Economics and Electronic Access to Scholarly Information

Jeffrey K. MacKie-Mason[1] and Juan F. Riveros[2]

First Version: 23 January 1997

Current Revision: 19 May 1997

Published in B. Kahin and H. Varian, eds., Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property (MIT Press: Cambridge, Mass., 2000).

Table of Contents

Economics and Electronic Access to Scholarly Information

Jeffrey K. MacKie-Mason[1] and Juan F. Riveros[2]

First Version: 23 January 1997

Current Revision: 19 May 1997

Published in B. Kahin and H. Varian, eds., Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property (MIT Press: Cambridge, Mass., 2000).

1. Introduction

Dramatic increases in the capabilities of computers and communication networks, accompanied by equally dramatic decreases in cost, have fomented revolutionary thoughts (if not the revolution itself) in the scholarly publishing community. Much attention is focused on upstart electronic-only journals, and whether they will successfully displace or competitively discipline the journals managed by traditional professional publishers. Meanwhile, the traditional publishers are developing and testing electronic access schemes to their body of literature. This paper concerns a controlled field experiment to investigate the effects of product bundling and pricing structures for electronic access to scholarly literature.

We and our colleagues at the University of Michigan are implementing the field experiment at the time of this writing. The host service team is receiving and preparing digital content. The marketing team is recruiting trial participants. We are finalizing the details of the bundling and pricing schemes, and the technical structures necessary to provide authorization and authentication, accounting, and other services are being developed. The experimental treatment and data collection have not yet begun, however.

Our primary research objective for the field trial is to generate rich empirical evidence on user behavior when faced with various bundling and pricing schemes. This field trial will complement the recent theoretical research of Bakos and Brynjolfsson (1996), and Chuang and Sirbu (1996). However, authors in the bundling literature, including these recent papers, have restricted their models to rather simplified bundling structures. Although we are severely limiting the design of our trial in order to obtain good data and testable hypotheses, for both practical and intellectual reasons our experimental design calls for bundling structures that have not yet been explored theoretically. We also are working to extend the theoretical work to our more realistic and general structures.

In this paper we first provide background on the market economics of the publishing business, and describe some important economic problems facing both research libraries and scholarly journal publishers. We then explain how the opportunity provided by electronic access to use innovative

bundling and pricing structures offer some hope of easing the library and publisher problems. New bundling and pricing schemes can uncover new value from both old and new scholarly content. In the next section we describe our utility-theoretic framework for analyzing the consumer response to the trial conditions, and then briefly survey some related literature on bundling and nonlinear pricing. In the final section, we describe the main features of the experiment, and the rationale behind the design.

2. Economics and Electronic Publishing

2.1 Publishing economics

Publishing is a value-adding business. This point has been underemphasized by some recent authors. Utopians have suggested that by putting technology in the hands of authors, for-profit publishers can and inevitably will be bypassed.[3] Electronic scholarly journals will arise that are edited, produced, marketed, and distributed by scholars working in the service of scholarship, rather than by professional publishers in the service of profit (e.g., Harnad 1996; Odlyzko 1995).

This striking view conflates two issues: whether or not publishing as a business adds value to authorship, and the industrial organization of the business of publishing. Of course, the way in which value is added, and the structure of costs underlying that value may determine the efficient and sustainable organization of the industry. Nonetheless, sources of value and industrial organization must be examined separately when technology, costs and service offerings are drastically changing.

Publishing adds significant value to authoring. The digital revolution is changing sources and amounts of value added, but not eliminating it. The costs of some functions are decreasing rapidly, but other costs are not. Meanwhile, as new information services are developed, new opportunities for publishing value-added are also created.

Publishers provide many services to authors and readers. For example, they perform copyediting, proofing, typesetting and layout, printing and binding. The publisher chooses and controls the quality of the medium (whether paper stock, SGML tagging, &c.), and production of offprints and supplemental bundles (e.g., CD-ROMS). The publisher also handles fulfillment: billing, accounting, and distribution. Even electronic journals require distribution skill: server administration, backup maintenance, network management. And even electronic journals have costs that must be covered through some mechanism: publishers specialize in cost management and recovery.[4]

Perhaps most importantly, the publisher brings readers to the author: the marketing function. It is quite evident that scholarly journal authors are not writing for direct cash compensation, but to obtain readership (from which indirect compensation may follow). Good scholars are good at research, not at finding readers.[5]

There are at least two implications from the considerable value publishing adds to authorship:

- There is likely a role for an independent publishing industry; and,
- Competition should ensure efficient provision of publishing services, with only normal (not monopolistic) profits over time.

We expect to see independent publishing because there is enough value-added separate from authorship itself that specialization will continue to be efficient: authors aren't the most fit to provide publishing services. If transactions costs between authors and publishers were high, then we might expect integration into single organizations, but there is no obvious reason to think such transactions costs are important. There are some clear economies of scale that publishers can achieve, however (such as in the employment of typesetters, copy editors and printers), which are an advantage for large, multi-journal publishers.

However, we believe that our first claim is not as important (nor as compelling) as the second: whatever organizational form the industry takes, competition should lead to efficient provision of publishing services. Those providing publishing services should not be able to earn above-normal profits over time.[6] The argument here is quite fundamental, and well-understood in the economics of industry: there are low entry barriers and few proprietary advantages to participation in publishing. When entry barriers are low, above-normal profits will attract entry, and entry will continue until the ensuing competition is sufficient to drive returns down to a normal rate.

One type of entry barrier is proprietary control over distinctive intellectual property. It might seem that control over intellectual property is a prevalent characteristic of scholarly publishing: authors' work, over which the publisher usually has copyright, is unique and valuable. Clearly different research articles are valued differently, or we wouldn't see such wide variation in prices and reputations between journals within similar fields and with similar production quality. However, the key to understanding the steady state performance of the publishing industry is to recognize that publishers don't *initially* control content: authors do. Publishers who are currently more successful at bringing readers to authors might have an advantage in bargaining for the best new content, but they do not have control. If the publisher charges too much to readers, or degrades production quality, then another publisher can easily step up and offer authors a journal that will be more accessible or more appealing to readers.

Of course, reputations rarely depreciate overnight. Thus, a publisher who produces a journal that is considered the best in its field is likely to maintain its subscription base and readership for at least some time. However, academic editorships turn over relatively frequently, and publishers have to compete to attract new editors, who will pay attention to quality and pricing of the journal. Likewise, libraries and other subscribers do not mindlessly subscribe to every journal. Indeed, there are very few journals, even among the very best, to which *every* potential subscriber subscribes. There are always marginal purchasers who are willing to forego a subscription if the price is a bit too high or the production quality a bit too low.

The point is simple: there is little unique or proprietary that publishers bring to the table that cannot be well imitated by others. A publisher that is lazy (quality degrades) or fat (profits are abnormally high for a sustained period) will find effective, lean and hungry competitors luring away their academic editors, pool of talented authors, and readership.

The natural competitiveness of the market for publisher services is quite important. First, it is a well-known result in economics that if the market participants behave competitively, then the outcome will maximize *consumer plus producer surplus*.[7] Surplus is the excess between value received and price paid. The consumers in the market for publishing services are both readers and

authors; the producers are the providers of publishing services. However, competition also ensures that the producers will not receive *excess profits*. That is, producers will generally only receive enough to cover their costs (including a competitive rate of return on invested capital). The rest of the maximized surplus will be enjoyed by readers and authors. A competitive market also has the property that it is *Pareto efficient*, which means that no alternative allocation of goods and services can make at least one agent better off without making at least one agent worse off.

In more colloquial terms, the performance of a competitive market in publishing services is quite attractive: readers get the best combination of quality, price and quantity possible given the technology and costs of production, and authors get the best readership possible. When new sources of value arise (such as searching over a digital document database) much of the value ultimately accrues to consumers. Of course, at any given moment, a market may not be performing optimally, but in the absence of structural problems (e.g., barriers to entry and externalities) we have a lot of evidence that competitive markets perform very well on average, over time. Many readers will no doubt wonder whether this optimistic view is justified, given apparent problems for readers and authors in the scholarly publishing business. We now discuss some of these problems.

2.2 Electronic access and research libraries

Research libraries have faced a number of recent difficulties. They are often located in institutions of higher education, which have experienced severe budget pressures. In many cases, centrally-provided infrastructure services -- such as libraries provide -- have borne more than their share of budget reductions. Concurrent with budget reductions, libraries have faced rapid increases in subscription prices, with some publishers recently increasing prices more than 10% in a year even after adjusting for inflation. The increasing demand for new services based on digital and network technology have perhaps making matters worse (although also creating excitement for information professionals), as they call upon already overtaxed human and financial resources.

Given this rather dismal set of problems, we should consider what opportunities might be provided by electronic access to scholarly materials. The first effect that many have hoped for is a reduction in service cost. Unfortunately, although many costs are lower when distribution is electronic, these costs tend to represent only a modest fraction of the total costs of publishing. In any case, even if there are significant cost savings to be made, the question is not very well posed. Electronic access changes both the profile of services available, and their quality (which in some cases increases, but in others decreases). Any attempt to compare the costs of paper and electronic publishing must carefully account for these differences in quality and the services provided: the classic "apples and oranges" problem.

2.3 Electronic access and publishers

The fundamental problem facing publishers is clear: high first copy costs. Odlyzko (1995) reports that it costs $900 - $8700 to publish a single math article, with a median of $4000. Of this, 70% is editorial and production first-copy cost: that is, the cost before reproduction and distribution. Thus, most of the the cost to be recovered by a going concern is fixed.

Pricing at marginal cost will not recover first-copy costs. Competition in publishing, however, creates pressure to price at marginal cost. When most publishers are printing on paper, and thus have similar cost structures, an uneasy equilibrium with prices above marginal cost appears to be stable, possibly supported by significant lags in the movement of editors and journal content between publishers in response to competitive pricing changes. However, the advent of electronic publishing brings a medium with potentially lower first-copy costs. In the U.S. telecom industry, local access firms face similar situation when "bypass" operators who need not pay the fixed costs of universal service are able to offer large customers lower rates for connection to long distance networks. Likewise, in publishing there is an increasing threat from"bypass publishing".[8] Increased electronic competition pressures print-on-paper publishers to find new ways to recover first-copy costs without loading them all onto per copy (or per subscription) pricing.

In part, print-on-paper publishers respond to electronic competition by themselves seeking to develop electronic delivery media. They also are investing to develop various new value-added services based on the investment in first-copy scholarly literature creation, in order to share the recovery of fixed costs across more activities. However, these incremental activities involve development expenditures, and also increase the publisher's risk since the new products are often untested and of uncertain value.

2.4 Opportunities with electronic access: bundling and nonlinear pricing

Although libraries and publishers both face significant problems, some of them the result of developments in electronic access, such access also creates opportunities for dealing with those problems. Electronic access enables both publishers and libraries to engage in new product bundling and nonlinear pricing schemes. The first will often involve unbundling of traditional journal components, and then rebundling in a greater variety of packages, some of them customized for customers, or customizable by them. Nonlinear pricing is facilitated by lower transactions costs for fine-grained purchases, and feasible direct usage monitoring by the publisher or library.

New bundling and pricing schemes are enabling technologies: they liberate previously unrealized sources of value from existing content and from new value-added services. For example, bundling can do a better job of extracting revenue from users who value the same content differently. Nonlinear pricing can "sort" users by their preference for new services, and extract more of their differing values. Publishers earn returns on their innovations that can offset fixed publishing costs, and by spreading those fixed costs over more streams of revenue, can recover them with less quantity distortion and vulnerability to competition. In a competitive publishing environment, consumers get to keep most or all of the newly liberated surplus (excess of value over cost).

An example of finding new value from old content would be differential charges to various user types for access to traditional content. Differential charging might be hard to implement for print-on-paper publication because there is little opportunity to observe actual differences in usage. With electronic access, however, it can be possible to distinguish between, say, retrieval of abstracts or bibliographic records, versus retrieval of full text. Since users typically value different uses of content differently, a system of differential charges to recover cost of service can extract more value from a heterogeneous set of uses. There are a variety of different schemes that fall under the general

rubrics of *price discrimination* or *nonlinear pricing*. The general result in the economics literature is that when it is possible to distinguish between different user or usage types, and when resale between parties or uses (*arbitrage*) is costly or preventable, then nonlinear pricing extracts more value than does uniform pricing.

The simplest example of nonlinear pricing is *perfect price discrimination*. Imagine that it is possible for the service provider to perfectly observe exactly how much every different transaction is worth to every user, and to charge accordingly. Obviously, the service provider is then able to extract the maximum possible revenue to recover costs and support new services. In general, such perfect differentiation is not possible. The key to differential pricing is to find observable characteristics (of users, or their uses of content) that are *correlated* with their willingness to pay, and then to base charges on the value information that is partially revealed. Electronic access provides an opportunity to observe a variety of such characteristics, such as user type (education, business, subscriber or occasional, determined from authentication records); use type (immediacy of access to current information, volume, full documents or components); and quality (display resolution; text-only or images; plain text or formatted).

The second type of opportunity is to extract value from *new service provision*. The literature is full of new services that may be enabled by electronic access. We will summarize a few here just to illustrate the opportunities:

- Hyperlinks. It is possible to prepare documents with embedded hypertext links to referenced material. The electronic version of this article is a simple example: references within the document can be followed quickly; for example, from the table of contents to section text. When more documents are available in electronic archives, it will be possible to embed links to more external references as well, as we have done where possible. For example, some research working papers archived by the Economics Working Papers Archive project embed references to other papers stored in the archive, making it possible to simply click on a reference and retrieve a copy.
- Dynamic Commentary. Many authors have puzzled about the interaction between electronic publishing and peer review. The main conflict observed is that the peer review process in most disciplines is quite lengthy, and defeats some of the publication timeliness that can be gained by electronic publication and distribution. One novel theme that has emerged is that the peer review process may be replaced in part by dynamic, public commentary and response. [Harnad 1990, Harnad 1996] Readers could post comments and critiques directly to the archive where an article is stored; authors could reply. In response to some comments, revised versions of a paper might be posted. Thereby, the review process might be more open, inclusive, timely and dynamic (see, e.g., *Psycoloquy*).

- Social Filtering. One of the obvious impacts of widespread communication networks is that the geography of community can change. One opportunity provided by *virtual communities* is the application of networks of like-minded people to the problem of information filtering. Every scholar faces the problem of selecting which articles to read from the vast flow of new material.[9] Social filtering systems collect the ratings of networked users, and then based on some form of cluster analysis dynamically match one user's preferences to the preferences of others to prepare recommendations. For example, Prof. X is matched to 10 other professors

who have given similar ratings on articles they have mostly all ready; the system then recommends to Prof. X a previously unread article that the match group has liked. Such a system might, for example, supplement asking only one's local colleagues for advice. Firefly (tm) is a working example of a social filtering system (providing recommendations on audio recordings and films).

3. Bundling and Nonlinear Pricing

3.1 Demand Model

In this project we will implement an experiment on electronic access product bundles and nonlinear prices. We wish to learn something about the extra value that can be extracted from existing content. To estimate the responsiveness of consumer demand to new bundling and pricing schemes we specify a general utility-theoretic model of article demand. This model will form the basis for our econometric estimates once the experimental data are collected. We also review some of the bundling and nonlinear pricing literatures as they relate to the questions raised by the experiment, and describe gaps in the theoretical literature that we hope to fill.

We expect the share of income spent on journal articles to be a small fraction of total income. Further, price and quantity data for demand on other goods will probably not be available. Therefore, we will adopt two-stage budgeting: the consumer first allocates expenditure across broad budget categories (e.g. journal expenditure), then determines budget shares for the commodities within each group.

Two-stage budgeting requires the first-stage utility function to be weakly separable,

(1) $u(y,x) = U(T_y(y), T_x(x,t))$

where x is the vector of journal articles demanded, y is the vector of demand on other goods, t is a vector of types and U is an increasing function of T_y and T_x. With two commodity groups, weak separability imposes hardly any restrictions, other than requiring that article demand and "other expenditure" be substitutes (MacKie-Mason and Lawson, 1993).

Gorman (1959) derives the conditions under which is possible to solve the first-stage problem without knowledge of all of the individual prices. One sufficient condition for top-level budgeting with composite commodities and group price indices is homotheticity of the group subutility functions. This implies that every commodity within a group must have the same income elasticity. Unfortunately this contradicts all known household budget studies (Deaton and Muellbauer, 1980).

A less restrictive solution to the problem is that the group indirect utility functions take the generalized Gorman Polar Form (GGPF):

(2) $V_g (M_g, P_g) = F_g [M_g / B_g(P_g)] + A_g(P_g)$

for some monotone in increasing function $F_g(\ )$, combined with an additively separable top-level utility function:

(3) $U = T_1(Q_1) + T_2(Q_2) + \ldots + T_n(Q_n)$

with g = 1,...,G indexing groups.

Under this formulation, the top level utility function for journal articles and other goods has the following form:

(4) $U = T_y(y) + T_x(x,t)$.

The GGPF does not impose strong restrictions on the shape of preferences within commodity groups and provides necessary and sufficient conditions for broad group allocation within the framework of weak separability. This is the most general formulation possible for the analysis of the experimental data in this project.

3.2 Product Bundling

As discussed below, we have selected three different bundle types as the products for the experiment. In this section we review the bundling literature as it applies to each of the products.

Most of the literature on bundling has focused on the case of two goods. Recent research in the area has extended some of the results to more general cases and provided a theoretical framework to analyze the bundling of information goods.

The literature on bundling has concentrated on the demand side incentives that makes it a profitable strategy. Examples of these incentives include: negative correlations in consumer valuations, complementarity in consumption, and uncertainty in the valuations of the quality of the goods (Hanson and Martin, 1990). On the production side, various cost efficiencies may also provide a basis for bundling.

Most of the bundling literature has focused on the two good case. Stigler (1963) studies commodity bundling strategy for movie distribution. A distributor could increase profits by bundling movies in packages when reservation prices for individual movies are negatively correlated. Adams and Yellen (1976) determine when it is profitable for a firm to bundle, or not. Using a graphical tool they generalize Stigler's example for two goods. They identify three strategies: pure bundling when goods are sold only in package form; mixed bundling when the goods are sold separately as well as in packages; and, pure unbundling or component selling. They compare these different strategies using examples with a discrete number of customers. When the reservation values for the elements of the bundle are negatively correlated, bundling can serve the purpose of sorting customers into groups of different characteristics. This allows the firm to extract additional consumer surplus.

Schmalensee (1982) studies whether a monopolist can profitably bundle its good with a good provided by a perfectly competitive market. Schmalensee (1984), assuming a bivariate normal distribution for individual demands, shows that bundling can be also profitable when demands are

uncorrelated or even positively but not perfectly correlated. By adding goods, the variance of the consumer s valuation for the bundle is reduced, allowing the seller to extract more surplus.

McAfee, McMillan and Whinston (1989) show that mixed bundling is always (weakly) better than pure bundling. A more interesting question posed in their study is whether mixed bundling dominates component selling. They consider two cases: the monopolist can and cannot monitor purchases. When purchases can be monitored, the monopolist can charge more for the bundle than the sum of the component prices. Without monitoring, the bundle price is constrained to be no more than the sum of the component prices. McAfee et al. show that mixed bundling almost always dominates pure unbundling when the seller can monitor. They also derive conditions under which mixed bundling dominates pure unbundling when monitoring is not possible.

Salinger (1995) analyzes the profitability and welfare consequences of bundling. When bundling does not lower cost, it tends to be profitable when reservation prices are negatively correlated and high relative to marginal costs. However, when bundling lowers cost, the incentive to bundle increase when reservation prices are positively correlated and costs are high relative to average reservation values.

The bundling problem becomes increasingly complex as we depart from the two-good formulation. Hanson and Martin (1990) find the optimal bundling strategy by formulating the question as a mixed integer linear program. Their solution method is tested on bundles with up to 21 components. Nevertheless, a typical academic journal contains around 100 articles per subscription bundle. This requires making simplifying assumptions to make the problem analytically tractable.

Bakos and Brynjolfsson (1996) show that pure bundling of zero marginal cost goods with independent and identically distributed valuations dominates pure unbundling. This approach reduces the average deadweight loss and increases profits as the number of goods in the bundle increases. They also analyze the consequences of relaxing these assumptions.

Chuang and Sirbu (1996) address the problem by assuming that all articles have identical marginal costs and are priced identically. This reduces the calculation from 2^n optimal prices to just 2 (a per article price and a subscription price). Their analysis shows that mixed bundling is the dominant strategy: "By offering a menu choice which includes both the original and the components, the producer can extract consumer surplus more completely via consumer self-selection." They also derive conditions under which pure unbundling outperform pure bundling, even when bundling reduces costs. They conclude that more bundling is not always better for the producer.

Chuang and Sirbu model consumer heterogeneity in two dimensions. Each consumer ranks the N articles in the journal in decreasing order of preference. A linear demand function is assumed for all positive-valued articles. The consumer's valuation of the n-th article is:

(5) $W(n) = \min \{ 0, W_0 [1 - (1/K)(n/N)] \}.$

W_0 represents the willingness to pay for the user's favorite article and K indicates the fraction of articles in the journal that have a non-zero value to the individual. This linear demand equation is consistent with our general formulation. The demand for journal articles doesn't depend on income,

and it is derivable from a quasilinear utility function which is a particular case of the Generalized Gorman Polar Form.

The literature to date on bundling is not sufficiently rich to encompass the products in our field trial. We will discuss the gaps below after we detail the bundles we are implementing.

3.3 Nonlinear Pricing

As discussed below, we will introduce some nonlinear pricing scheme in the experiment. Due to the complexity of the field trial, and issues of participant acceptability, we will limit experimentation with nonlinear pricing parameters to individual, single article purchasing.

Wilson (1993) constructs the optimal multipart tariff by using a menu of optional two-part tariffs, and also characterizes the optimal fixed fee. In the general nonlinear pricing problem formulation, a monopolist seller charges P(q) for an n-vector bundle q of its products. A customer of m-vector type t has a predicted benefit: W(t) = max {U(q,t) - P(q)} (The maximum is over q in Q, where Q is the set of possible bundles). The distribution of type parameters has a density function f(t), with upper support T.

By the Revelation Principle, the monopolist can choose the vectors q(t), W(t). Wilson considered the Ramsey problem of a welfare-maximizing monopolist subject to the constraint of cost recovery. We are interested, instead, in a profit-maximizing monopolist. Thus, for us the seller's problem is to

(6) max over {q(t), W(t)} the integral from 0 to T of { (U(q(t),t) - W(t) - C(q(t))) f(t) dt}

subject to:

- W (t) = U_t(q(t),t) (incentive compatibility)
- W(t) >= U(0,t) = 0 (customer s participation constraint)
- q(t) >= 0 (feasibility constraint).

Tirole (1988) analyzes the welfare effects of nonlinear pricing. In the two-type case he shows that that, with optimal pricing, the quantity purchased by the high-demand consumers is socially optimal (the marginal utility of consumption of the good is equal to the marginal cost). If the monopolist serves both types of consumers, the quantity purchased by the low-demand consumers is suboptimal. Faced by a continuum of types, a monopolist induces consumers to purchase a suboptimal quantity. The marginal willingness to pay for the good exceeds the marginal cost, except for the highest demand consumer. These two cases assume quasilinear utility functions.

Katz (1989) shows that non-linear pricing may yield too little or too much output in comparison with the social optimum. If the single crossing condition hold, the monopolist generally produces too little output.

4. PEAK: A Field Trial

The University of Michigan has negotiated with Elsevier a pricing field trial, "Pricing Electronic Access to Knowledge" (PEAK). This negotiated trial permits Michigan to provide a host service for three years (1996-1998) of all approximately 1200 Elsevier Science scholarly titles. Once the trial is underway, articles will become available as quickly as they do for any Elsevier Electronic Subscriptions customer, which is to say within a few weeks of the mail distribution of the print-on-paper version. Michigan will provide Internet-based delivery to participants in the trial, which will include users both on campus and off.

The PEAK project will be implemented by the same team responsible for the Michigan's collaborative digital library enterprise, including its Humanities Text Initiative, the TULIP project, JSTOR, and other efforts. The underlying technologies have been utilized in the TULIP and NSF/ARPA/NASA UM Digital Library projects, among others. The University of Michigan has implemented a full, commercial digital production service, and currently hosts more than a terabyte of indexed document data on high-speed disk systems.[10]

For PEAK, Michigan will create a variety of access models, and administer a pricing system. The agreement is designed explicitly to support experimental field research, and thus there will be experimental variation in the bundles and prices offered to clients. In this section we shall describe in some detail the structure of the experiment. There are three major design components:

1. Economic design: Specification of the product bundles and the price structure.
2. Experimental design: Implementation of the economic design to obtain statistically informative data.
3. Technical implementation.

4.1 Economic Design

4.1.1 Bundling

Electronic access directly provides the opportunity to unbundle and rebundle scholarly literature. A print-on-paper journal is a bundle of issues, each of which is a bundle of articles and other items, each of which is a bundle of bibliographic information, an abstract, references, text, figures, and so forth. It is often straightforward (though not necessarily costless) to rebundle any of these elements in different ways when the source material is archived in electronic form. For example, one might obtain all abstracts matching a given keyword search; or all citations appearing in a particular article; or just the bibliographic headers from the articles appearing in a given year.

From the beginning the PEAK project was designed to explore bundling alternatives as well as pricing structures. The specification of product bundles turns out to be a quite difficult task because the space of possible bundles is extraordinarily large, even if treats each component of a document as identical across documents (e.g.., do not define articles by two different authors to be different commodities for the purpose of bundling). To see this problem, consider bundles defined over just three possible dimensions for electronically accessed scholarly documents:

- *article component* (abstract, references, text, &c.)
- *time limit on usage* (unlimited, per use, per year, &c.)
- *usage rights* (read only, read and print, &c.).

From a rather simple population of the matrix of possibilities, we can postulate 160 possible different bundles (see Figure 1), *before even specifying different bundle quantities*. For example, we might offer a one-year right to read-only N abstracts. If bundles were priced differently for, say, N={1, 10, 100, unlimited} then the space increases to 640 possible bundles.

Figure 1: Hypothetical bundling possibilities

Although an explicit goal of the field trial was to explore new opportunities afforded by electronic access, feasibility constraints led to a rather limited selection of bundle types to implement. One important constraint is experimental variation: with a somewhat limited number of different observational units we cannot obtain sufficient variation to explore very many bundle types. A second constraint is customer acceptance: although this is being operated as a research project by the University of Michigan, the experimental subjects are operational research libraries and individual users who must find the offerings palatable enough for participation to be worthwhile. After numerous meetings with library collection experts and potential participants, the project team decided on three bundles to offer:

- *Per article*. A user can purchase unlimited access to a specific article for a fixed price. This option is designed to closely mimic a traditional interlibrary loan (ILL) product. With ILL the individual receives a printed copy of the article that can be retained indefinitely. This is different from the "per use" pricing model often applied to electronic data sources. The article is retained on the PEAK server, but the user can access a paid-for article as often as desired. This is a buyer-chooses scheme, in that the buyer selects the articles before paying for them.

- *Traditional subscription*. A user or a library can purchase unlimited access to a set of articles designated as a *journal volume* by the publisher. These volumes will correspond to the Elsevier print-on-paper journal titles. Access continues at least until the end of the project. This is a sellers-chooses bundle, in that the seller (Elsevier and its editorial associates) selects which articles are delivered to the user, after the user has subscribed.
- *Generalized subscription*. A library can pre-purchase unlimited access to a set of 120 articles selected by the user. (The number of articles was chosen to correspond approximately to the average number of articles included in a traditional subscription.) Individual users can purchase personal blocks of *N* articles (where *N* is less than 120). This is a buyer-chooses bundle. In every regard is the same as a traditional subscription, except that the buyer selects which articles are accessed, from across all Elsevier titles, after the user has subscribed. This bundling approach allows users to capture value from the entire corpus of articles, without having to subscribe to all of the journal titles. This opportunity is justified by the approximately zero incremental cost of delivering additional articles once the server database is constructed.

The time duration for each of the product bundles is the life of the project plus one year.[11] The usage rights for a document will generally be the same as those for an Elsevier article distributed as print-on-paper (e.g., U.S. users can print a copy for personal archival purposes, but may not redistribute it further except as provided under the fair use doctrine).

The first and second products are intended to match as closely as possible the best known traditional product bundles.[12] This conservatism should assist in customer acceptance of the trial, and should also allow better comparisons between electronic and print-on-paper experience. The third option is more novel. The generalized subscription is designed to test the opportunity provided by unbundling and rebundling without straying too far from the other product bundles. Limiting product variations to only one dimension will improve statistical inference about the effects of the different bundling schemes.

Although the planned product bundles do not deviate far from traditional bundles, they do permit some interesting user choices. In particular, generalized subscriptions allow users to prepay and thus obtain the administrative advantage of budget predictability, yet also allow users the flexibility to choose only articles they want. Indeed, users can bundle articles across journals, and even across traditional disciplines. This shares the risk of paying for unwanted articles with the publisher, Consequently, the value per article accessed should generally be higher to the user than under a traditional subscription, and the per article average price can be correspondingly higher for generalized subscriptions. In short, users trade off an upfront commitment against the benefits of greater customization, and the ability to select articles post-publication. Publishers gain by having a significant revenue stream that is predictable, and also by direct feedback on which articles, from across several journals, readers most highly value.

For the experiment, we will vary the mixed bundling opportunities available to users. Some institutions will be offered only traditional subscriptions plus per article purchase; others will be offered only generalized subscriptions plus per article purchase. Yet a third group will be offered the opportunity to construct a portfolio from all three of traditional subscriptions, generalized subscriptions and per article purchase.

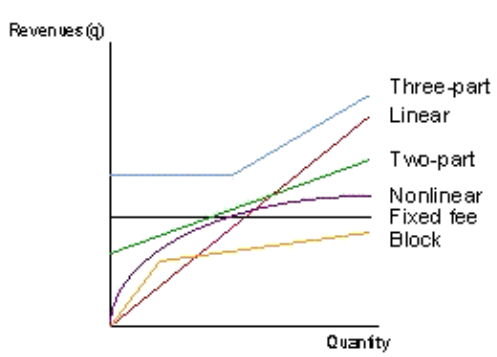
Our products are more complex than those studied thus far in the bundling literature. The differences are largely due to our attempt to maintain plausible comparability to existing print-on-paper offerings, while still designing a controlled experiment. For example, our "mixed bundling" is more general than the form studied in the literature. Customers do not need to choose just subscriptions, or alternatively just individual articles. Our customers might purchase some subscriptions and supplement them with per article purchases. This feature highlights the other necessary complexity: the fact that subscriptions are actually "sub-bundles". In the Chuang and Sirbu (1996) and Bakos and Brynjolfsson (1996) papers, only one bundle at a time is considered, which consists of either all or a subset of all articles. In practice, publishers offer several bundles, each as a different journal title. Sizable publishers never limit their bundles to a single, all-article bundle. Thus, the value of per article and mixed bundling to the provider will necessarily be conditional on the particular set of sub-bundles (journal titles) the publisher offers.[13]

We are focusing on another difference between our formulation and the standard modeling in the bundling literature: uncertainty. With a traditional subscription, the buyer does not know the quality of the bundle before making the purchase decision, but may know the distribution of past quality. With a generalized subscription, the buyer does not know the quality of the individual articles before deciding how many bundles to purchase, but then chooses the articles to include after their quality has been (at least partially) revealed. With per article pricing, the buyer knows the quality before committing to any purchases. Therefore, we expect consumer demand for the different bundles to depend on risk preferences and expectations of the heterogeneity of perceived article quality. We are formulating a formal model of consumer demand that incorporates uncertainty, upon which to base our econometric analysis of the trial. We believe that the ease of offering a variety of different risk-sharing schemes may be an important advantage for electronic access.

4.1.2 Pricing Structures

The variety of nonlinear pricing structures is essentially unbounded: any nonlinear, monotonically increasing relationship between total payment and quantity (of something) purchased is admissible. Although a field trial offers some possibility to explore the space of pricing structures, we again face feasibility and participation constraints. In particular, because we are also varying product bundles, the number of different pricing schemes that we can implement and still obtain sufficient data for statistical inference is very limited.

Within the class of general nonlinear pricing schemes a large number have been specifically studied and implemented in various industries. These are illustrated in Figure 2, which shows the relationship between revenues and quantity for a linear scheme, second-degree price discrimination (block pricing), a two-part and a three-part tariff, a fixed fee, and a generic nonlinear scheme.



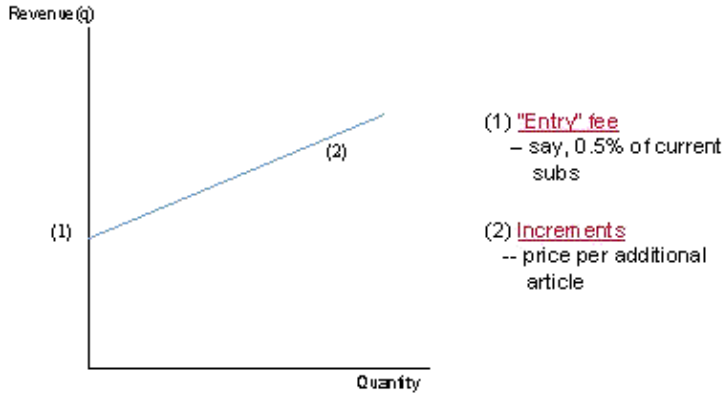


Figure 2: Various nonlinear pricing schemes **Figure 3: Two-Part Tariff Explained**

The project team plans to implement two-part tariff pricing (see Figure 3). There is an *entry fee*, or initial payment merely to participate. Incremental access is charged at a flat per unit price.

4.2 Experimental Design

The PEAK field trial provides an unprecedented opportunity for exploring bundling and nonlinear pricing opportunities afforded by electronic access. The key research objective, however, is generalizability to other user populations and other scholarly materials. Experimental design considerations are therefore paramount, and impose significant constraints on the range of bundling and pricing schemes than can be studied. The fundamental principle of experimental design is that we learn from variation: how do users respond differently when confronted with different alternative bundles or price structures? To obtain a sufficiently rich set of observations to draw inferences from the effects of design parameter variation, we must significantly restrict the number of dimensions that vary. These concerns were reflected in the discussion of bundles and price structures above. There are a number of other interesting features of the experimental design that we describe below.

1. *Product diversity*. Elsevier is making available approximately 1200 different journal titles for the trial. This corpus includes titles in virtually every academic discipline. The style, quality, price and other features of these journals cover a very wide range. Further, scholarly usage styles appear to differ significantly across disciplines. Therefore, product diversity potentially creates a source of difficult-to-control nonexperimental variation. To limit this problem, the experimental design calls for limiting the trial to a few disciplines in which Elsevier titles have an especially strong presence: engineering, medical science and economics.[14]
2. *Customer diversity*. Scholarly journals have narrow, specialized audiences. Even within a community as large as the University of Michigan (some 30,000 students, of whom about 10,000 are in graduate school, and about 3200 research faculty), there are likely to be only a few readers for many speciality journals. Only a fraction of these are likely to participate in an electronic access trial. Therefore, in order to obtain sufficient participation and usage, the project was designed to include clients from a number of organizations outside the University of Michigan. The preliminary participant list includes other research universities and industrial research facilities. There may also be participating four-year undergraduate colleges. Although the sample size of users for each subset of literature is increased by

including multiple institutions, broader scope also creates design problems. There will be greater diversity in physical facilities, marketing and communications to participants, and institutional arrangements. The latter is of special concern. Participation in the trial will occur both by institutional and individual decisionmakers. For example, both will be able to subscribe, and both will be able to purchase on a per article basis. The behavior of a rational and well-informed individual user will be conditioned on the behavior by her associated institution.[15] For example, an individual might subscribe to a journal only if the organization's library does not. This complication is probably unavoidable, and will pose challenges for the proper statistical analysis of the results.

3. *Duration and learning*. This trial involves a rather novel access mode for scholarly literature. Although Michigan has substantial institutional experience from TULIP, JSTOR, UMDL and other projects, most individual users of scholarly literature on campus have little or no experience with electronic access to traditional scholarly journals. Users at the other participating institutions -- both individual and institutional -- are likely to be even less experienced. The learning process will complicate efforts to uncover generalizable results. The project team will actively educate potential users about the trial, the products and the pricing. Further, we will be collecting data over a two or three year period, and thus will have some chance to isolate learning effects.

Although not really part of the experimental design, there are also a number of technical issues that must be overcome to complete the implementation of the PEAK project. These include the implementation of an accounting system and a payment mechanism, and user authentication to the access system. As much as possible we are relying on off-the-shelf or already implemented local solutions. For example, we have negotiated an agreement with First Virtual, Inc., for bulk provision of user IDs that our client library participants can then redistribute to their individual users. Users need make only one automated phone call to authorize First Virtual to post charges to their credit card, and send one email message to link their user ID and their email address. They then have a fully-refutable and secure method for purchasing bundles and individual articles beyond those paid for by their institutional library.

5. Conclusion

It is hard to assert a conclusion when the data have not yet been collected, nor analyzed. We offer two significant observations based on our experience and theoretical research to date:

- The space for electronic access product bundling and pricing structures is immense. Many field trials will be required to explore even a limited set of important alternatives.
- The literature on bundling, and to a lesser extent nonlinear pricing, has not yet developed to the point where the models fit very well to actual circumstances. We have identified some of the theoretical framework that needs further attention.

Notes

[1] Jeffrey K. MacKie-Mason is an Associate Professor of Economics and Information at the University of Michigan, Ann Arbor, Michigan, USA. Professor MacKie-Mason is the Project

Director for the economic research aspects of the PEAK project described in this article. [return to text]

[2] Juan Riveros is a Ph.D. student in the Department of Economics, University of Michigan. [return to text]

[3] Kling and Lamb (1996) offer an insightful discussion of utopian discourse on information technology and electronic publishing. [return to text]

[4] See Fisher (1996) for a nice overview of the value-added functions provided by publishers. [return to text]

[5] Hayes (1996) nicely characterizes the publisher's role in bringing readers to authors. [return to text]

[6] "Normal" profits are the ordinary rate of return, adjusted for risk, that an investor could obtain in a competitive financial instrument (such as a government bond). It is a return for the use of capital in the enterprise, necessary since the capital could be elsewhere invested to obtain a normal return. [return to text]

[7] The competitive equilibrium result assumes that there are no externalities. This means that the amount of publishing services received by any one agent does not directly affect how well off is any other agent. A typical externality is water pollution: the more that I produce, the worse off are downstream residents. [return to text]

[8] Some publishers are sufficiently concerned with the effect of electronic distribution that they are treating posting of drafts and working papers on the Internet as prior publication, and are refusing to consider such articles for their journals. Most American Psychological Association journal editors have adopted this policy, as have *Neuroscience* and *The New England Journal of Medicine*. [return to text]

[9] Odlyzko (1995) points out that about half of all scholarly literature ever published has been published in just the last 10 years. [return to text]

[10] The University of Michigan was also one of the original host sites and developers of TULIP (The University Licensing Project), sponsored by Elsevier Science. [return to text]

[11] The University of Michigan and Elsevier anticipate that access to paid-for documents and bundles might persist beyond the life of the original project agreement, but that is not guaranteed. [return to text]

[12] Pay per article is traditionally available through at least three vehicles: by paying a per article royalty to the Copyright Clearance Center; by purchasing an article from a document delivery service (which in turn remits royalties); or, when fair use is invoked, by paying the cost of photocopying. [return to text]

[13] Of course, the selection of the number of sub-bundles, or journal titles, and the guidelines for their specialty content, are endogenous. At a given point in time mixed bundling might be superior to "traditional subscription only" because reader tastes are sufficiently aligned with journal boundaries. If the publisher responds to this mismatch, the relative value of traditional subscriptions may increase. [return to text]

[14] In 1996, the University of Michigan subscribed to approximately 69 engineering titles at an average cost of about $1000/title; 13 economics titles (~$900/title); and 134 medical titles (~$1200/title). [return to text]

[15] The causality may also run in the opposite direction, of course, but we expect that during the relatively short duration of this trial the purchasing decisions of the institution will likely be relatively predetermined with respect to individual decisions. Over a greater time horizon, the chance that library purchasing would respond to individual choices would likely increase. [return to text]

References

Adams, W. and Yellen, J. (1976) "Commodity Bundling and the Burden of Monopoly", *Quarterly Journal of Economics*, 90, 475-498

Bakos, Y. and E. Brynjolfsson (1996) "Bundling Information Goods: Pricing, Profits and Efficiency", University of California, Irvine Working Paper, 1-30. Presented at the *Conference on Economics of Digital Information and Intellectual Property*, Harvard University, January 1997.

Chuang, J. and M. Sirbu (1996) "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.

Deaton, A. and John Muellbauer (1980). *Economics and Consumer Behavior*. Cambridge University Press. Cambridge.

Fisher, Janet (1996). "Traditional Publishers and Electronic Journals," in *Scholarly Publishing: The Electronic Frontier*, R. Peek and G. Newby, eds. (Cambridge: MIT Press).

Gorman, W. M. (1959), "Separable utility and aggregation", *Econometrica*, 27, 469-481.

Hanson, W. and R. Martin (1990) "Optimal Bundle Pricing", *Management Science*, 36, 155-174.

Harnad, Stevan (1990). "Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry." *Psychological Science* 1: 342 - 343.

Harnad, Stevan (1996). "Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals," in *Scholarly Publishing: The Electronic Frontier*, R. Peek and G. Newby, eds. (Cambridge: MIT Press).

Hayes, Brian (1996). "The Economic Quandary of the Network Publisher", in *Scholarly Publishing: The Electronic Frontier*, R. Peek and G. Newby, eds. (Cambridge: MIT Press).

Katz, M. (1983). "Nonuniform Pricing, Output and Welfare under Monopoly", *Review of Economic Studies*. 50: 37-56.

Kling, Rob and Roberta Lamb (1996). "Analyzing Alternate Visions of Electronic Publishing and Digital Libraries", in *Scholarly Publishing: The Electronic Frontier*, R. Peek and G. Newby, eds. (Cambridge: MIT Press).

MacKie-Mason, J. and D. Lawson (1993) "Local Telephone Calling Demand Whan Customers Face Optional and Nonlinear Price Schedules", Working Paper, 1- 33.

MacKie-Mason, J. and J. F. Riveros (1997). "Economics and Electronic Access to Scholarly Information", Presented at the *Conference on Economics of Digital Information and Intellectual Property*, Harvard University, January 1997.

McAfee, R., McMillan, J. and Whinston, M. (1989) "Multiproduct Monopoly, Commodity Bundling and Correlation of Values", *Quarterly Journal of Economics*, 114, 371-384.

Odlyzko, Andrew (1995). "Tragic Loss or Good Riddance? The Impending Demise of Traditional Scholarly Journals", *International Journal of Human-Computer Studies*, 42: 71-122.

Salinger, M.A. (1995), "A Graphical Analysis of Bundling", *Journal of Business*, 68: 85-98.

Schmalensee, R (1982). "Commodity Bundling by a Single Product Monopolist", *Journal of Law and Economics*. 25: 67-71

Schmalensee, R (1984). "Gaussian Demand and Commodity Bundling", *Journal of Business*. 57: S211-S230

Stigler, G. (1963), "United States vs. Loew's Inc. : a note on block booking", *Supreme Court Review*, 152-157.

Tirole, J., (1988) *The Theory of Industrial Organization*, MIT Press, Cambridge, MA.

Varian, H. (1992), *Microeconomic Analysis*. Norton. New York.

Wilson, R. (1993) *Nonlinear Pricing*. Oxford University Press. New York.

Zahray, W. and M. Sirbu (1990) "The Provision of Scholarly Journals by Libraries via Electronic Technologies: An Economic Analysis", *Information Economics and Policy*, 4, 127-154.