Pricing and Bundling Electronic Information Goods: Field Evidence

Jeffrey K. MacKie-Mason[*]
Juan F. Riveros[♦]
Robert S. Gazzale[♦]

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 report initial results. We found that although there is a steep initial learning curve, decision-makers rapidly comprehended our innovative pricing schemes. We also found that our novel and flexible "generalized subscription" was successful at balancing paid usage with easy access to a larger body of content than was previously available to participating institutions. Finally, we found that both monetary and non-monetary user costs have a significant impact on the demand for electronic access.

---

[*] 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.
[♦] Juan F. Riveros and Robert S. Gazzale are Ph.D. students in the Department of Economics, University of Michigan.
[♦] Juan F. Riveros and Robert S. Gazzale are Ph.D. students in the Department of Economics, University of Michigan.

1   Introduction[1]

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 support ongoing innovation. Consequently, although publishers and libraries face a number of challenges, they also have promising new opportunities.[2] 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 creating new pricing schemes and content bundles to take advantage of the characteristics of digital duplication and distribution.

The University of Michigan has completed 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 1996–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 then 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.

The products in PEAK are more complex than those studied in the theoretical literature on

bundling. For example, Bakos and Brynjolfsson (1998) and Chuang and Sirbu (1996) consider selling a complete bundle, which consists of all the goods available; selling the individual components; or letting consumers choose between these two options. Their analysis showed that when consumers have similar average valuations for the information goods, profits are highest from selling only a single, complete bundle. When consumers have different average values for the articles and value a small fraction of the goods, users will prefer to purchase individual items. When marginal costs are high relative to the goods valuations, unbundled selling is the seller-preferred strategy.

PEAK customers could buy producer-defined "sub-bundles" of articles, user-defined sub-bundles, or individual articles. Scholarly journal consumers are very heterogeneous in their preferences for the sub-bundles available and also only value a small fraction of the articles. For example, chemists and sociologists value chemistry articles quite differently. Therefore, in practice publishers do not offer a single complete bundle in print or electronic form.[3] Instead, they define "journals" which are article sub-bundles; we reproduced this option for electronic access to Elsevier content. We also created a novel option, "generalized subscriptions," with which users get a sub-bundle of articles at a discount, but the users select which articles are included in the bundle (ex post, or after the articles are published and as they become aware of them, not in advance). As a third option, users could purchase access to individual articles not included in a traditional or generalized subscription. We expect that these and other new product offerings will liberate sources of value previously unrealized by sorting customers into groups that value the content differently.

---

[2] See MacKie-Mason and Riveros (1999) for a discussion of the economics of electronic publishing.
[3] Not a single academic institution in the world subscribes to all 1200 Elsevier journals in print. Even the largest customers subscribe to only about two-thirds of the titles.

Publishing is an industry that deals directly in differentiated products that are protected as intellectual property. It is also an industry subject to substantial recent merger activity, entry, and exit. Therefore, understanding emerging product and pricing models for digital content and distribution is important for competition and intellectual property policy.[4] New value recoverable through intelligent pricing and bundling of electronic delivery also has implications for the value of broadband data transport as a component of the telecommunications industry.

Our initial analysis of the PEAK data sheds some light on this subject. We found that, while there is a steep initial learning curve, decision-makers rapidly develop an understanding of innovative pricing schemes. Further, they utilize the new usage data that digital access makes available to improve purchasing decisions. We have found that the user-defined sub-bundle (generalized subscription) enabled institutions to provide their communities with easy access to a much larger body of content than they previously had from print subscriptions. Finally, we have gathered substantial evidence that user costs—both monetary and non-monetary—have a very real impact on demand for electronic access. The effects of these costs must be taken into account both when designing and selecting electronic access products.

2   The Problem

Information goods such as electronic journals have two defining characteristics. The first and most important is low marginal (incremental) cost. Once the content is created and transformed into a digital format, the information can be reproduced and distributed at almost zero cost. Nevertheless information goods often involve high fixed ("first copy") costs of production. For a

---

[4] See, e.g., McCabe (1999).

typical scholarly journal most of the cost to be recovered by the producer is fixed.[5] The same is true for the distributor in an electronic access environment. With the cost of electronic "printing and postage" essentially zero, nearly all of the cost of distribution consists of the system costs due to hardware, administration, and database creation and maintenance—all costs that must be incurred whether there are two or two million users. Our experience with PEAK bears this out: our only significant variable operating cost was the time of the user support team which answers questions from individual users. This was a small part of the total cost of the PEAK service.

Electronic access offers new opportunities to create and extract value from scholarly literature. This additional value can benefit readers, libraries, distributors, and publishers. For distributors and publishers, capturing some of this new value can help to recover the high fixed costs. Increased value can be created through the production of new products and services (such as early notification services and bibliographic hyperlinking). Additional value that already exists in current content can also be delivered to users and in part extracted by publishers through new product bundling and nonlinear pricing schemes that become possible with electronic distribution. For example, journal content can be unbundled and then rebundled in many different ways. Bundling enables the generation of additional value from existing content by targeting a variety of product packages to customers who value the existing content differently. For example, most four-year colleges subscribe in print to only a small fraction of Elsevier titles. With innovative electronic bundling options content can be accessible immediately, on the desktop, by this population that previously had little access.[6]

The underlying economic motivation for the PEAK experiment is to learn how additional

---

[5] Odlyzko (1995) estimates that it costs between $900–$8700 to produce a single math article. 70% of the cost is editorial and production, 30% is reproduction and distribution.

[6] All participants in PEAK had immediate access to all content from all 1200 journals, under various payment

value can be extracted from existing content by means of innovative electronic product offerings and pricing schemes. In this article we present initial analyses of usage and economic behavior based on the eighteen months of data from the just-completed experiment. Over the next year we will study how users responded to different pricing schemes and assess the additional value created from new forms of bundling. We will analyze the impact of the different pricing schemes on publisher revenues. We will also compare our empirical evidence with the predictions from the economic literature on bundling of information goods.[7]

3 Access Models Offered

Participants 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 for this user-defined subscription on demand, after they are published, as users request articles that are not otherwise already paid for, until the subscription is exhausted.[8] All authorized users at an institution may access articles selected for an institutional generalized subscription.

3. *Per Article:* Unlimited access for a single individual to a specific article. If an article is not available in a subscribed journal, or in a generalized subscription, nor are there un-

---

conditions.

[7] New research results will be posted on http://www.lib.umich.edu/libhome/peak/ and on http://www-personal.umich.edu/~jmm/.

[8] 120 is the approximate average number of articles in a traditional printed journal for a given year. We refer to this bundle of options to access articles as a set of tokens, with one token used up for each article added to the generalized subscription during the year.

used generalized subscription tokens, then an individual may purchase access to the article for personal use.

The per article and generalized subscription options allow users to capture value from the entire corpus of articles, without having to subscribe to all of the 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 price and bundling schemes affect both how 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 for both libraries and individuals was that the "tokens" cost substantially less per article than the per-article license price, and less than the full cost of interlibrary loan. By predicting in advance how many tokens will be used (and thus bearing some risk), the library can essentially pre-pay for interlibrary loans, at a reduced rate. There is an additional benefit: unlike an interlibrary loan, all users within the community have ongoing, unlimited access to the articles that were obtained with generalized subscription tokens. Generalized subscriptions provide some direct revenue to publishers, whereas interlibrary loans do not. In addition, unlike commercial article delivery services, generalized subscriptions produce a committed flow of revenue at the beginning of each year, and thus shift some of the risk for usage (and revenue) variation from the publisher to the users. Another advantage is that they open up access to the entire body of content to all users, and by thus increasing user value from the content, provide an opportunity to obtain greater returns from the publication of that content.

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

Table 1: Access Models

Participating institutions in the experiment were assigned randomly to one of three different experimental treatments, which we labeled as the Red, Green and Blue groups. Users in every group could purchase articles on a per article basis; in the Green group they could also purchase institutional generalized subscriptions; in the Blue Group they could purchase traditional subscriptions; in the Red group they could purchase all three types of access. Twelve institutions are participating in PEAK: large research universities, medium and small colleges and professional schools, and corporate libraries. Table 1 shows the distribution of access models and products offered to the participating institutions.

4   Pricing

Pricing electronic access to scholarly information is far from being a well-understood practice. Based on a survey of 37 publishers, Prior (1999) reported that when both print-on-paper and electronic versions were offered, 62% of the publishers have had a single combined price, with a surcharge over the paper subscription price of between 8% and 65%. The most common surcharge is between 15-20%. Half of the respondents offer electronic access separately at a price between 65% and 150% of print, most commonly between 90% and 100%. Fully 30% of the participating publishers have changed their pricing policy just this year. In this section we will describe the pricing structure chosen in the PEAK experiment and the rationale behind it.

For content that can be delivered either on paper or electronically, there are three primary cost elements: content cost, paper delivery cost, and electronic delivery costs. The price levels chosen for the experiment reflect the components of cost, adjusted downward for an overall dis-

count to encourage participation in the experiment.

The *relative prices* between access options were constrained by arbitrage possibilities that arise because users can choose different options to replicate the same access. In particular, the price per article in a per-article purchase had to be greater than the price per article in a generalized subscription, and this price had to be greater than the price per article in a traditional subscription. These inequalities impose the restriction that the user cannot save by trying to replicate a traditional subscription by subscribing to individual articles or a generalized subscription, or save by replicating a generalized subscription by paying for individual articles.

To participate in the project, each institution paid the University of Michigan an individually negotiated institutional participation license fee (IPL), roughly proportional to the number of authorized users. In addition, the access prices for articles were:

1. *Traditional Subscription:* The library pays an annual fee per traditional subscription. The fee depends on whether the library previously subscribed to the paper version of the journal, as follows:

   - If the institution previously subscribed to a paper version of the journal, the cost of the traditional subscription is $4 per issue[9] regardless of journal title. Since the content component is already paid, the customer is only charged for an incremental electronic delivery cost.[10]

---

[9] An "issue" is identical to a print issue. For most Elsevier Journals there are several volumes per subscription year, where the volume equals a standard measure (depending on the journal) of 2-4 issues. The range in the number of issues in a year is from 4 to 129, with the number of volumes in the year for these titles ranging then from 1 to 61. The actual prices were adjusted to reflect more than a full year of content during the first project year, and less than a year of content the second project year.

[10] The institution must continue to subscribe to the paper version. If a library cancelled a paper subscription during the life of PEAK, it was required to pay the full paper cost plus 10% for the electronic subscription, to make it uneconomical to use electronic subscriptions to replace previously subscribed paper subscriptions. This was not intended to represent future pricing schemes, but to protect Elsevier's subscription base during the experiment since the PEAK prices were deeply discounted.

- If the institution was not previously subscribed to the paper version, the cost of the traditional subscription is $4 per issue, plus 10% of the paper version subscription price. In this case, the customer is charged for the electronic delivery cost plus a percentage of the content cost.

2. *Generalized Subscription:* A library pays $548 for access rights to 120 articles ($4.56 per article). These articles are selected on demand, after publication. A library may purchase any number of generalized subscriptions it wishes, but all generalized subscriptions must be purchased within the first 60 days after the start of the billing year. Once accessed, articles may be used any number of times by all members of the institution, for the life of the project.

3. *Per-article licensing:* A library or individual pays for access to individual articles. The per-article fee is $7 per article.[11] Once licensed, an article may be used any number of times by the individual licensor for the life of the project. Most electronic delivery services charge *per use*, not per article.

The mapping of costs to prices is not exact, and because there are several components of cost the relationship is complicated. For example, although electronic delivery costs are essentially zero, there is some incremental cost to creating the electronic versions of the content (especially under Elsevier's current production process which is not fully unified for print and electronic publication). This electronic publication cost plus user support costs underlie the $4 per issue price for electronic delivery of traditional subscriptions.

---

[11] The per-article fee is the same whether paid by a library on behalf of an individual, or paid by the individual directly.

5 Revenues and Costs

	Traditional Subs		Generalized Subs		Individual Articles				
Year	Subs	Revenue	Subs	Revenue	Articles	Revenue	All Access Revenue	IPL Revenue	Total Revenue
1997-98	1939	$216,018	151	$82,748	275	$1,925	$300,691	$140,000	$440,691
1999*	1277	$33,608	92	$50,416	3186	$22,302	$106,326	$42,000	$148,326
Annualized 1999§	1277	$78,996	138	$75,624	4779	$33,453	$188,073	$84,000	$272,073
Total† 1997-1999	3216	$295,014	289	$158,372	5054	$35,378	$488,764	$224,000	$712,764
*Article use through August 1999									
§ Annualization done by scaling the quantity of Generalized Subscriptions and per article purchases. Traditional subscriptions priced at the full year rate.									
† Annualized									

Table 2: Revenues

In Table 2 we summarize the revenues received during the PEAK experiment. The total revenue was over $580,000.[12] The first and third rows report the annual revenues, with 1999 adjusted to reflect an estimate of what revenues would have been if the service were to run for the full year.[13] We can see that between the first and second year of the service, the number of traditional subscriptions was substantially decreased: this occurred because two schools cancelled all of their (electronic) subscriptions. By reducing the number of journal titles under traditional subscription, the users of these libraries needed to rely more heavily on the availability of generalized subscription tokens, or they had to pay the per article fee.

A full calculation of the costs of supporting the PEAK service is difficult, given the mix and dynamic nature of costs (e.g., hardware). We estimate expenditures reached nearly $400,000 during the 18 month life of the project. Of this cost, roughly 35% was expended on technical infrastructure and 55% on staff support (i.e., system development and maintenance, data loading,

---

[12] The University of Michigan received $182,000 in IPL fees for providing the service. Elsevier Science received the remainder, net of payment processing costs, for the value of accessing the content.

[13] Due to delays in starting the project, the first revenue period covered content from both 1997-98, although ac-

user support, authentication/authorization/security, project management). Participant institution fees covered approximately 45% of the project costs, with vendor and campus in-kind contributions covering another 20-25%. UM Digital Library Production Service resources were also devoted to this effort, reflecting the University of Michigan's contribution to providing this service to its community and also its interests in supporting the research.

In the following sections, we present preliminary results on the usage of the PEAK service. We summarize some demographics of the user community, and then analyze usage and economic behavior.

6   User Demographics

In the PEAK project design, substantial amounts of content were freely available to participants. We call this "unmetered" content: not-full-length articles, plus all content published pre-1997 during 1998, and all pre-1998 content during 1999.[14] Unmetered content and articles covered by traditional subscriptions could be accessed by any user from a workstation associated with one of the participating sites (authenticated by the computer's IP address). This user population consisted of about 340,000 authorized users. If users wanted to use generalized subscription tokens, or to purchase individual articles on a per-article basis, they had to obtain a password and use it to authenticate.[15] We have more complete data on the 3546 users who obtained and used passwords.

cess was available only during 1998. For this period, prices for traditional subscriptions were set to equal $6/issue, or 1.5 times the annual price of $4 per issue, to adjust for the greater content availability.

[14] A substantial amount of material, including all content available that was published two calendar years prior, was available freely without any additional charge after an institution paid the IPL fee to join the service. We refer to this as "unmetered". Full-length articles from the current two calendar years were "metered": users could access it only if the articles were paid for under a traditional or generalized subscription, or purchased on a per article basis.

[15] Through an onscreen message we encouraged all users to obtain a password and use it every time in order to provide better data for the researchers. Only a small fraction apparently chose to obtain passwords based solely on our urging; most apparently obtained passwords because they were necessary to access a specific article

	Status					
Division	Faculty	Staff	Grad. Student	Undergrad	Other	Total
Engineering, science and medicine	408	214	1032	211	38	1903
Architecture and urban planning	103	11	47	16	19	196
Education, business, informa-tion/library science and social science	91	43	287	46	2	469
Other	178	240	350	176	34	978
Total	780	508	1716	449	93	3546

Table 3: Users who obtained and used PEAK password

In Table 3 we report the field and status distribution of the users who obtained passwords and who used PEAK at least once. Most of the users are from engineering, science, and medicine, reflecting the strength of the Elsevier collection in these disciplines. 70% of these users were either faculty or graduate students, although the relative fractions of faculty and graduate students vary widely by discipline. Our sample of password-authenticated users is probably not representative of all electronic access usage, but it represents an important group of users who are more motivated (and the sample includes all of those who obtained articles via either general-ized subscription tokens or by per article purchase).

7   Usage

Group	Un-metered	Trad'l Article: 1st Use	Trad'l Article: 2nd or higher use	Gen'l Article: 1st Use	Gen'l Article: 2nd or higher use	Per-Article: 1st Use	Per-Article: 2nd or higher use	Total Accesses
Green	24,632	N/A	N/A	8,922	3,535	194	108	37,391
Red	96,658	27,140	11,914	9,467	4,789	75	26	150,069
Blue	13,911	2,881	597	N/A	N/A	3,192	63	20,644
All	135,201	30,021	12,511	18,389	8,324	3,461	197	208,104

Table 4: Unique content accesses by group and access type: Jan 1998–Aug 1999. N/A indicates "Not Applicable" because that access option was not available to participants in that group.

In Table 4 we summarize usage of PEAK through August 1999. There have been 208,104

different accesses to the content in the PEAK system.[16] Of these, 65% were accesses of "un-

metered" material. However, one should not leap to the conclusion that users will access schol-

arly material much less when they have to pay for it, though surely that is true to some degree.

First, to *users* much of the "metered" content appeared to be free: the libraries paid for the tradi-

tional subscriptions and the generalized subscription tokens. Second, the quantity of "un-

metered" content in PEAK was substantial: as of January 1, 1998, all 1996 content, and some

1997 content was in this category. On January 1, 1999, all 1996 and 1997 content and some

1998 content was in this category.

Generalized subscription "tokens" were used to purchase access to 18,389 articles ("1st to-

ken"). These articles were then accessed an additional 8,324 ("2nd or higher tokens"). The

number of subsequent accesses per generalized subscription article were significantly higher than

the subsequent accesses for traditional subscription articles. First, we compare subsequent ac-

cesses by users in the Green group, which had generalized but no traditional subscriptions, to

subsequent accesses by users in the Blue group, which had traditional but no generalized sub-

scriptions. The subsequent accesses per article were 0.4 for generalized, and 0.21 for traditional.

To control for cross-institutional differences in users, we can compare the subsequent accesses

by users in the Red group, which had both generalized and traditional subscriptions. Here the

subsequent accesses were 0.51 for generalized and 0.44 for traditional. The difference is not as

large for Red users, but this is to be expected: Presumably the Red group librarians selected their

traditional subscriptions to be those journals with the highest expected use, and generalized to-

---

[16] We limited our scope to what we call "unique accesses"—counting multiple accesses to a given article by an individual during a PEAK session as only one access. For anonymous access (i.e. access by users not entering a password), we define a "unique" access as any number of accesses to an article within 30 minutes from a particular IP address. For authenticated users, we define a "unique" access as any number of accesses to an article by an authenticated user within 30 minutes of first access.

kens were used only for *other* journals. Thus, the result is quite striking: despite the bias towards having the most popular articles in traditional subscriptions, repeat usage was still higher for generalized subscription articles. These results confirm our prediction that the generalized subscription is valuable because it allows users rather than the publishers and editors to select the articles purchased, and thus even among the subset of articles that are read at least once, the articles in generalized subscriptions are more popular.

A total of 3,461 articles were purchased individually on a per article basis; these were accessed 1.06 times per article on average. This lower usage than for generalized subscription articles is not surprising: Articles purchased per item can be subsequently viewed only by the particular user who purchased them, whereas once selected a generalized subscription article can be viewed by every authorized user at the institution. Thus, given an overall subsequent use rate of 0.45 for generalized subscription articles, we can estimate that *initial* individual readers accessed individually paid (by token or per-article purchase) articles 1.06 times, and *additional* users accessed these articles .39 times. It appears on average there is at least one-third additional user per article under the more lenient access provisions of a generalized subscription token.[17]

---

[17] Note that we could only measure *electronic* accesses to an article. Users were permitted to print a single copy of an article for personal use, so the total accesses – including use of printed articles – is likely to be higher.
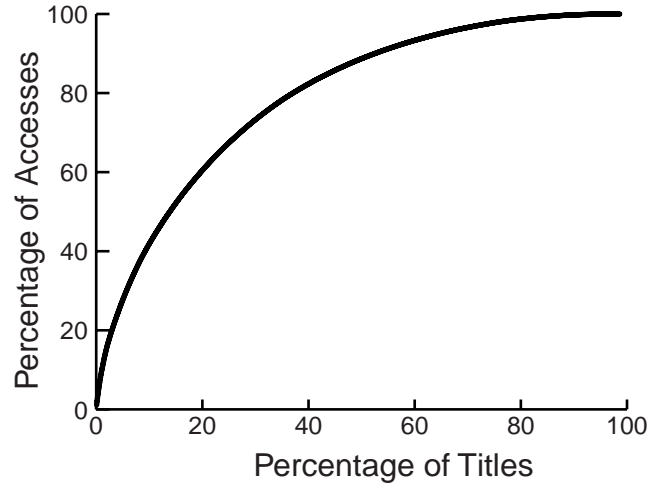
Figure 1: Concentration of Accesses

In we show a curve that reveals the concentration of usage among a relatively small number

of Elsevier titles. We sorted journal titles by frequency of access. Then we calculated the small-

est number of titles that together comprised a given percentage of total accesses. For example, it

only required 37% of the 1200 Elsevier titles to generate 80% of the total accesses. 40% of the

total accesses were accounted for by only about 10% of the journal titles.
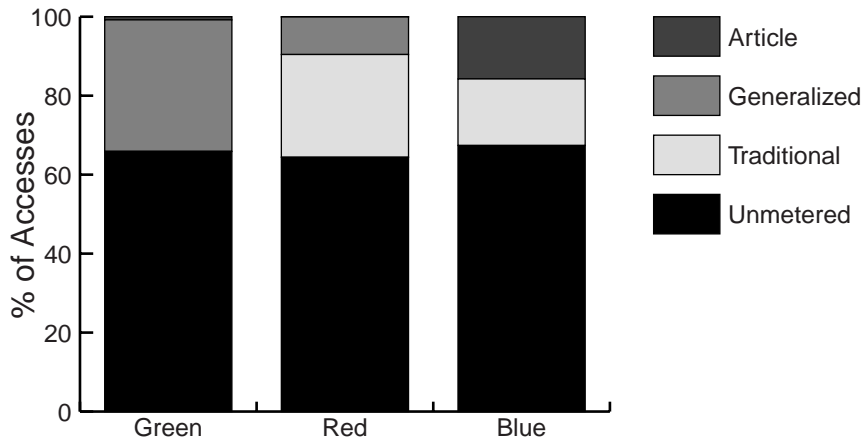
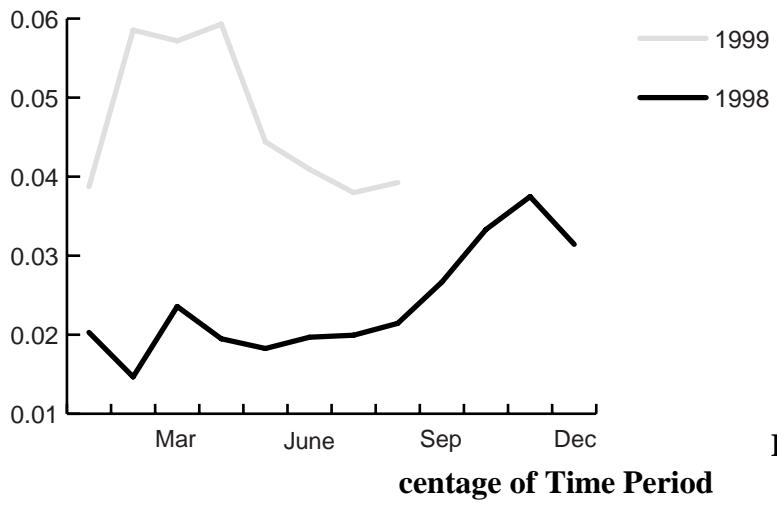**Figure 2:** : Percentage of Access by Access Type and Group: Jan 1998–Aug 1999
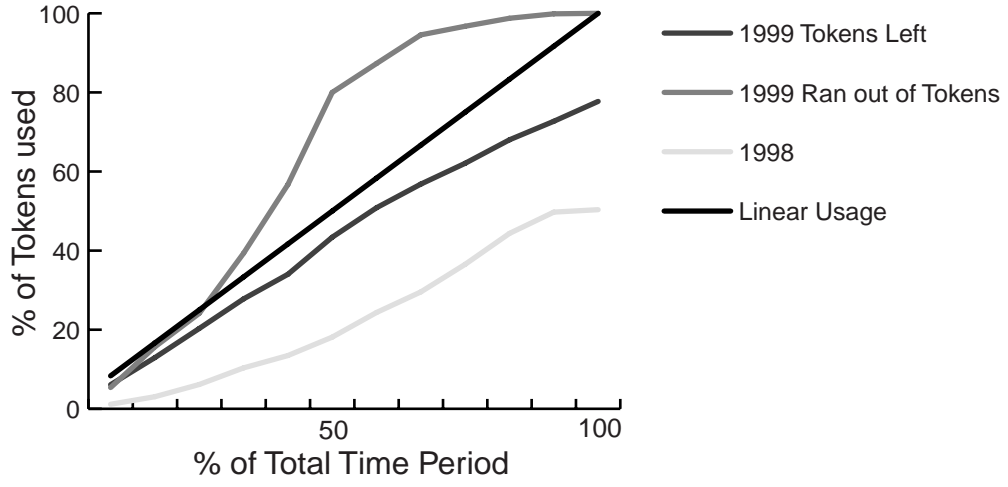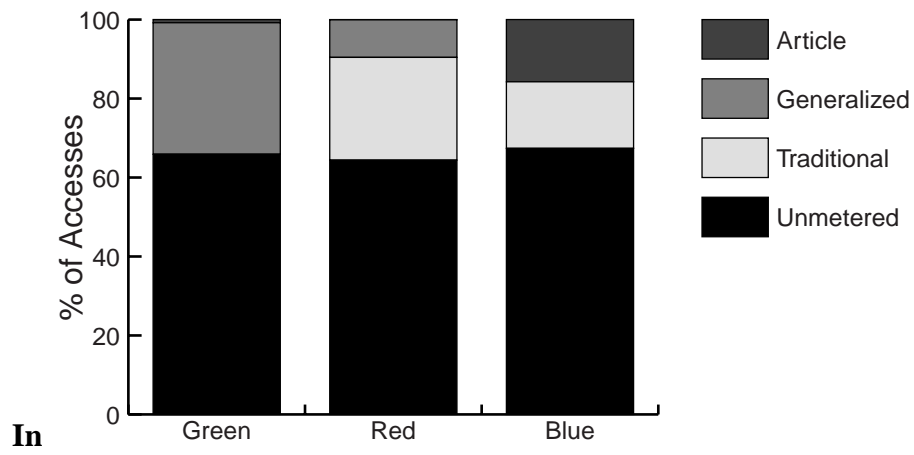
16

Figure 4: Token Use as Percentage of Time Period

In

Figure 2: : F   **Figure 3:** Total Accesses Per Potential User: Jan   Aug 1999
1998–Aug 1999

17

we compare the fraction of accesses within each group of institutions that are accounted for by traditional subscriptions, generalized subscriptions and per article purchases. Of course, the Green and Blue groups only had two of the three access options. We observe that when institutions had the choice of purchasing generalized subscription tokens, their users purchased essentially no access on a per article basis. Of course, this makes sense as long as tokens are available: it costs the users nothing to use a token, but it costs real money to purchase on a per article basis. What the data also indicate is that institutions that could purchase generalized subscription tokens tended to purchase more than enough to cover all of the demand for articles by their users; i.e., they didn't run out of tokens in 1998. We show this in aggregate in
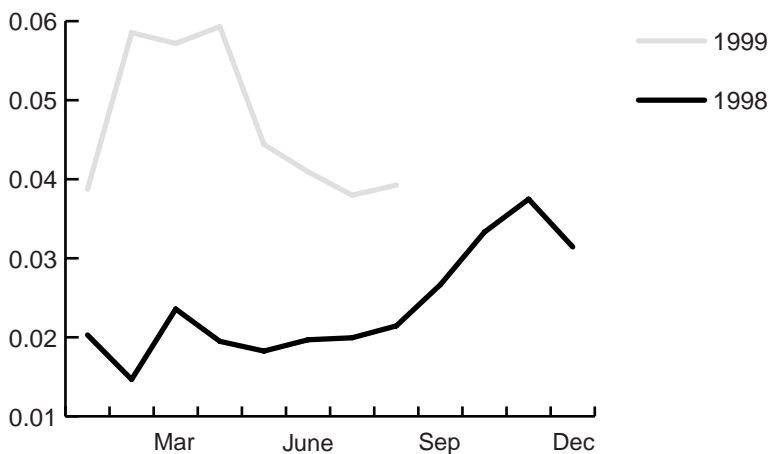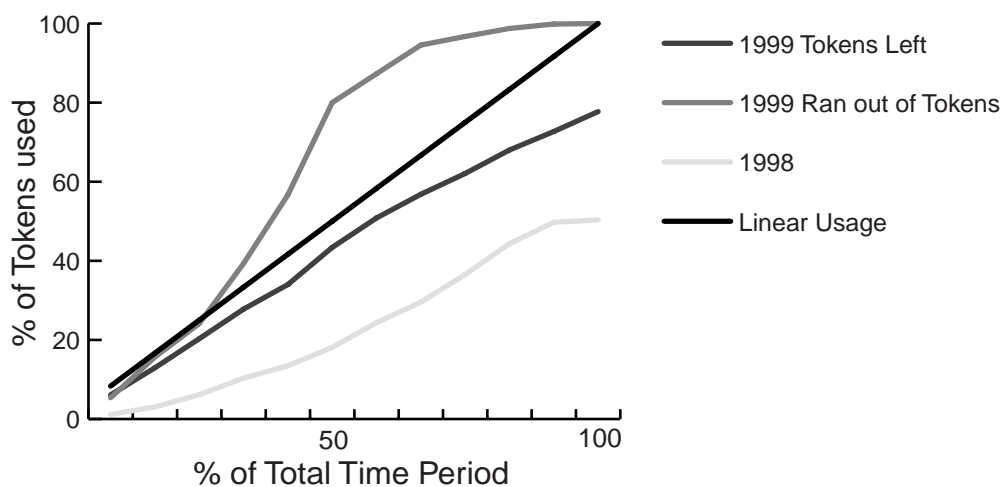

Figure 4: only about 50% of the

tokens purchased for 1998 were in fact used. Institutions who did not run out of tokens in 1999

appear to have done a better job of forecasting their token demand for the year, as 78% of the

tokens purchased for 1999 were used. Institutions who ran out of tokens used about 80% of the

tokens available by the beginning of May.

Articles in the "unmetered" category constituted about 65% of use across all three groups, regardless of which combination or quantity of traditional and generalized subscriptions an institution purchased. The remaining 35% of use was paid for with a different mix of options depending on the choices available to the institution.

Figure 5: Total Accesses Per Potential User: Jan 1998–Aug 1999

8   Seasonality and Learning Effects

We show the total number of accesses per potential user for 1998 and 1999 in Figure 5. We divide by potential users (the number of people authorized to use the computer network at each of the participating institutions) because different institutions joined the experiment at different times. This figure thus gives us an estimate of learning and seasonality effects in usage. Usage per potential user was relatively low and stable for the first 9 months. However, it then increased to a level nearly three times as high over the next 9 months. We expect that this increase be due to more users learning about the existence of PEAK and becoming accustomed to using it. Note also that the growth begins in September, 1998, the beginning of a new school year with a natural bulge in demand for scholarly articles. We also see pronounced seasonal effects in usage: local peaks in March, November and April and decreases of usage in May and December.
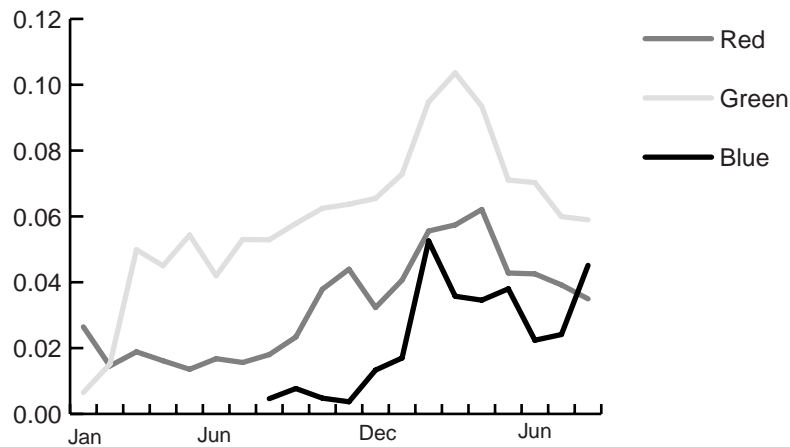
Figure 6: Total Accesses Per Potential User by Group: Jan 1998–Aug 1999

In **Figure 6** we show the accesses per potential user by group. Uses have increased over time across groups reflecting the learning effect. The Green group has the highest access per user across the period. This can be explained by the fact that the Green group includes two corporate libraries in which the user heterogeneity is lower thus increasing the number of accesses per potential user. Uses for the Blue group are the lowest and show a late surge in learning (the Blue institutions started using the system in the second half of 1998). Generalized subscriptions had a much lower user cost per access because the library paid for the tokens, while the individual user had to pay for per article purchases. Our hypothesis, which we will explore in more detail in later sections, is that when generalized subscriptions were available, more articles were accessed than when individuals could only access an article through an individual purchase.

Year	Unmetered	Traditional	1st Token	1st Per-Article	2nd or higher Token	2nd or higher Per-Article	Total
Mar–Aug 1998	7.61	2.54	1.44	0.02	0.53	0.01	12.16
Mar–Aug 1999	19.32	4.63	2.06	0.75	1.12	0.03	27.90
Percent Change	153%	82%	42%	3140%	109%	206%	129%

Table 5: Learning: Two-year comparison: March-August. (Access per potential user in hundredths)

To see the learning effect without interference from the seasonal effect, we calculated usage by type of access per average potential user over the same six-month (March-August) period of 1998 and 1999; see Table 5. Overall, usage increased 129% from the first year to the second. First token use increased by 42% and first per article purchases increased by 3,140 %. The increase in per article purchases could be explained in part by the fact that the institutions in the Blue group started using PEAK after June 1998.

9   Repeated Access and Recency

To study repeated accesses, we selected only those articles first accessed in 1998 that were accessed three or more times through the end of the project ("high use articles"). In Figure 7, we show that only 7% of articles accessed were accessed three or more times. We counted the number of times high use articles were accessed in each month subsequent to the initial access; see Figure 8. What we see is that almost all access to high use articles occurred during the first month. In the second and later months, there was a very low rate of use that persisted for about 7 more months, then faded out altogether. Thus, we see that, even among the most popular articles, recency was very important.
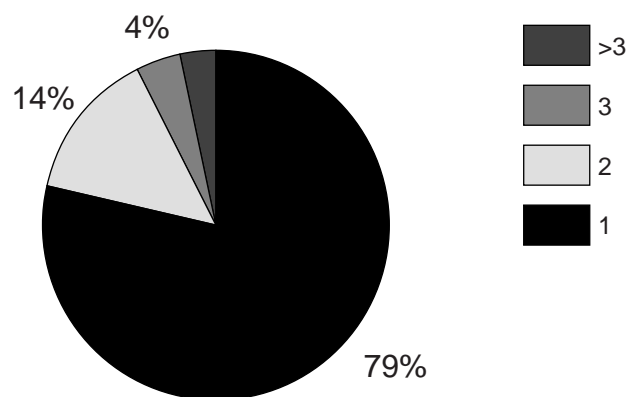
Figure 7: Percentage of Articles by Number of Times Read
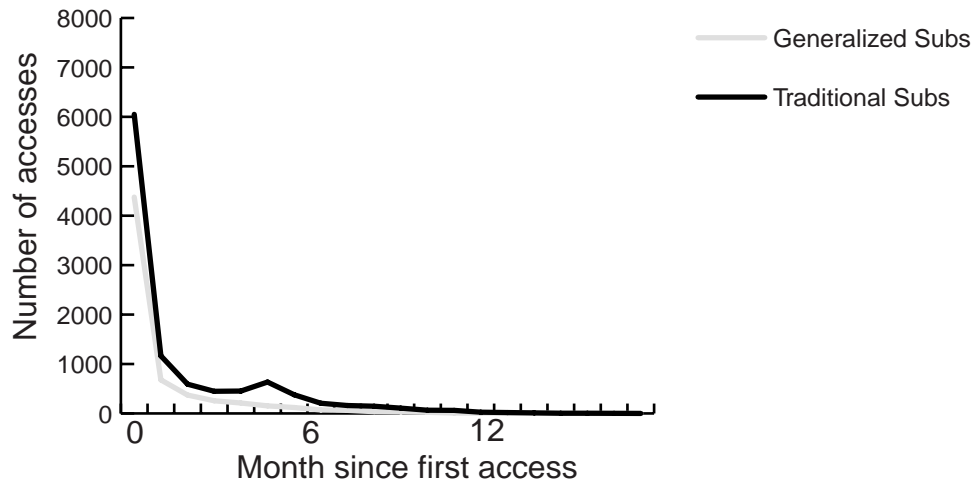
Figure 8: Subsequent Accesses by Month for 1998 High Use Articles

Although recency appears to be quite important, we saw in Table 4: **Unique content accesses by group and access type: Jan 1998–Aug 1999. N/A indicates "Not Applicable" because that access option was not available to participants in that group.**

 that over 60% of accesses were for content in the "unmetered" category, most of which was over one year old. Although monetary price to users for most "metered" articles was zero (if accessed via institution-paid traditional or generalized subscriptions), user costs are generally higher. To access an article using a generalized subscription token, the user must get a password, remember it (or where she put it), and enter it. If the article is not available in a traditional subscription and no tokens are available, then she must do the above plus pay for the article. (If the institution subsidizes the per article purchase, it might require filing paperwork for reimbursement). There are real user cost differences between the "unmetered" and "metered" content. The fact that usage of the older, "unmetered" content is so high, despite the clear preference for recency, supports the notion that users respond strongly to costs of accessing scholarly articles.

Purchasing an article with a generalized token rather than on a per article basis offers a distinct benefit to institutions: the ability for others at an institution to access the article without ad-

ditional monetary cost. One would expect that this benefit would be of most value to institutions with either a large research community or a community with very homogenous research interests. We investigated the pattern of subsequent accesses to articles purchased with a token at all academic institutions where this method was available. Due to the fact that institutions started their participation at various times throughout 1998, we looked at data for 1999. We divided the academic institutions into 2 groups: large research universities and other academic institutions.[18]

Institution	Group	Total Subsequent Access Per Token Spent				
		Subsequent Accesses	Anonymous Access	Initial User	Other Authenticated	Total Non-Initial
3	red	0.76	0.60	0.11	0.05	0.65
5	green	0.52	0.23	0.21	0.08	0.31
7	green	0.40	0.21	0.15	0.04	0.25
8	green	0.26	0.08	0.14	0.04	0.12
9	red	0.36	0.17	0.13	0.06	0.23
10	red	0.42	0.29	0.09	0.04	0.33
11	red	0.83	0.61	0.14	0.08	0.69
12	red	0.42	0.15	0.22	0.04	0.19
Total		0.45	0.27	0.14	0.05	0.32
Total Red		0.52	0.35	0.12	0.05	0.40
Total Green		0.38	0.17	0.15	0.05	0.22
Lg. Research Univs.		0.58	0.39	0.13	0.06	0.45
Other Academic		0.38	0.20	0.13	0.04	0.25

Table 6: Subsequent Unique Access to Articles Purchased with Tokens: Jan-Aug 1999

We present our results in **Table 6**, which details the number of subsequent unique accesses per token used in 1999. It is apparent that the majority of subsequent access to articles previously purchased with generalized subscription tokens was anonymous.[19] This further suggests that, for the most part, users incurred the cost of password use only when necessary. We note that password-authenticated subsequent access to articles by the initial user does not appear to depend on institution size, while subsequent access by other authenticated users is in fact higher

---

[18] The large research universities are institutions 3, 9, and 11.
[19] Password authentication was required to spend an available token to access an article. Anyone using a workstation attached to that institution's network could thereafter access that particular article anonymously by anyone.

at large research institutions than at other academic institutions.[20] The difference in subsequent access becomes sizable when one considers anonymous access. Unique anonymous subsequent access is markedly higher for large research institutions, and, in light of the results for subsequent access by initial users, there is no reason to believe that this result is driven by anonymous subsequent access by the initial user. We therefore added unique anonymous accesses to accesses by subsequent authenticated users to derive a statistic measuring total access to articles by users other than the initial user. Subsequent access by other users is significantly greater at large research universities, about .42 versus .24 subsequent accesses per article.[21]

These results demonstrate two benefits of our implementation of generalized subscriptions. First, users will, for the most part, avoid the costs of password use if possible. In subsequent papers, we will analyze the transaction logs to study user behavior when presented with the need to enter or obtain a password. Second, the purchase of a token created a positive externality, namely the opportunity for others to access the article without incurring the costs associated with password use. The benefit of this externality is more pronounced for larger institutions.

10   Actual vs. Optimal Choice

In determining to which scholarly print journals to subscribe, librarians are in an unenviable position.[22] They must determine which journals best match the needs and interests of their community subject to two constraints. Their first constraint is budgetary, which has become increasingly binding of late as renewal costs have tended to rise faster than serial budgets (Harr, 1999). The second constraint is that libraries have incomplete information in terms of community needs.

---

[20] The *t* statistic for the null hypothesis of equality of the means is -2.8628. The *p* value is 0.42%.
[21] The t statistic for the null hypothesis of equality of the means is -9.7280. The null hypothesis is rejected at any meaningful level of significance

At the heart of this problem is the fact that a traditional print subscription forces libraries to purchase publisher-selected bundles (the journal), while users are interested primarily in the articles therein. The library generally lacks information about which articles will constitute the small fraction of all articles[23] their community valued. 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 and 50% in 1999 were outside the institution traditional subscription base.) Second, the transaction logs that are feasible for electronic access enabled us to provide libraries with detailed monthly reports detailing which journals and articles their community accessed. Detailed usage reporting should enable libraries to provide additional value to better allocate their serials budgets to the most valued journal titles or to other access products.

Instit.	Traditional		Generalized		Per Article		TOTAL			
	Actual	Optimal	Actual	Opti-mal	Actual	Opti-mal	Actual	Optimal	Savings	Percent
3	25,000	17,000	2,740	3,836	7	133	27,747	20,969	6,778	24.43%
5	N/A	0	15,344	6,576	0	169	15,344	6,745	8,599	56.04%
6	N/A	0	0	548	672	0	672	548	124	18.45%
7	N/A	0	24,660	12,604	0	0	24,660	12,604	12,056	48.89%
8	N/A	0	13,700	2,740	0	0	13,700	2,740	10,960	80.00%
9	0	556	13,700	6,576	0	56	13,700	7,188	6,512	47.53%
10	4,960	323	8,220	7,672	0	483	13,180	8,478	4,701	35.67%
11	70,056	5,217	2,192	13,700	0	84	72,248	19,001	53,247	73.70%
12	2,352	107	2,192	1,096	0	98	4,544	1,301	3,243	71.37%
13	28,504	139	N/A	0	952	1,120	29,456	1,259	28,197	95.73%

[22] For an excellent discussion of the collection development officer's problem, see Haar (1999).

[23] The percentage of articles read through August 1999 for academic institutions participating in PEAK ranged from .12% to 6.40%. An empirical study by King and Griffiths (1995) 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.

14	17,671	0	N/A	0	294	504	17,965	504	17,461	97.19%
15	18,476	0	N/A	0	0	1,176	18,476	1,176	17,300	93.63%
Red	102,367	23,203	29,044	32,880	7	854	131,418	56,937	74,481	56.67%
Green	0	0	53,704	22,468	672	169	54,376	22,637	31,739	58.37%
Blue	64,651	139	0	0	1,246	2,800	65,897	2,939	62,958	95.54%

Table 7: Actual vs. Optimal Expenditures on PEAK Access Products: 1998

In order to estimate an upper bound on how much the libraries could benefit from better usage data, we determined each institution's optimal bundle for 1998 had they been able to *perfectly forecast* which articles would be accessed. We compared the cost of the optimal bundles with the institutions' actual expenditures.[24] Obviously even with extensive historical data, libraries would not be able to perfectly forecast future usage. The realized benefits from better usage data would clearly be less than the "upper bound" we present in Table 7. We can identify, however, a few trends. Institutions in the Red and Blue groups purchased far too many traditional subscriptions, and 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 decision making. A preference for non-variable expenditures would induce a library to rely more heavily on traditional and generalized subscriptions, and less on reimbursed individual article purchases or interlibrary loans.[25]

---

[24] An appendix describing the optimal cost calculation is available from the authors or the PEAK website: http://www.lib.umich.edu/libhome/peak/.

[25] 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.

Some of the excess expenditure can thus be viewed as an insurance premium.

Institution	Traditional		Generalized	
	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 8: 1999 Expenditures: Acutal Increase/Decrease vs. Optimal Cost Predicted

We might expect to see that in determining 1999 expenditures, institutions' access product decisions would conform to a simple learning dynamic—increasing expenditures on products they underbought in 1998 and decreasing expenditures on products they overbought 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,[26] and compared this change with the change recommended by the learning dynamic. We present the results in Table 8.

Six of the nine institutions adjusted the number of generalized subscriptions in a manner con-sistent with what we predicted based on our learning dynamic. (It is interesting to note that one of the institutions whose increase in token purchases was the opposite of the learning dynamic ran out of tokens less than six months into the final eight-month period of the experiment.) This adjustment of expenditures has not taken effect to the same degree for traditional subscriptions.

---

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

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. It is possible that libraries wanted 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.

11   Effects of User Cost on Access

In a test of the impact of user cost on usage, 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, while Blue had only traditional. We compared "paid" access to individual articles (paid by generalized tokens or per article), as paid article access requires user cost either in terms of password entry or $7.00 per article. We normalized paid article access for the number of traditional subscriptions, as users at an institution with a larger traditional subscription base are less likely to encounter content they must purchases. 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.[27] We thus compared normalized paid article access per unmetered access across institutions in the Red and Blue groups.[28]

---

[27] Recall that "unmetered" means access to material for which no payment scheme is applied. Such content includes all articles more than one year old. We *are* able to measure "unmetered" transactions: with several different access pricing schemes in place it is hard to devise a transparent vocabulary to describe all contingencies.

[28] Normalized paid access per unmetered access is equal to $\dfrac{A_{pi}}{A_{fi}}\left(\dfrac{1200}{1200 - T_i}\right)$, A$_{pi}$ is paid access for institution

Institution	Group	Normalized Paid Accesses Per Unmetered Access
3	Red	0.06
9	Red	0.20
10	Red	0.30
11	Red	0.08
12	Red	0.23
	Red §	0.11
13	Blue	0.39
14	Blue	0.11
15	Blue	0.02

§ Average of Red institutions weighted by number of unmetered accesses.

Table 9: 1999 Normalized Paid Access Per Unmetered Access

We present the results for January through August 1999 in Table 9. In evaluating these statistics, one must keep two things in mind. First, institution 3 had a much larger number of traditional subscriptions than any other institution in our sample (875 traditional subscriptions for institution 3 compared with 205 for the next highest in our sample, institution 13). As the traditional subscriptions were selected to include most of the most popular titles, we would therefore expect a lower demand for paid access even after normalization. Second, institutions 9 and 11 ran out of generalized tokens in 1999. This severely throttled demand for paid access, as we will discuss below.

We can rank the institutions in the Blue group based on marginal cost to gain access to paid content, and compare these institutions to the Red group. Users at institution 13 faced no appreciable marginal cost to access paid content as users did not need to authenticate and paid access was invisibly subsidized by the institution. We would expect a level of paid access higher than that of the Red group, where most users would face the marginal costs of authenticating.[29] This

---

i, T_i is number of traditional subscriptions for institution i, and A_{fi} is the number of unmetered accesses for institution i.

[29] Only 27% of Red group unmetered access in 1999 was authenticated.

is in fact the case.[30]

Paid access at institution 14 was similarly subsidized by the institution, but password authentication was required. We would therefore expect a rate of paid access similar to that of the Red group. This in fact does seem to be the case, as both this institution and the Red group accessed approximately 11 paid articles per 100 unmetered articles.

Finally, per article access for users at institution 15 was not directly subsidized. Thus, users faced very high marginal costs for paid content: a $7.00 per article fee, credit card entry, and password entry. We would therefore expect that the rate of paid access to be lower than that of the Red group. This is the case, as users at institution accessed paid articles at one-fifth the rate of users at Red institutions.

We gain further understanding of the degree to which differences in user cost affects the demand for paid article access by looking at those institutions that depleted their supply of tokens at various points throughout the project. There were three institutions that are in this category: institution 3 ran out of tokens in November 1998; institution 11 in May 1999; and institution 9 in June 1999. Once the tokens were depleted, a user wanting to view a paid article not previously accessed by the institution would then have 3 choices. 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. Third, she could also request the article through an interlibrary loan, which also involves higher costs (from filling out the request form and wait-

---

[30] This result is even more striking when one considers the that this institution had the second largest traditional subscription base in our sample and we are, if anything, under-correcting for the self-selection of popular journals for traditional subscriptions.

ing for the article to be delivered) than spending a token.[31]

	Institution 3	Institution 9	Institution 11
30 days prior token depletion	0.0950	0.2020	0.1603
30 days after token depletion	0.0018	0.0000	0.0035
Decrease from base	-98.11%	-100.00%	-97.82%

Units: Normalized paid access per unmetered access.

Table 10: Effect of Token Depletion on Demand for Paid Content

For each of the institutions that ran out of tokens, we present in Table 10 the normalized paid access per unmetered access for the thirty days prior and subsequent to token depletion. The results clearly demonstrate that when users are faced with these increased costs for electronic access, demand for these articles plummets.

The on-line user survey we conducted in October and November 1998 provides further evidence that password use is a real non-pecuniary cost.[32] Of the respondents who had not yet obtained a password, the lack of need was cited by 70%. This percentage decreases in usage. The more frequently one uses PEAK, the likelihood of the need of a password to access an article increases, as does the willingness to bear the fixed cost of obtaining a password. Once the fixed cost of obtaining a password is borne, users report that password use is a true cost. Ninety percent of users who report password use of less than 50% attribute non-use to an issue of cost,[33] while only approximately 3% cite security concerns. Access data bolsters this finding. In 1999, only 33% of all accesses to unmetered articles were password authenticated. Clearly users generally do not use their passwords if they do not need to.

---

[31] 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.

[32] The 297 survey respondents are biased towards users who have passwords and use them often. All users were alerted to the fact that authenticated users who complete the survey would no longer be presented with and need to "click-through" the survey before subsequent authenticated sessions.

[33] These cost reasons were: password too hard to remember, lost password, and password not needed.

11   Conclusions

It is too early to draw firm conclusions from the PEAK research project: we are continuing to collect data through August 1999, and have only completed preliminary analysis of the data currently available. However, we have observed several interesting features of user behavior and the economics of access to scholarly literature:

- The innovative access model we introduced—the generalized subscription—is only feasible in an electronic environment and, apparently, was quite successful. Users at all institutions, even the largest, gained easy and fast access to a much larger body of content than they previously had from print subscriptions, and they made substantial use of this opportunity.

- The user cost of access, consisting of both monetary payments and non-pecuniary time and effort costs, had a significant effect on the number of articles that readers access. It appears that usage was increasing even after a year of service. By the end of the experiment, usage was at a rather high level: approximately five articles accessed per month per 100 potential users, with potential users defined broadly (including all undergraduate students, who rarely use scholarly articles directly). The continued increase in usage can be explained by a substantial learning curve during which users become aware of the service and accustomed to using it, as well as improvements in the underlying service over the life of the project.

- There is also a learning curve for institutions, both in terms of understanding the pricing schemes as well as their users needs. Institutions apparently made use of access data from the first year to improve their purchasing decisions for the second year.

- It has long been known that overall readership of scholarly literature is low. We have

32

seen that even the most popular articles are read only a few times, across 12 institutions. Of course, we could not simultaneously measure how often those articles were being read in print versions.

- Recency is very important: repeat article access dropped off considerably after the first month.

We will undertake more careful analyses of the data over the next year. Thus far, we think the most important findings are that access can be expanded through innovative schemes like the generalized subscription while maintaining a predictable flow of revenue to the publisher, and that non-pecuniary costs of electronic access systems can be as important as prices.

Two of the general lessons we have learned deserve special attention. First, as has been shown for usage of printed journals, usage per article is quite low. Articles that were read in PEAK were accessed less than two times each on average, summing over all twelve institutions. (Of course, most articles were not accessed at all.)

Second, the economics of access decisions by users at an institution (university, corporation, etc.) are complicated. The librarian in charge of participation made most of the purchasing decisions at our client institutions. Some access decisions involved a mixed decision: individuals paid for access, but then got reimbursed. And, perhaps most importantly, the hard-to-quantify non-pecuniary costs seemed to be as important as the prices in determining user behavior. Although other system designs might reduce some of the non-pecuniary costs, the University of Michigan has considerable experience in delivering digital library services, and in our opinion the implementation was about average in terms of user convenience.

References

Bakos, Yannis and Erik Brynjolfsson, "Bundling Information Goods: Pricing, Profits and Effi-

ciency", University of California, April 1998.

Chuang, John and Marvin Sirbu, "The Bundling and Unbundling of Information Goods: Economic Incentives for the Network Delivery of Academic Journal Articles", Presented at the Conference on Economics of Digital Information and Intellectual Property, Harvard University, January 1997.

Harr, John, `'"Project PEAK: Vanderbilt's Experience with Articles on Demand," NASIG Conference, June 1999.

King, Donald W. and Jose-Maria Griffiths, "Economic issues concerning electronic publishing and distribution of scholarly articles," Library Trends 43, no. 4, 1995, pp. 713–740.

MacKie-Mason, Jeffrey K. and Juan F. Riveros, "Economics and Electronic Access to Scholarly Information, in *The Economics of Digital Information*, (tentative title), D. Hurley, B. Kahin and H. Varian, eds., MIT Press, forthcoming 1999.

McCabe, Mark J., "Academic Journal Pricing and Market Power: A Portfolio Approach", Presented at American Economics Association Conference, Boston, January 2000.

Odlyzko, Andrew, "Tragic Loss or Good Riddance? The impending demise of Traditional Scholarly Journals", International Journal of Human-Computer Studies, Vol. 42, 1995, pp. 71–122.

Prior, Albert, "Electronic journals pricing - still in the melting pot?," UKSG 22nd Annual Conference, 4th European Serials Conference, April 12-14 April 1999.