Ownership and pricing of information: a model and application to open access

James D. Campbell*
University of Toronto Scarborough and Brown University

October 2, 2015

Abstract

We consider a spillover externality in the use of information. When a person consumes an information source, she enjoys private benefit but also creates social value. Her use of the source enhances society's understanding of it, which is 'in the air' and freely accessible by others, but which decays in quality over time. We ask how efficiently information will be used in the presence of this externality effect under four cases: exogenously given prices, ownership and pricing by a profit-maximizing monopolist, ownership and pricing by a firm restricted by competition or mandate to earn zero profit, and open access. Information is overpriced and underused under both monopoly and zero-profit ownership, and underpriced and overused under open access. However, as the cost falls toward zero, outcomes under zero-profit ownership and open access both tend toward the efficient level, and for-profit monopoly ownership is increasingly inefficient. Potential efficiency gains from open access and zero-profit ownership are therefore greater as advances in technology reduce the costs of delivering information to consumers.

JEL: D40, D62, D8, L11

Keywords: information goods, spillovers, intergenerational externality, pricing, open access

*James_D_Campbell@brown.edu

1 Introduction

In many situations we produce output that refers to previous work done by others. In this paper we consider a spillover externality that can arise in the process of consuming an information source to produce output. An individual bears the cost of accessing and using a piece of information, but by using it they not only enjoy the private value of understanding the source but may also create social value by enhancing others' understanding of the source. We present a model that captures this externality, and we apply the model to the question of how we should store, price, and assign property rights to information.

We will consider three broad classes of regime for the ownership of information, and compare the use and pricing of information under each regime to the socially efficient outcome. First, we consider proprietary ownership by a single entity who can make profit. Second, we consider proprietary ownership by a firm that is restricted to earn zero profit, either because they compete with other owners of the information, are a not-for-profit institution, or operate under regulatory enforcement of a zero profit condition. Third, we consider the case in which the information source is freely available to anyone, for example through some repository under an open or online access policy.

In the model consumers arrive one by one in sequence. Each consumer would like to understand an information source, and can choose from two options. The first option is to access and use the original piece of information at a cost, either by paying a price to access the source or having paid for subscription access to it. The second option is to rely on a freely available but imperfect understanding of the original source that is 'in the air'. The tradeoff for a consumer is therefore the cost of accessing the original information source versus the lower quality of the freely available information about the source. The key feature of the model is that the quality of information in the air degrades over time: the more consumers in a row who choose the freely available information rather than accessing the original source, the less a subsequent consumer will learn from the freely available information. This general idea of decay in knowledge of the original source could reflect, for example, such specific examples as received wisdom being gradually forgotten as generations pass, a hard copy of information degrading over time, or congestion in learning when there are more learners per teacher. An analogy for the decay concept is the telephone game: a phrase is whispered down a line of people, but as it passes along it becomes garbled, so that the phrase heard by the person at the end of the line may not be at all like

the phrase at the beginning.

In this model, is information used efficiently, or it is used too much or too little? Does this depend on who owns and prices the information? Does it depend on whether information is sold for a price per use or under a library subscription setup? How does it depend on the costs incurred by those who store and deliver the information? Consider one concrete example of the idea of the model as a whole. An agency produces a report on a country by researching and referring to original sources, incurring some costs. At this point the agency knows as much about the country as it ever will. Occasionally a need arises for a firm or organization to know something about the country to inform a course of action. The organization can access the agency's report, paying some price or subscription fee, to learn something about the country. Over time the information in the report grows dated and obsolete—it decays—so that when it is consulted it is less useful than it was before. At what point will the agency produce a new report? Does it produce too few or too many? Does it matter whether the agency is public or private?

Later in the paper we will consider an application of the model to the question of whether scientific research should be published by for-profit copyright holders, or instead published under an open access policy or made freely available online. In this interpretation of the model, scientists produce original research. Another researcher, a practitioner, or a member of an institution may seek to use this research to address a problem or inform their own work. If they or their institution subscribe to a journal or database that includes the research, or are willing to pay to access it, or it is available freely online, they can read the research directly. If not, they can rely on the understanding of the research that is in the air as a result of others having used it, perhaps by talking to colleagues or referring to a simplified textbook explanation. The fewer people read the original, the worse the quality of information that is in the air. Should government mandate that publicly-funded research be published freely online or under open access rules? Would the use of information under this regime be more or less efficient than if the information is published for profit?

Since the general idea of the model can be applied to a variety of settings, let us briefly consider some other concrete examples:

Example 1: citations

A researcher is producing a paper. He must decide whether to study a particular prior paper in the

field directly, or to rely on a textbook explanation of it. To read the original paper is more costly but will provide him with a better understanding. For him to read the original is socially valuable since subsequent researchers will be able to use his work to improve their own understanding of the original source.

Example 2: research fields

A researcher is choosing the direction of her research agenda. She can choose to make an incremental contribution to an existing strand of research, or to pursue a novel idea. Pursuing the novel idea is more costly, but the reward is greater. If she pursues the novel idea, she creates new opportunities for subsequent researchers to mine the new area.

Example 3: content aggregators

A website creates content for internet users. It can choose whether to engage in original reporting on a topic, or to summarize and aggregate content created by others. Original reporting is more costly but generates more information. If it engages in original reporting, it generates new content which can be discussed and aggregated by other outlets.

Example 4: legal appeals

A court is hearing an appeal of a prior decision. It must decide whether to hear directly from original witness by calling them again, or rely on the record of their original testimony. Calling witnesses to appear again is more costly but gathers better evidence. A court hearing a subsequent case can in turn make use of the testimony collected if the witnesses are called again.

Example 5: legal precedent

A court is hearing a case. It must decide whether to apply a precedent established in a previous case, or to work from first principles. Applying precedent is easier, but may be less accurate since the precedent may have contained errors or be an imperfect analog of the new case. Courts considering subsequent cases will be able to use the new case as precedent in the future.

Example 6: decisions based on evidence

A person is deciding between two options. He must choose whether to evaluate the two options comprehensively, or to rely on a recommendation from a friend. Using the friend's recommendation is quicker, but their tastes may not be a perfect match. If he chooses to do his own research, then when another friend asks him in turn for his recommendation, it will be more informed.

The model we propose is designed to capture the spillover externality that is the common feature of these examples. However, to focus ideas we will return frequently to the application of a 'library' that controls access to an information source by setting prices or subscription fees, and consumers who must decide between paying for the high quality original source, or relying on an imperfect free copy.

First we consider this model when the price to access an information source is exogenous. In general the effect of the spillover externality in the consumption of information causes information to be used inefficiently infrequently: individuals over-rely on free, imperfect, recent information, and too seldom access the costly, higher quality original source, so that the length of 'branches' from the original source is too long. The outcome is more inefficient when the price to access information is higher, and more inefficient when the decay in information quality along branches is smaller. These are equivalent to saying that the outcome is more inefficient the longer are the branches. A Pigouvian subsidy to the cost of accessing information can restore efficiency.

Next we consider the model when the price to access information is set by an owner, who faces a fixed cost of storing the information source and incurs a variable cost of delivery each time a consumer accesses the information source.[1] This allows us to ask how who owns and prices information affects the way it is used and accessed over time. We consider the case in which the owner of the information has a monopoly over it and seeks to maximize profit, and the case in which the owner of the information is restricted to earn zero profit, either because the information can be supplied by many competitors or because of regulatory mandate. For each case we consider what the resulting price and use of information would be under two types of price menu: a flat price per piece of information accessed, or a subscription fee that allows subscribers access to information for free.[2]

2 Model

There is a single original information source. An individual arrives at each date $t = 1, 2, ..., \infty$. Each individual seeks to learn about the original information source, and does so by choosing at most one previous source to link with to receive information from. This can be the original source or a previous

[1] Depending on the application, it may be feasible and more appropriate to consider the variable costs as a cost of producing the information source.

[2] These are somewhat related: if a subscription is non-binding, so that it can be begun and ended freely at any time by consumers as their need for information arises, it is similar to a per-access price.

individual. For example, the individual who arrives at time 1 can link only to the original source, while the individual who arrives at time 2 can choose to link to the original source or to individual 1.

Individual j receives a payoff that depends positively on the value of the information she receives from the source she connects to, and negatively on the price of the link she forms. The value of the information she receives depends on the length of the shortest path between j and the original source 0. If the shortest path has length m, then j obtains information $1 - \delta m$, where δ is between 0 and 1[3]. If she does not connect to a source, she receives zero information.

The parameter δ therefore corresponds to the degradation or decay in the quality of information as it is passed through more steps before reaching the individual. This captures the idea that a person gets a better understanding of the source by consuming it first-hand rather than second-hand, and so on. As we mentioned earlier, we can be quite flexible in how to interpret the decay parameter. In particular, depending on the application it is possible to interpret δ as capturing decay over time—a collective forgetting as information lies dormant and unused—or decay due to dilution—a collective ignorance due to congestion in the number of users relative to the number of readers.

The cost to j of forming a link depends on the generation gap between j and the source she links to. The price to j of linking to her immediate predecessor i is 0 and the price to link to any other source is p, where p is between 0 and 1. That is, it is costless to receive information from an immediate predecessor (information that is contemporary, 'in the air'), but costly to access any older information. This captures the idea that an individual can learn directly from an immediate predecessor, but must access an archive of information to read previous generations.

3 Related literature on spillovers

Although we are interested in the microeconomic question of how market structure affects the use of information sources, spillover externalities have been studied in various other settings in the previous literature. Of particular relevance to our application of the model to open access, Jeon and Rochet (2010) presents a two-sided market model of academic journal pricing in which readers of academic papers generate positive externalities. The focus in that model is on the relationship between open

[3]The main implications of the model on the relative use of information under different market structures do not depend on this assumption of linear decay in the value of information over time. It does, however, have specific implications for how average revenue per time period depends on price, which we will discuss in detail below.

access policies and the quality standard of journals, while our focus is on the efficiency of pricing and usage of a given, exogenous repository of sources. We return to these distinctions and compare the implications of the model to the broader prior literature on the issues of journal pricing and article usage under open access in Section 7 below.

Another broad example of a spillover externality is the notion of knowledge accumulation in growth models (following Arrow, 1962, Lucas (1988) and so on). For example, Freeman and Polansky (1992) studies a growth model with overlapping generations in which there arises "a market price of knowledge that fails to reward the seller with the full value of his addition to the stock of knowledge", an outcome that is similar in spirit to our observation in our baseline case (without prices set by a firm) that the ownership of an information source is under-rewarded relative to the social optimum. Another example is Kemnitz and Wigger (2000), which demonstrates in a similar context to Freeman and Polansky (1992) an intervention to establish an unfunded social security system can be efficiency-enhancing by increasing incentives to acquire human capital, again similar to our observation that in our baseline case a subsidy to information owners can be efficiency-enhancing.

The distinction with our setting is that we are interested in source-specific externalities for exogenous sources, rather than aggregate effects of an expanding stock of endogenously-acquired knowledge. In our model, when an individual uses a source she generates an externality that is both particular to that source (independently to any potential value of her own contribution based on that source) and decaying over time: her successors can get an imperfect sense of the source by seeing her state-of-the-art use of it, and so on recursively, but eventually some future generation may seek to revisit the original source to understand it more fully than in its handed-down version. Therefore one way to view the concept of information in our model is that information can be owned and its replication restricted (the original source in our model), while knowledge based on that information is 'in the air' and cannot be controlled but dissipates over time (second-hand sources in our model). This distinction allows us to consider the effect of spillover externalities under different ownership structures for information sources, which is relevant to the kind of applications we suggested above.

4 Exogenous price

In this section we analyze the model with an exogenous linking price p (in Section 5 we consider price-setting by an owner who incurs costs to store and deliver information). An individual j can link to her immediate predecessor i at a price of zero, or link to sources before i at a price of p. Let p be no greater than $1-\delta$, to ensure that an individual would prefer to pay to access the source than to get zero information. It follows immediately that for each individual j, linking to the original source 0 at a price p dominates linking to any other source $s \in [1, i-1]$ at a price p. The optimal choice for j is thus either source 0 or source i, her immediate predecessor. Since the value of linking to i is decreasing in the length of the path from 0 to i and the value of linking to the original source 0 is constant, in equilibrium individuals link to their immediate predecessor until the length of the path from 0 to i reaches a threshold, at which point individual j links to the original source and a new branch begins. Figure 1 illustrates the spirit of this process.

Figure 1

The outcome of this process is in general not socially efficient. This is because when an individual revisits the original source, she conveys a positive externality on her successors, for whom the information they can access at zero price is now better than it would otherwise have been.[4] The individual who

[4] An analogy to this externality would be performing maintenance on a machine: subsequent users will benefit from the maintenance, but the individual does not capture this value.

revisited the source, however, does not capture any of this benefit.

To find the threshold at which an individual will go back to the original source, we find n such that the nth individual prefers to use the information in the air, but the $(n+1)$th individual prefers to use the original. That is, we seek n such that

$$1 - p - \delta < 1 - \delta n, \tag{4.1}$$

$$1 - p - \delta > 1 - \delta(n+1). \tag{4.2}$$

The equilibrium path length is therefore the smallest integer greater than $\frac{p}{\delta}$. Equivalently, equilibrium path length is n^* if

$$p \in [\delta(n^* - 1), \delta n^*]. \tag{4.3}$$

To arrive at the socially efficient path length we calculate the average payoff per individual, which we will denote $\bar{u}$, when path length is n:

$$\bar{u} = \frac{1}{n}\left[(\sum_{i=1}^{n}(1 - \delta i)) - p\right] \tag{4.4}$$

$$= 1 - (\frac{1}{2}n + \frac{1}{2})\delta - \frac{p}{n} \tag{4.5}$$

By analogous reasoning to the equilibrium path length, the socially efficient path length is therefore m if

$$p \in \left[\frac{m(m-1)}{2}\delta, \frac{m(m+1)}{2}\delta\right]. \tag{4.6}$$

Proposition 1. *Under an exogenous price p to access information:*

i. *Equilibrium path length is longer than the socially efficient branch length if $p > 2\delta$.*

ii. *In the case in which $p > 2\delta$, so that equilibrium path length is inefficiently long, equilibrium path length exceeds the socially efficient path length by more when p is larger and when δ is smaller.*

Any equilibrium path length of three or more is inefficiently long, and the inefficiency is larger the

longer is the path length. To see this, take some price p and consider δ decreasing. As δ gets smaller, the equilibrium path length grows more quickly than the socially efficient path length grows. The gap in average payoff per individual, given by 4.5, between the equilibrium and efficient path lengths therefore gets larger as δ decreases.

Figure 2 illustrates equilibrium and efficient path lengths for various prices p.

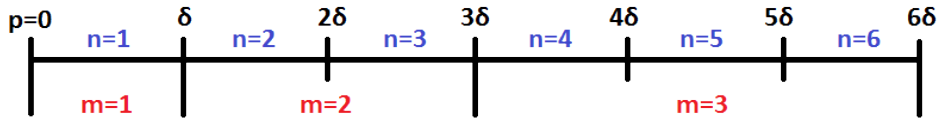

Figure 2: Equilibrium path lengths n and efficient path lengths m for various exogenous prices

To enforce socially efficient branch length would therefore require a Pigouvian subsidy whose amount is decreasing in the rate of decay of information and increasing in the price of accessing information. In this example, the socially efficient branch length can be enforced with a Pigouvian subsidy τ that modifies the cost to j of accessing sources prior to i to an amount $p - \tau$:

Corollary 1. *Under a subsidy of* $\tau = \begin{cases} 0 & \text{if } p \leq 2\delta \\ p - \sqrt{2p\delta} & \text{if } p > 2\delta, \end{cases}$ *equilibrium path length is socially efficient, conditional on the fixed price.*

The required subsidy is decreasing in δ, the rate of decay in information as the path to the original source increases, and is increasing in p. This reflects that the magnitude of inefficiency is higher for higher equilibrium path length. The intuition for this result is similar to that for Proposition 1. As δ increases to approach p (or equivalently as p decreases to approach δ), both equilibrium and efficient path lengths are decreasing. But the gap between them is also decreasing, since equilibrium path length decreases faster than efficient path length. Therefore the amount of the subsidy required to bring them into harmony decreases.

Before we move to the case in which prices are chosen by an owner of information, consider an application of the exogenous prices case. A researcher's institution has a site license to access a repository of information. However, the repository could be paper journals in library stacks, or electronic,

searchable and deliverable online. The time and effort that the researcher must put in to access a source is captured by the exogenous price she faces. The model with exogenous prices suggests that the researcher will access more original sources under the electronic delivery method, that information use will be less inefficient, and equivalently, the optimal social subsidy to information distribution will be lower. In the interpretation of the branches in the model as the mining of research subfields, the model predicts that electronic distribution would be associated with less recursive mining of a topic and more short strands of mining built from the same foundation.

5 Prices chosen by a firm

5.1 Monopoly

Next we consider the case in which the access price p is chosen by a firm. We assume that the opportunity to access the information at price p is always available; that is, at no time does the firm not allow a consumer to purchase access to the information for price p. Assume also that access to the original source is proprietary and controllable by the firm, but information 'in the air' is outside of the firm's control. First let p be set at time 0 by a monopolist that seeks to maximize its per-period average payoff. If it is active, the firm incurs a fixed operating cost of $f < \delta$ per period, and a cost $c \geq 0$ that it must pay each time it delivers a piece of information.

From the previous results, when the firm sets a price $p = n\delta$, it receives that amount $n\delta$ once every n periods. For the profit-maximizing monopolist, we can rule out $p \in ((n-1)\delta, n\delta)$ without loss of generality, since such p have a lower payment and the same frequency as $p = n\delta$. Since the price required to induce a sale once per n periods is proportional to n, average revenue per period is unchanging in the n implied by the firm's strategy.

Average payoff per period for the firm is then given by

$$\bar{\pi} = \frac{p}{n} - \frac{c}{n} - f. \tag{5.1}$$

Invoking $p = n\delta$ this becomes

$$\bar{\pi} = \delta - \frac{c}{n} - f, \tag{5.2}$$

which is maximized at the longest possible chain length, induced by $p = 1 - \delta$ and thus of length n_M equal to the largest integer not more than $\frac{1}{\delta}$.[5] This is because average revenue per period is not changing in chain length, but the firm must pay a cost each time it delivers information.

The fact that average revenue per period is not changing in chain length follows from the linear decay specification in the model. If demand for information is more elastic than implied by linear decay, then it may be that average revenue changes with price, which would affect the solution to the tradeoff between frequency of delivery and revenue for the monopolist.

At this point we may note that the monopolist's profit is increasing in δ. The faster the rate of decay in the quality of information available in the air, the more a consumer arriving at a given position in a chain is willing to pay to access the original information rather than rely on the information in the air. This willingness to pay effect outweighs the partially offsetting effect that the monopolist must deliver information more often, since the maximum chain length is decreasing in δ. In the context of our motivating examples, this implies that the monopolist would ideally like to impose constraints or inconveniences on consumers seeking to recount information second-hand. Whether it is possible in practice to influence δ will depend on the precise application we have in mind—to harass or invoke legal action against pirates of copyrighted material is readily possible, yet preventing colleagues from discussing a paper one has read is not. The preference for high δ by the monopolist has some affinity to the literature on a durable goods monopolist preferring a high depreciation rate (as in, for example, Schmalensee, 1974 and Waldman (1996)).

5.2 Firm facing competitive pressure

Next consider a restriction that the firm make zero profit. This captures a situation in which the owner of the information faces potential competition: the zero-profit firm is that which cannot be profitably

[5] This is such that the $(n_M + 1)$st individual weakly prefers to pay to access the original source than to receive zero information from the air.

undercut.[6] The zero profit restriction could also capture the situation in which the firm is a not-for-profit entity or is required to make zero profit by an enforceable regulation.

To find the zero-profit path length, we seek a branch length n_0 such that the firm's profit, as given by Equation 5.2, is greater than zero at n_0 but would be smaller than zero at $(n_0 - 1)$. The zero-profit path length n_0 is therefore the smallest integer no less than $\frac{c}{\delta - f}$. Equivalently, branch length is n_0 if

$$c \in [(n_0 - 1)(\delta - f), n_0(\delta - f)]\,. \tag{5.3}$$

This is the path length such that to induce a path shorter than n_0 is certainly loss-making. The zero-profit price to induce n_0 is $p = c + fn_0$, which is in the range from $(n_0 - 1)\delta + \epsilon$ to $n_0\delta$, the range of prices which induce n_0.

Competitive firms thus induce shorter path lengths than the monopolist provided that the fixed and variable costs f and c are sufficiently small to make zero profit feasible. Where the for-profit monopolist captures the surplus created by falling costs, under zero-profit ownership prices fall as costs fall and so the surplus created by falling costs accrues to consumers.

5.3 Efficient pricing

To find the socially efficient path length in this case we combine average payoff per period to the firm and individuals to arrive at a measure of total welfare per period W:

$$W(m) = \bar{u} + \bar{\pi} \tag{5.4}$$

$$= \underbrace{1 - (\frac{1}{2}m + \frac{1}{2})\delta - \frac{p}{m}}_{\text{consumer welfare per period}} + \underbrace{\frac{p}{m} - \frac{c}{m} - f}_{\text{firm profit per period}} \tag{5.5}$$

$$= 1 - (\frac{1}{2}m + \frac{1}{2})\delta - \frac{c}{m} - f \tag{5.6}$$

Unsurprisingly, the welfare maximizing δ is zero—if it were possible, the ideal level of decay in re-transmission of information would be no decay. This is because retransmission is free by assumption, accessing the information source is costly by assumption, and the firm's revenue is simply a transfer

[6] For example, say that two firms play a two stage game in which they simultaneously set prices and then simultaneously decide whether to stay in the market or exit. Another possibility is that the market is perfectly contestable.

from consumers.

From our expression for welfare per-period, we find that the welfare-maximizing path length is m^* if

$$c \in \left[\frac{m^*(m^*-1)}{2}\delta, \frac{m^*(m^*+1)}{2}\delta \right]. \tag{5.7}$$

Note that whenever path length exceeds m^*, welfare is strictly decreasing in path length: the magnitude of welfare loss is increasing in the amount by which paths are 'too long'.

Proposition 2 summarizes the results of this section:

Proposition 2. *When the firm sets a per-use price for information:*

i. *The socially efficient outcome is for the access price to be equal to $\sqrt{2c\delta}$, less than the cost of delivering information, and so for the firm storing the information to incur a per-period loss.*

ii. *When $c < 1 - \frac{f}{\delta}$, a monopolist can be profitable. Both the monopolist and the zero-profit firm set prices too high and induce paths too long relative to the social optimum. As c tends to zero, the magnitude of the inefficiency increases under a monopoly structure and decreases toward zero under a zero-profit structure.*

iii. *When $c > 1 - \frac{f}{\delta}$, the owner of the information cannot be profitable. The monopolist sets prices too high and induces paths too long relative to the social optimum. As c tends to 1, the magnitude of the inefficiency decreases towards zero under the monopoly structure.*

Figure 3 illustrates the relationship between the monopoly, competitive, and socially efficient path lengths as a function of the variable cost of delivering information, c.

For c higher than $1 - \frac{f}{\delta}$, no firms can profitably exist, and so for the market to be served privately requires barriers to exit for a firm. As c grows larger, the (loss-making) path length for the monopolist is inefficiently long, but tends toward the efficient path length. This is because at very high costs, the socially efficient outcome is to have information accessed as seldom as possible. Here this coincides with the monopolist's preference to minimize the frequency of paying the cost to deliver information.

For c high but below $1 - \frac{f}{\delta}$, the competitive and monopoly prices converge at the zero-profit price, since the monopolist's profit at the optimum is decreasing as c increases. Both the monopoly and

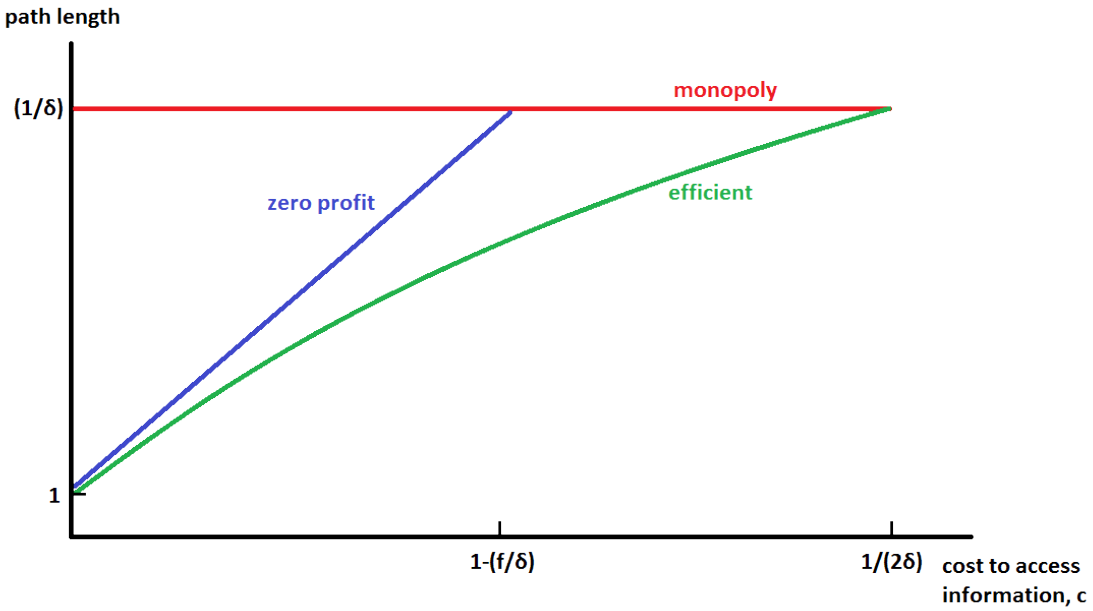

Figure 3: Path lengths by variable cost c

competitive prices are higher than the socially efficient price at these values of c, and so chains are 'too long': information is underused under both market structures.

As c tends to zero, the competitive price and path length decrease, converging to the socially efficient values. However, the gap between the monopoly and socially efficient path length and price increases. The efficient outcome is realized only when either the variable cost of delivering information is low and the market structure is competitive, or when the variable cost of delivering information is high and the market is served by a loss-making monopolist with a barrier to exit. If firms are required to break even, so that a loss-making monopolist cannot exist, then the monopoly outcome is never more efficient than the competitive outcome.

6 Subscription fees

The previous section considered a firm setting a price for the use of a piece of information. In this section we consider a firm setting a subscription fee that a consumer pays in advance for the right to access information at any time. Subscription fees are of particular relevance to many of the motivating examples for the model. For example we can see the subscription fee case as capturing the library model

of access to journal articles; an individual or institution pays a subscription to a journal for the right to access articles from it at any time.

We assume that a consumer must commit to subscribing or not *before* learning where they will arrive in the line of consumers for a given information source.[7] We will make a simplifying assumption about how path length from original sources relates to the number of subscribers in the population. We assume that if there is one subscriber per n individuals in the population, then the path length is precisely n: after a subscriber accesses an information source, precisely $n-1$ non-subscribers will follow before another subscriber arrives. This simplification allows us to avoid calculating the expected number of non-subscribers who will arrive between subscribers. We also assume a 'large' population of consumers, so that if there is one subscriber per n individuals, then if there was one fewer subscriber, there would be one subscriber per $n+\epsilon$ individuals, with ϵ arbitrarily small.

The problem for a consumer who is deciding whether or not to subscribe is therefore whether it is better to pay the subscription fee in order to always be precisely one degree removed from the original information source, or not to pay and be at some point between 2 and n degrees removed from the original source. That is, the consumer can choose not to subscribe and instead receive information that is 'in the air' for free, but the expected quality of that information will be worse the fewer subscribers there are. For a given individual, the relevant tradeoff is now not the *marginal* information available 'in the air' at a given moment versus the price to access the original source, but instead the *average* information available 'in the air' across all possible moments versus the subscription fee to access the original source.

This admits a new possibility for the length of chains of information. Recall that the value of information obtained by the nth individual in a chain is $1-\delta n$. For large enough n, so that n exceeds $\frac{1}{\delta}$, this value dips below zero. Let y be the largest integer that does not exceed $\frac{1}{\delta}$. If the per-access price set by the firm was so large as to make the $(y+1)$th individual was not willing to pay to access the original source, then it is certainly the case that no subsequent individual would be willing to pay such a price, since the marginal value of information 'in the air' is zero for all individuals beyond y. Therefore, chains must either not exceed y or be unboundedly long.

By contrast, under subscription fees, chains can indeed be finitely long but longer than y. The

[7] Note that a non-binding subscription fee, that a consumer could freely opt in and out of as her need for information arises, would be equivalent to a per-access price and so the analysis of the preceding section would apply in that case.

reason is that the *average* value of information obtained by individuals 2 through n is strictly decreasing in n for all n, even $n > y$. This means that there always exists a subscription fee such that an individual prefers to be one of $y + i$ non-subscribers than to pay to subscribe, but prefers to subscribe rather than be one of $y + i + 1$ non-subscribers. Any length of chain is possible, even those for which some individuals in the chain receive no information 'in the air'; that is, even those whose length exceeds y.

With this in mind, we can first calculate the relationship between subscription fee S and the length of chains—equivalently the proportion of individuals who are subscribers—that the fee induces. The payoff to a subscriber when the subscription fee is S is

$$u_{sub} = 1 - \delta - S. \tag{6.1}$$

The payoff to a non-subscriber when there is one subscriber per n individuals in the population is precisely the average payoff to positions 2 through n in a path of length n from the original source. There are two distinct possibilities for the nature of this average payoff, depending on whether or not $n > y$. First, consider the case in which $n \leq y$; that is, the number of non-subscribers relative to subscribers is sufficiently small that each non-subscriber receives information of positive value by consuming what is 'in the air'. Equivalently, the value of information obtained by the ith individual, $1 - \delta i$, is positive for all i. In this case,

$$u_{non}(n) = \frac{1}{n-1}\left[(\sum_{t=2}^{n}(1 - \delta t))\right], \tag{6.2}$$

$$= 1 - \frac{\delta}{2}(n + 2). \tag{6.3}$$

For path lengths to be n in equilibrium, it must be the case that when there is one subscriber per n individuals, both subscribers and non-subscribers prefer their choice than to switch. This requires that $u_{sub} \geq u_{non}(n)$ and $u_{non}(n) \geq u_{sub}$. That is,

$$u_{sub} = u_{non}(n), \tag{6.4}$$

$$1 - \delta - S = 1 - \frac{\delta}{2}(n + 2), \tag{6.5}$$

$$S = \frac{\delta}{2}n. \tag{6.6}$$

We can see that the subscription fee that induces a given path length in the range $n \leq y$ is lower than the per-use price to induce the same path length that we derived in Section 4. The reason for this is that the per-use price incentivizes the marginal consumer to switch back to the original source, while the subscription fee incentivizes the average consumer to subscribe to the original source. The per-use price must be sufficiently low for a consumer to prefer to pay to use the original source than to be the $(n+1)$th in line from the original, while the subscription fee must be sufficiently low for a consumer to prefer to pay to use the original rather than to be somewhere from 2nd to nth in line from the original.

The second possibility for the nature of the average payoff is that $n > y$; that is, the number of non-subscribers relative to subscribers is sufficiently large that some non-subscribers receive no value from the information that is 'in the air'. Equivalently, in the count of non-subscribers per subscriber there exists some i such that $1 - \delta i$ (and therefore $1 - \delta j$ for all $j > i$) is negative. In this case,

$$u_{non}(n) = \frac{1}{n-1}\left[(\sum_{t=2}^{y}(1-\delta t))\right], \tag{6.7}$$

$$= \frac{y-1}{n-1} - \frac{\delta}{2}\frac{(y-1)(y+2)}{n-1}. \tag{6.8}$$

For path lengths to be y in equilibrium,

$$u_{sub} = u_{non}(n), \tag{6.9}$$

$$1 - \delta - S = \frac{y-1}{n-1} - \frac{\delta}{2}\frac{(y-1)(y+2)}{n-1}, \tag{6.10}$$

$$S = (1-\delta)(1 - \frac{y-1}{n-1}) + \frac{\delta}{2}y\frac{y-1}{n-1}. \tag{6.11}$$

With this result we can go on to again analyze the outcome under monopoly and zero-profit ownership.

6.1 Monopoly

Consider the case in which the information source is controlled by a monopolist which chooses a subscription fee. Average payoff per period for the monopolist is, as before, given by

$$\bar{\pi} = \frac{S}{n} - \frac{c}{n} - f. \tag{6.12}$$

Invoking $S = \frac{\delta}{2}n$ this becomes

$$\bar{\pi} = \frac{\delta}{2} - \frac{c}{n} - f. \tag{6.13}$$

Similarly to the case in which the monopolist set a price for access, the monopolist receives revenue more often when it chooses a lower fee, but the extra frequency and lower fee offset. Therefore from the set of possible paths $n \leq y$, the monopolist again prefers to induce the longest possible path length, y. This involves setting a subscription fee of $S = \frac{1}{2}$.

We also must consider the case in which the monopolist chooses a subscription fee so as to induce a path length longer than y. In the previous section we derived in equation 6.11 an expression for the subscription fee required to induce a given path length $n > y$. Substituting this expression into the payoff function for the monopolist yields

$$\bar{\pi} = \frac{(n-y)(1-\delta)}{n(n-1)} + \frac{\delta y(y-1)}{2n(n-1)} - \frac{c}{n} - f. \tag{6.14}$$

Among those paths such that $n > y$, we can find that the highest payoff for the firm is associated with paths of length n if

$$c \in \left[\frac{1}{n-2} \left(\delta y(y-1) + (1-\delta)(n-2y) \right), \frac{1}{n-1} \left(\delta y(y-1) + (1-\delta)(n-2y+1) \right) \right]. \tag{6.15}$$

We can compare the value of profit at $n = y$, governed by equation 6.13, with the value of profit at $n > y$, governed by equation 6.14. It is the case that the monopolist can in some cases earn a higher profit by inducing paths of length beyond y. That is, it can be profitable for the monopolist to induce so few people to subscribe that not all non-subscribers can obtain any positive value from information 'in the air'.

To see why, consider the tradeoff faced by the firm. Inducing longer chains has the benefit of saving on average costs per period, as the firm must deliver the information less frequently. For $n \leq y$, this is always good for the firm, since average revenue per period is unchanging in chain length. However, for $n > y$, average revenue per period is decreasing in chain length. This is because consumers' willingness to pay declines faster as n increases after $n > y$ due to the extra chance of getting zero information as a non-subscriber.

The following result compares the outcome under a monopolist in the per-access price and subscription fee cases:

Proposition 3. *When information is owned and priced by a monopolist:*

i. *If $c < \delta$, path lengths are identical under per-access prices and subscription fees.*

ii. *If $c > \delta$, path lengths are longer under subscription fees than per-access prices, so that some consumers receive zero information under subscription fees.*

iii. *The monopolist's preferred subscription fee per use of information is less than its preferred per-access price.*

6.2 Firm facing competitive pressure

Next we return to the situation in which the firm controlling the information and setting the subscription fee is restricted to earn zero profit, due to competitive pressure or legal status. In the setting with subscription fees, we must again consider the possibility that paths can be longer than y. By analogy with the approach for the case with per-access price, if there is a zero-profit path length shorter than y then it will be found where profit, as given by Equation 6.13, is positive for branches of length n_0 but negative for branches of length $(n_0 - 1)$. In this way we can derive that the zero-profit path length n_0 is the smallest integer not less than $\frac{c}{\frac{\delta}{2}-f}$, or, equivalently, that branch length is n_0 if

$$c \in \left[(n_0 - 1)(\frac{\delta}{2} - f), n_0(\frac{\delta}{2} - f) \right]. \tag{6.16}$$

The zero-profit path length is longer for any c under subscription fees than under a per-access price, again reflecting the lower willingness to pay by the average non-subscriber for a given n than by the marginal nth consumer in a chain.

Under subscription fees it may be that there exists a zero profit path length longer than y. For paths beyond y, again we seek the path length such that

$$\bar{\pi} = \frac{S}{n} - \frac{c}{n} - f \tag{6.17}$$

is greater than zero for that path length, but less than zero for any shorter path length. Invoking the subscription fee required to induce some $n > y$, this payoff function is equivalent to

$$\bar{\pi} = \frac{(n-y)(1-\delta)}{n(n-1)} + \frac{\delta y(y-1)}{2n(n-1)} - \frac{c}{n} - f. \tag{6.18}$$

From this we can derive that if there is a zero profit path length longer than y, it occurs at n_0 if

$$c \in \left[\frac{1}{n_0-2}\left((n_0-y-1)(1-\delta) + \frac{\delta}{2}y(y-1) - f(n_0-1)(n_0-2)\right),\right.$$
$$\left.\frac{1}{n_0-1}\left((n_0-y)(1-\delta) + \frac{\delta}{2}y(y-1) - fn_0(n_0-1)\right)\right]. \tag{6.19}$$

The following result compares outcomes under a zero-profit firm in the per-access price and subscription fee cases:

Proposition 4. *When information is owned and priced by a zero-profit firm:*

i. *Path lengths are longer under subscription fees than per-access prices.*

ii. *The zero-profit subscription fee per use of information is smaller than the zero-profit per-access price.*

6.3 Efficient subscription fees

The efficient path length under subscription fees is identical to the efficient path length under per-access prices. This is because the payment from consumer to firm is only a transfer and is not part of the social tradeoff, which balances the payoff to individuals from better information—path length—against the technological cost of delivering information. Total social welfare is as before:

$$W(m) = \bar{u} + \bar{\pi} \tag{6.20}$$
$$= \underbrace{1 - (\frac{1}{2}m + \frac{1}{2})\delta - \frac{S}{m}}_{\text{consumer welfare per period}} + \underbrace{\frac{S}{m} - \frac{c}{m} - f}_{\text{firm profit per period}} \tag{6.21}$$
$$= 1 - (\frac{1}{2}m + \frac{1}{2})\delta - \frac{c}{m} - f \tag{6.22}$$

Therefore, as before, the welfare-maximizing path length is m^* if

$$c \in \left[\frac{m^*(m^*-1)}{2}\delta, \frac{m^*(m^*+1)}{2}\delta \right]. \tag{6.23}$$

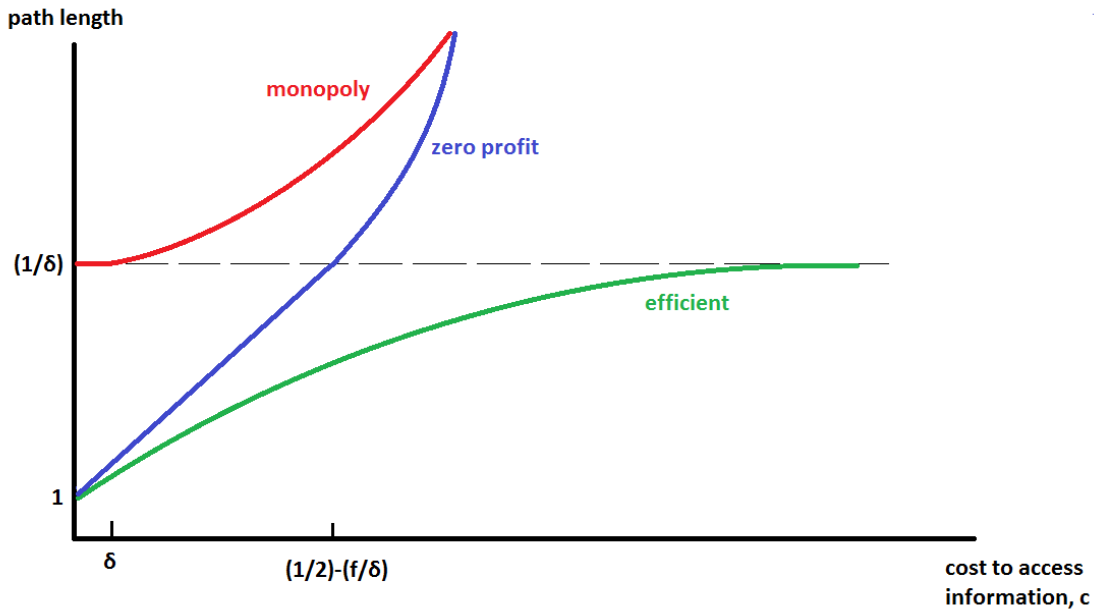

Figure 4: Path lengths under subscription fees by variable cost c

Figure 4 illustrates the relationship between the monopoly, zero-profit, and efficient outcomes for the case of subscription fees. The direction of inefficiency is similar to the per-access price case: once again, information is used inefficiently infrequently under both monopoly and zero-profit ownership. The reason is that once again firms have no way to recoup the spillover externality associated with the use of information. Again we see that as the cost c falls, the zero-profit path lengths tend toward the efficient level, while the gap between the monopoly path lengths and the efficient level grows.

A notable difference between the access fee and subscription fee cases is that subscription fees can decrease usage: paths are longer for a given cost c under subscription fees, for both monopoly and zero-profit structures. A second notable feature is that we observe cases under both structures in which paths are so long as to mean some consumers receive zero information. Third, the point at which no firm can make positive profit (in the figure this is where the monopoly and zero-profit curves intersect) is reached at a lower cost under subscription fees than under per-access prices. All of these

implications follow from the fact that consumers' willingness to pay for access to information is smaller under subscription fees than access fees for a given ratio of payers to non-payers.

Proposition 5. *When the firm sets a subscription fee for information:*

i. *The socially efficient use of information is identical to the per-use price case, but the subscription fee per use to induce it is smaller than the per-use fee that would yield the same usage.*

ii. *Both the monopolist and the competitive firm set subscription fees too high and induce paths too long relative to the social optimum. As c tends to zero, the inefficiency of the monopoly outcome increases relative to the inefficiency of the zero-profit outcome, which tends toward zero.*

6.4 Rationalizing the subscription model

Since subscription fees perform poorly for the firm relative to per-access prices, why are they such a pervasive model for distributing information? One clear rationale for the subscription model is that it eliminates the frictions and transaction costs associated with per-access prices. This is likely to benefit both sides of the market: the information owner is not forced to record consumption and collect and process payments at a rate proportional to the number of pieces of information a consumer accesses, and the consumer does not endure a payment in their consumption experience. Moreover, this rationale is all the more clear when the cost to deliver information, c, is zero, since the owner then has no reason to worry about frequent access by subscribers.

A subscription model may also be attractive to consumers with more subtle preferences than those in the model. Consumers in our model had very straightforward motivations, considering average quality of information obtained as a subscriber or non-subscriber and the size of the subscription fee. A richer conception of consumers' motivation may include a preference for flat rates over per-use fees, an empirical regularity well documented since Train et al. (1987) and with many plausible behavioral roots (Lambrecht and Skiera, 2006). Another possibility could be that subscriptions serve as a commitment device to consumers with time inconsistent preferences who may otherwise be tempted to eschew paying for a piece of information in favor of procrastination or some other use of their budget. Osten and Morton (2005) finds evidence for this kind of rationale in the pricing of magazine subscriptions.

The case of institutional subscriptions to information archives offers further rationale for the subscription model. In the model we can be quite flexible about whether a 'consumer' represents an individual or a collection of individuals. The key feature of the model was the information spillover that occurs from consumers of the information to non-consumers, which can be impossible for the owner of the information to price and capture. If the subscription is paid and consumed by an institution, for example a university or a firm, then it may be that the subscription fee captures some of the information spillover that occurs *within* the institution, even if the portion spilling outside the institution cannot be captured. The subscription model may therefore carry a higher willingness to pay from the institutional subscriber than our baseline model, and so admit a smaller efficiency loss than the one we have derived based strictly on external spillovers. Institutional subscriptions also have the potentially attractive feature of permitting egalitarian access to information within the institution, allowing for implicit cross-subsidies from units generating high revenue from the use of information to desirable but less lucrative units.

Finally, while so far we have drawn a distinction between price-per-access and subscription fees, it is common that owners of information adopt a hybrid pricing model that allows consumers to obtain information by either method. A price menu of this type naturally suggests price discrimination in the presence of consumer heterogeneity. If consumers vary according to how often they need to use information or in how much they value information, the owner can in general do better by offering both subscription access and a price per use designed to induce self-selection by different types of consumer via a standard second-degree price discrimination argument. Those consumers with a frequent need or high valuation select a subscription model, and those consumers with an infrequent need or low valuation select pay per use.

7 Open and online access

An example of a policy-relevant application of the model is the question of whether research output should be made available under open or online access policies rather than under proprietary control of an owner. In a typical example of an open access policy, a research funder requires "unrestricted access to the published output of research" and distribution of research papers through "free, online access" (Wellcome Trust, 2014). Researchers funded by the National Institutes of Health have since 2008 been

required to make "electronic version of their final, peer-reviewed manuscripts... publicly available no later than 12 months after the official date of publication" (National Institutes of Health, 2014). Does open access result in a more efficient use of information sources than proprietary ownership?

In the context of the model, an open access policy implies prices and subscription fees of zero to any consumer who seeks information. Consumers in the model choose whether to link to the original source based on its cost and its quality relative to the information freely available in the air. When there is no cost to access the original source, the consumer will always prefer to do so. Therefore the original source is accessed every period, with no consumer relying on information in the air. Note that since the model considers the pricing and trade of a given information source, the usage predictions in the model do not depend on how the open access repository is funded or administered. An open access repository could be run in a variety of ways, for example relying on public funding, recouping costs through author fees, or simply involving links to ungated copies of research papers on researchers' websites. This is not to say that the funding and administration structure will be irrelevant for the distribution of surplus between producers, owners, and consumers of information under open access. However, once a piece of information is in an open access repository, these considerations do not affect the cost to a consumer of accessing it, and so our conclusions on the efficiency of trade of existing information sources are independent of how open access is implemented. On a similar note, recall that information sources in the model are exogenously present and of exogenous quality. We therefore remain agnostic on the question of how the quality of information sources produced may change under an open access policy, a consideration that is studied in McCabe and Snyder (2004), McCabe and Snyder (2007), Jeon and Rochet (2010), and McCabe and Snyder (2010).

We can represent the open access situation alongside the other regimes by adding to Figures 3 and 4, which showed information usage by regime under per-access prices and subscription fees. Following our analysis above, we will be able to compare open access outcomes to proprietary ownership by either a monopolist, or by a competitive or not-for-profit firm. Figure 5 adds the open access case to the previous results.

The differences we observe in the model among these three ownership structures are consistent with those from prior literature on academic journals. Bergstrom (2001) and Dewatripont et al. (2007) document that prices charged for journals by commercial publishers are significantly higher than prices

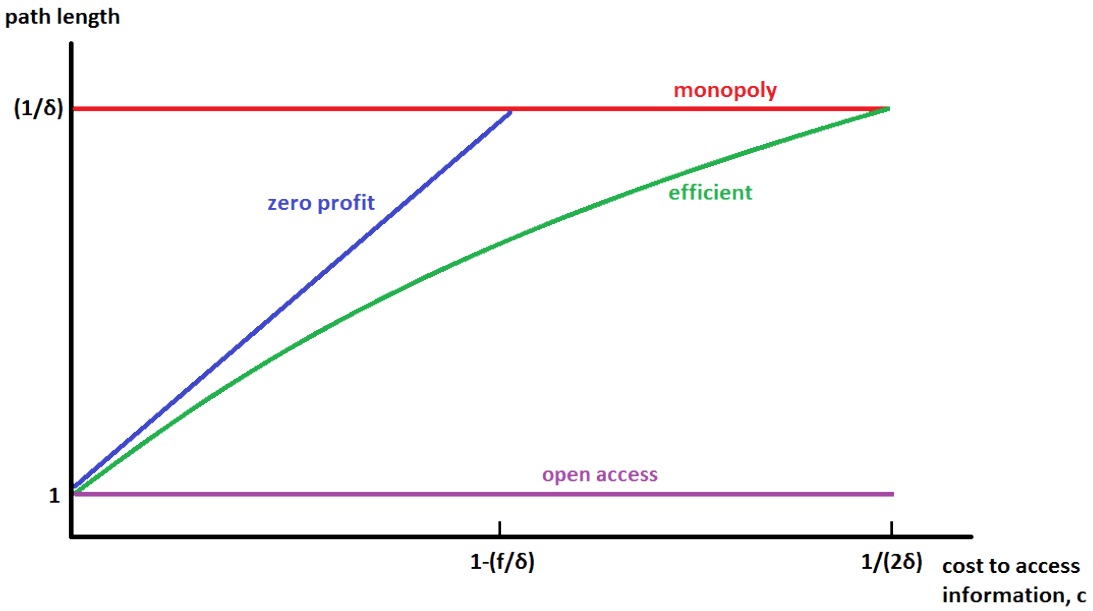

Figure 5: Path lengths under per-access fees, including open access

charged by university presses and professional societies, and McCabe (2002) shows a relationship between concentration in publishing and price increases for journals. White and Creaser (2004) shows that the price gap between commercial publishers and university presses has increased between 2000 and 2004, a period when it is plausible that the increased use of online distribution led to lower distribution costs. Similar conclusions on price trends as internet technology has advanced are made in Bergstrom and Bergstrom (2004) and McCabe and Snyder (2010). This evidence is consistent with the relationship between monopoly and zero-profit prices in the model.

Similarly, usage patterns in the model fit with prior evidence. Whether open access increases citation rates is a topic of some debate in the literature. Lawrence (2001) found higher citation rates for articles that are made freely available online when compared to other articles published in the same outlet. Antelman (2004) and Eysenbach (2006) find more citations for papers published under open access, but Gaulé and Maystre (2011) finds that this effect is not present when controlling for time trends and self-selection into open access based on quality. One way to reconcile this is to consider simply how often articles are consumed—read and downloaded—and to seek evidence of increased usage by consumers who do not necessarily themselves produce research papers that cite what they have read.

Randomized controlled trials in Davis et al. (2008) and Davis (2011) and a natural experiment on JSTOR access in Depken and Ward (2009) suggest that open access results in more downloading and reading of articles, even if not more academic citations.

Rather than academic citations, Bryan and Ozcan (2014) considers citation of academic medical literature in patent applications and demonstrates that articles published open access are cited in this way more often. The survey of inventors in Jaffe et al. (2000) suggests that around half of citations in patents reflect genuine knowledge transfer, so this would amount to meaningful marginal use of the information. This suggests that the marginal consumers of research—those who consume under open access but who did not consume under the pay model—may be those that Davis (2011) describes as communities "that consume, but rarely contribute to, the corpus of literature". These could be such communities as private sector individuals or institutions operating without the benefit of site subscriptions to journals. In a similar vein Evans and Reimer (2009) finds that open access has a large impact on citations in developing countries. In sum, there do appear to be important marginal consumers of information under open access, even if they are not necessarily producing academic research that cites the information. This is consistent with the model's prediction that some consumers receive zero information in the subscription fee model. The marginal consumers under open access being non-academic would follow naturally from heterogeneous valuation of institutional journal subscriptions by academic and non-academic institutions.

Armed with our simple model that is consistent with empirical usage and pricing patterns, we can then ask whether we should expect open access to deliver more efficient outcomes than other ownership structures. The use of the information source is only efficient under open access for small costs: truly free access will result in inefficient overuse until costs fall so much that the efficient outcome is always to access the original. As costs tend to zero, the socially efficient outcome is for the original source to be accessed as often as possible, which is precisely what open access delivers. As we already saw in the earlier analysis, this is also delivered by zero-profit ownership for costs close to zero. Recall also from the previous analysis that zero-profit proprietary ownership of information is always more efficient than monopoly ownership (so long as costs are not so high that profit cannot be as high as zero). Open access is more efficient than monopoly ownership when costs are small, but as distribution costs increase away from zero, non-profit ownership yields usage closer to the efficient level than open access.

For still higher costs, proprietary monopoly ownership can also yield more efficient use of information than open access. These results are because open access induces too much use of the original source whenever costs are large enough to make exploiting the spillover externality socially valuable. Zero profit ownership must account for costs, which of course are part of social welfare, but open access does not.

Although, as we have noted, the ownership of an open access repository does not affect these conclusions, one relevant point on the funding of open access repositories follows trivially. In any case in which a single proprietary owner of information can be profitable, the loss to that owner in moving to an open access regime exceeds the loss to a zero-profit owner in moving to an open access regime. Therefore, if a proprietary owner of information is mandated to offer open access to it, a transfer to the owner to precisely compensate for lost revenue will be quite significantly higher than a transfer to precisely compensate for the loss incurred in operating the open access archive. The magnitude of the transfer to the operator of the open access repository and its source, for example through public funding or the charging of author fees, clearly delineates the division of surplus between the operator of the archive, consumers of information, and the general public. Therefore selecting an approach to the funding of open access repositories will have clear distributional implications, even though the issue of funding does not impact rates of information use under open access.

Finally, in the open access application we have implicitly focused on thinking of the model as capturing the consumption of journal articles. Another aspect of research that open access may affect is in the amount of mining of research fields, which we mentioned as example 2 in the introduction. Evans (2008) suggests that

> "forced browsing of print archives may have stretched scientists to anchor findings deeply into the past and present scholarship. Searching online is more efficient and following hyperlinks quickly puts researchers in touch with prevailing opinion, but this may accelerate consensus and narrow the range of findings and ideas built upon."

This maps to the exogenous price case of the model from Section 4, since the user of the information faces a fixed transaction cost to seek out and use the print version of an article. This argument has much in common with inefficiently short branches under open access in the model. There is then too much reference to the original information source, and not enough recursive referencing—long branches—that

captures, in this application, the progress in mining a strand of research more deeply.

8 The case with $c = 0$

The case in which there is no cost to the information owner to deliver information merits special consideration. In some applications this will be the natural assumption, particularly in a world of digital delivery of information. In the context of the examples of applications of the model proposed earlier, if to revisit the original source means to perform an update of a study, $c > 0$ would be a sensible assumption. But if to revisit means to deliver an online reprint of an article, $c = 0$ makes sense. Zero cost therefore relates well to the evidence in the previous section on the comparison between open online archives and online archives controlled by a for-profit publisher.

As we can see in Figures 3, 4, and 5, at $c = 0$ the efficient outcome is to have information accessed maximally often. The zero profit market structure delivers this: the price (or subscription fee) need only cover the per-period operating cost f, which is less than δ and so results in access of the original source (or subscription) by each individual. Open online archives naturally deliver the same outcome, with the feature that they must make a loss on the operating costs that may be recovered by donations or grants.

By contrast, since the for-profit monopolist sets prices in consideration of consumers' willingness to pay rather than costs, the benefits in prices and access of zero cost do not reach consumers under monopoly ownership. This makes the $c = 0$ case one in which the benefits of other ownership structures are particularly clear.

9 Conclusion

We have proposed a model that captures a simple externality effect in the use of information goods: when a person uses an information source, they contribute a way to help subsequent generations to understand the original source. This feature of the cumulative use of information is consistent with a range of applications, including academic research, the corpus of common law, content aggregation, and customer reviews. In a baseline case with an exogenous price for information, this externality implies that information is inefficiently underused, justifying a public subsidy for information.

We considered proprietary ownership of information by a monopolist, private ownership by competitive firms or not-for-profit firms, and free availability of information, for example through open access. A general result that is that the the socially efficient use of information is not realized by either the monopoly or zero-profit structure, under either access prices or subscription fees. The socially efficient price for information is, as we would expect, increasing as the cost to deliver information rises. However, due to the positive externality associated with consuming information, the socially efficient outcome is only realized when information is priced below the cost that the owner incurs to deliver information. This means that the efficient outcome involves the information owner incurring a loss equal to the fixed cost of storing information plus a portion of the variable cost to deliver it. Therefore under both market structures, prices for information are set too high, and information is used inefficiently infrequently: information is underused and overpriced. It is impossible for any type of firm, whether profit-making or non-profit, to collect and account for the spillover externality associated with the use of information. Neither type of firm is capable of pricing efficiently.

There are, however, important differences in the magnitude of this inefficiency according to market and pricing structure. First, the monopoly outcome approaches the efficient outcome only when a loss-making monopolist faces barriers to exit and costs to deliver information are very high. This is because while the monopolist here always has an incentive to induce infrequent access to information, at very high costs it is also socially optimal to have individuals access information very rarely. Second, at intermediate levels of cost, the monopoly and zero-profit outcomes are both inefficient, both with prices too high and usage too low, but close to each other, so that the efficiency loss of monopoly ownership relative to zero-profit ownership is small. Third, as the cost to the firm of delivering information tends towards zero, the zero-profit outcome tends towards the efficient outcome, but the monopoly outcome becomes more inefficient.

These observations suggest that the efficiency implications of ownership structure will change as the nature of costs changes. For example, say it is the case that printing and paper storage are more costly than online delivery and storage of information as a computer file. If so, then for-profit ownership and zero-profit ownership of information could deliver quite similar social welfare under legacy technology, but quite dissimilar social welfare as technology advances. As information is increasingly digitized and delivered online, the efficiency loss associated with zero-profit ownership of information would fall, and

the efficiency loss associated with monopoly ownership would rise, so that zero-profit ownership would deliver greater social welfare than monopoly ownership, and increasingly more so as costs continue to fall.

We consider an application of the model to the issue of open access journals in Section 7 by introducing the possibility of information being publicly available without any price or subscription fee. We show that the model's relative prices and information usage under monopoly, zero-profit, and open access regimes are consistent with theoretical and empirical findings in the previous literature. Open access in general results in inefficient overuse of information, but this inefficiency falls towards zero as distribution costs fall towards zero. Therefore as technology reduces distribution costs, outcomes under either open access or zero-profit ownership both converge on efficiency. For higher costs, however, zero-profit ownership yields outcomes closer to the efficient outcome than open access, and for very high costs, even monopoly ownership is closer to the efficient outcome than open access. This is because free availability completely untethers the price that consumers face—now zero—from the cost to store and distribute information. For this reason, in general zero-profit ownership of information, which must respect costs, delivers more efficient outcomes than open access.

Open access policies are in the model associated with greater social welfare in the use of information sources than profit-maximizing pricing as long as distribution costs are sufficiently small. How this increased social surplus will be divided between consumers of information and the owner will depend on how the open access repository is funded, for example whether the owner grants open access at no cost to the producer of information, or charges a fee in exchange for granting open access status. A caveat on this analysis is that the model assumes an exogenously existing information source. We therefore do not consider the potential effects of lower monopoly profits on the incentives to produce information. If the monopolist over information is also its producer, then we must consider this potential incentive cost.

In sum, in a low-technology, high-cost world, for-profit control of information is defensible since the efficiency loss is not too great relative to not-for-profit ownership, whose outcome is closer to the social optimum. But in a high-technology, low-cost world the efficiency gain of not-for-profit ownership is larger relative to monopoly ownership. This conclusion holds under either pay-per-use pricing or a subscription fee model. While profitable proprietary ownership of information by a monopolist may be

more socially desirable than open access when distribution costs are high, it is always outperformed by zero-profit ownership, and as costs fall becomes less desirable than open access. We therefore argue that as the cost of storing and delivering information goods continues to fall, ownership of information by a monopolist is increasingly untenable relative to not-for-profit ownership and open access.

Reconciling these forces that are generated by the externality in question with the incentive effects for the producers of information is an open question. For some applications it seems reasonable to assume that the incentives of information producers will not be degraded very much by a smaller return to the owner of the information, while in other applications the incentive effect may be much more important. Nevertheless, if we are concerned with social efficiency or the benefit of information to the interested public, the presence of spillover externalities in the use of information, particularly when combined of falling costs of information distribution, is one justification to discourage for-profit monopoly control of information sources in favor of competitive ownership, not-for-profit ownership, or open access policies. A final consideration is that the choice of how to fund the open access repository will be of acute importance for the distributional implications of adopting an open access policy, and for the effect on existing information-owning firms during a transition from proprietary ownership to open access.

References

Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*, 65(5):372–382.

Arrow, K. (1962). The economic implications of learning by doing. *Review of Economic Studies*, 29(3):155–173.

Bergstrom, C. T. and Bergstrom, T. C. (2004). The costs and benefits of library site licenses to academic journals. *Proceedings of the National Academy of Sciences*, 101(3):897902.

Bergstrom, T. C. (2001). Free labour for costly journals? *Journal of Economic Perspectives*, 15(4):183–198.

Bryan, K. and Ozcan, Y. (2014). The impact of open access mandates on innovation. mimeo.

Davis, P. M. (2011). Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *The FASEB Journal*, 25(7):2129–34.

Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., and Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. *British Medical Journal*, 337:331–345.

Depken, C. A. and Ward, M. R. (2009). Sited, sighted, and cited: The effect of jstor in economic research. SSRN Working Paper http://ssrn.com/abstract=1472063.

Dewatripont, M., Ginsburgh, V., Legros, P., and Walckiers, A. (2007). Pricing of scientific journals and market power. *Journal of the European Economic Association*, 5(2-3):400–410.

Evans, J. A. (2008). Electronic publication and the narrowing of science and scholarship. *Science*, 321(5887):395–399.

Evans, J. A. and Reimer, J. (2009). Open access and global participation in science. *Science*, 323(5917):1025.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, 4(5):e157.

Freeman, S. and Polansky, S. (1992). Knowledge-based growth. *Journal of Monetary Economics*, 30(1):3–24.

Gaulé, P. and Maystre, N. (2011). Getting cited: Does open access help? *Research Policy*, 40(10):13321338.

Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *The American Economic Review*, 90(2):pp. 215–218.

Jeon, D.-S. and Rochet, J.-C. (2010). The pricing of academic journals: A two-sided market perspective. *American Economic Journal: Microeconomics*, 2(2):222–255.

Kemnitz, A. and Wigger, B. U. (2000). Growth and social security: the role of human capital. *European Journal of Political Economy*, 16(4):673–683.

Lambrecht, A. and Skiera, B. (2006). Paying too much and being happy about it: Existence, causes, and consequences of tariff-choice biases. *Journal of Marketing Research*, 43(2):212–223.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature (correspondence)*, 411(6837):521.

Lucas, Jr., R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.

McCabe, M. J. (2002). Journal pricing and mergers: A portfolio approach. *American Economic Review*, 92(1):259–269.

McCabe, M. J. and Snyder, C. M. (2004). A model of academic journal quality with applications to open-access journals. NET Institute Working Paper No. 04-18.

McCabe, M. J. and Snyder, C. M. (2007). Academic journal prices in a digital age: A two-sided market model. *The B.E. Journal of Economic Analysis & Policy*, 7(1).

McCabe, M. J. and Snyder, C. M. (2010). The economics of open-access journals. SSRN Working Paper http://ssrn.com/abstract=914525.

National Institutes of Health (2014). Nih public access policy details. NIH. http://publicaccess.nih.gov/policy.htm. Accessed 30 Aug 2014.

Osten, S. M. and Morton, F. M. S. (2005). Behavioral biases meet the market: The case of magazine subscription prices. *Advances in Economic Analysis & Policy*, 5(1).

Schmalensee, R. (1974). Market structure, durability, and maintenance effort. *The Review of Economic Studies*, 41(2):pp. 277–287.

Train, K. E., McFadden, D. L., and Ben-Akiva, M. (1987). The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The RAND Journal of Economics*, 18(1):pp. 109–123.

Waldman, M. (1996). Durable goods pricing when quality matters. *The Journal of Business*, 69(4):pp. 489–510.

Wellcome Trust (2014). Open access policy. Wellcome Trust. http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm. Accessed 30 Aug 2014.

White, S. and Creaser, C. (2004). Scholarly journal prices: Selected trends and comparisons. LISU Occasional Papers No. 34.