40%	92.50%
10	1998	23	0.00%	100.00%	65.22%
11	1998	663	3.62%	0.00%	84.46%
12	1998	22	0.00%	100.00%	81.82%
13	1998	205	0.49%	0.00%	12.68%
14	1998	72	0.00%	N/A	36.11%
15	1998	102	0.00%	N/A	48.04%
3	1999	907	74.97%	7.73%	97.02%
10	1999	23	13.04%	76.92%	65.22%
13	1999	205	29.76%	62.58%	86.83%
14	1999	72	0.00%	N/A	20.83%
15	1998	102	10.78%	8.33%	84.31%

Table 14: Optimality of Subscription Choices

stitutions. While we see a wide variation both in terms of the percent of subscriptions that are optimal and the percent of optimal journals that the institution does not subscribe, there is clearly opportunity for improvement. We would expect to see better decisions as the institutions gained experience. It is also surprising the rather large percentage of subscribed to journals which were not accessed at all.

In determining optimal cost, we were trying to find the optimal bundle for a given set of accesses. This set of accesses is clearly not exogenous. As a simple example, let us assume that a subscription to a given journal requires 25 accesses in order to pay for itself. Now assume that the institution in question did not subscribe to that journal, and that 20 tokens were used to access articles in the time period. At first look, it appears as though the institutions did the optimal thing. Let us assume, however, that we know that accesses increase by 50% when no password is required. It now appears as though the institution should have subscribed to that journal.

Instid	Year	Trad. Subscriptions		Addit. Articles		Increase	Total
		Actual Optimal	Rescaled Optimal	Actual Optimal	Rescaled Optimal	Optimal Cost	Access Increase
3	1998	500	556	1099	1130	9.39%	12.53%
3	1999	737	805	236	146	4.85%	7.46%
11	1998	24	31	2532	3019	21.11%	21.09%
12	1998	1	1	254	287	17.76%	13.67%
14	1999	0	0	168	249	48.21%	48.21%
15	1999	12	17	242	366	47.56%	60.36%

Table 15: Optimal Bundles With Barrier Free Access: Selected Institutions

We do not, however, know by how much access have increased if no tokens were require. We do however, have estimates from 4. In 15 we present the optimal bundles for selected institutions after rescaling for barrier free access.[20] For most institutions, the optimal number of journals increases, as more journals pay for themselves. We estimate that in general, the cost of the optimal bundle would not increase greatly if all access were at minimum user cost. In those cases where the number of traditional subscriptions increase, the increase in optimal cost is generally less than the increase in total accesses. Further, the greatest increase occurs for the institution where tokens were not available and the institution did not directly subsidize the per article fee.

7 Conclusion

Recent disputes over the ownership of business processes determining the number of "clicks" to buy a product bring to the forefront the importance of user costs in e-commerce success. In the PEAK experiment, we have evidence that for the information goods in question, these non-pecuniary costs are of the same magnitude as significant pecuniary costs. In a two-tiered decision problem such as in this project, where intermediaries determine the user costs required to access specific content, both the quantity *and* composition of demand is greatly effected by users reactions to these costs. Therefore any determination of what the intermediary "ought" to do must take these effects into account. Furthermore, we have initial evidence that suggests that users who come to expect information at zero marginal costs are far less likely to pay these non-monetary costs when requested than their counterparts who expect these costs. This finding is of great import to both those who design electronic information delivery and pricing systems as well as any intermediaries controlling information access and costs.

References

[1] MacKie-Mason, J.K. and Riveros, J.F., "Economics and Electronic Access to Scholarly Information," forthcoming in *The Economics of Digital Information,* Kahin, B. and Varian, H. eds. MIT Press, 2000.

[20] Observations chosen had a well defined scaling factor from 4 and had enough article accesses to be meaningful.

[2] MacKie-Mason, J.K., Riveros, J.F., and Gazzale, R.S."Pricing Electronic Access to Knowledge (PEAK): A Field Experiment," forthcoming in *The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy,* Vogelsang, I., and Compaine, B.M. eds. MIT Press, 2000.

[3] Haar, J., "Project PEAK: Vanderbilt's Experience with Articles on Demand," NASIG Conference, June 1999.

[4] King, D.W., and Griffiths, J.M., "Economic issues concerning electronic publishing and distribution of scholarly articles," *Library Trends*, Vol. 43, No. 4, 1995, pp. 713–740.