Open portals or closed gates?
Channeling Content on the World Wide Web[☆]

Eszter Hargittai*

*Sociology Department, Princeton University, 2-N-1 Green Hall,
Princeton, NJ 08544, USA*

Abstract

This paper explores what the tension between information abundance and attention scarcity implies for the diversity of information accessible to users of the World Wide Web. Due to limited user attention, there is a role for gatekeepers in the online content market. Sites that catalog Web content and primarily present themselves as content categorization services are identified as the gatekeepers in the new information age. Exploring the mechanisms by which they organize content is essential to understanding how user attention is allocated to information available on the Web. Theories about media content diversity are delineated to suggest what we may expect with respect to content diversity online. Methods for future empirical investigation are suggested. Finally, the policy implications of the argument are presented. © 2000 Published by Elsevier Science B.V. All rights reserved.

> "What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."[1]
>
> Herbert A. Simon (1971)

1. Introduction

Although the World Wide Web only became a popular communication medium in the mid 1990s, its publicly accessible segment already contains over one billion pages (Censorware, 2000). Finding the type of information one is looking for becomes increasingly complex with the rapid growth of available content. To address this problem, Web content classification services were developed by some users. Soon, they became profitable online ventures and some of the most popular Web sites.[2] For the purposes of this paper, I identify any site that classifies content and primarily presents itself as a one-stop point-of-entry to content on the Web as a portal site. Many Web sites have a section with links that may be categorized according to certain themes. The proposed definition only includes those where categorization is one of the most important features of the site. The section on the rise of navigational services describes the most popular portal sites in more detail.

A tremendous amount of information is available on the Web given how easily any person with access to the Internet can post data on his or her own initiative. This facility associated with the use of the network – both with respect to posting and retrieving information – has led many to speculate about the potential effects of this medium on all spheres of life including political, social, and economic consequences. Of interest here is the role of the medium in the cultural sphere with particular emphasis on the diversity of cultural offerings accessible to users. Although the argument presented here about the problems associated with the narrowing effects of portals can be extended to all types of content, this paper focuses on cultural content in particular to investigate the issue in detail with respect to that case study. For the purposes of this paper, it is important to distinguish between content that is merely present on the Web in contrast to content that users are readily exposed to. To make this distinction, I will use the word 'available' to refer to material that exists online and will use 'accessible' to denote content that is easily within

---

[1] The amount of attention any one person has does not decline in itself, that is constant over time. The 'poverty of attention' should be read to imply that the amount of potential attention for each piece of information declines dramatically. Thanks to Bob Wuthnow for pointing this out to me.

[2] A Web site is a collection of related Web pages maintained by one person, a community of people, or an organizational entity. The components of the site are usually housed in one location. A site is made up of a collection of interrelated Web pages that are not only connected in content (e.g. a common theme) or function (e.g. a common service such as auctions) but also via hyperlinks. These links allow direct movement (by the click of the mouse) from one page to another among all of the pages on the site. A Web site is the aggregate of such connected Web pages.

the reach of Web users. Whereas 'availability' means mere existence, 'accessibility' implies relative ease of reachability.

The Internet has the potential to create arenas for more voices than any other previous communication medium by dramatically reducing the cost of the replication and distribution of cultural products. However, information abundance still leaves the problem of attention scarcity. Ironically, even people who have recognized the importance of attention scarcity have suggested that any individual will be able to sidestep organizations and corporate packaging in an attempt to receive attention (Goldhaber, 1997). In contrast, I emphasize that attention scarcity leads individual creators of content to rely on online gatekeepers to channel their cultural products toward consumers of content. Web services that categorize online information can be considered gatekeepers on the World Wide Web. This paper looks at the implications of online gatekeeping mechanisms for the diversity of content accessible to Web users.

In the next section, I discuss existing theories about the production and distribution of cultural goods with particular emphasis on gatekeeping online. Then, I explore how some of the most popular portal sites attempt to deal with issues of uncertainty in the online media market. Next, I present theories about media content diversity and their implications for online content diversity. I then suggest some ways for empirically testing the proposed hypotheses. Finally, I discuss the policy implications of the line of argument and highlight questions for future research.

2. The production and distribution of cultural goods

Kurt Lewin first coined the term 'gatekeeper', arguing that news items flowing through communication channels get to points that function as gates blocking the flow of some material while allowing other information to pass through (White 1950). Previous studies on industries of cultural products such as books (Powell, 1985), news publications (Tuchman, 1978; Gieber, 1964), and popular music records (Peterson and Berger, 1975; Lopes, 1992) have explored the role of gatekeepers in influencing the type of cultural products that are produced and distributed on the market.

With previous cultural goods, the costs of production were so high that a vitally important gatekeeping step concerned the decision about what products should be produced. Studies have documented the coping mechanisms that firms adopted to deal with the uncertainty of large investments in cultural products. Paul Hirsch (1972) looked at how firms in the music industry cope with the uncertainty that surrounds the creation of musical products. His two main questions concerned the criteria that companies use to determine which products to pursue, and whether certain characteristics of the sponsoring organization – such as prestige or advertisement budget – affect the chances of a product being picked out by a firm to pursue on large scale production. Hirsch also looked at the mechanisms adopted by these organizations to cope with the difficulty of disseminating information about their products. In the case of the music industry, disk jockeys at radio stations play an impor-

tant role in deciding what products consumers learn about. The two questions Hirsch looked at underscore the two processes that a product must undergo before reaching the market: production and distribution. One common theme in all the studies cited above is that individual creators of cultural products have to go through both producers *and* distributors of their products to get attention on the market. The final link in the distribution chain – supermarket rack jobbers, movie critics, book review editors – can be a key figure in allocating people's attention to products.

On the World Wide Web, portal sites are the main gatekeepers between the creators and consumers of cultural products available on the network. People who think of the Web as a voice for any and all opinions base their assumption of equal access on the low barriers of entry regarding the production of goods that can be created and distributed on the Web.[3] By allowing a vast reduction in the replication and distribution costs of a product – whether text-based, audio, video, or multi-media – the Web puts product dissemination within the reach of the individual. This reduces the salience of the gate that functions between the creator of a product and the materialization of the product. Not only can a person create a product easily, it is also possible to make numerous copies of it available at very low cost. Moreover, because it is no longer necessary to transport these items physically, it is also nearly effortless to allow access to the product from various geographic locations.

However, because the barriers of entry are so low, a huge amount of cultural products flood the market leaving the user overwhelmed with the task of navigating through the vast amount of options offered online. Given this situation, the central concern is no longer what is produced, but what consumers hear and know about. Accordingly, gatekeeping activity still occurs, but now takes place at the level of product exposure. Its location has shifted from the decision about what should be produced to control of what products get to consumers and what they become aware of. Lewin had suggested that the first task in the research on gatekeeping is to identify the actual gatekeepers (White, 1950). In the online world, the most popular navigational sites are the most prominent gatekeepers.

3. The rise of navigational sites on the World Wide Web

Although the origins of the Internet were first constructed and used in the 1960s, the network was initially restricted to a small community of scientists and scholars in just a few nations. Moreover, the World Wide Web – the foundation of the Net's wide popularity – was invented only in 1991 and the graphical interface that made its use accessible to the layperson, the Web browser, was created only in 1993. It was this addition to the technology that significantly accelerated its spread. Thus,

---

[3] This approach also assumes that all people have equal access to online technologies. It is important to remember that access to the Internet is stratified among the population both in the United States (NTIA, 1999) and internationally (Hargittai, 1999, ITU 1999). In order to post one's opinion, one needs access to a network-connected device and has to have enough knowledge about online services to know how to post a message in a forum, on a mailing list or a Web page.

significant Internet diffusion can be observed only in the past few years. Between the years 1994 and 1999, the use of the system more than quadrupled in the United States with over 30 percent of Americans becoming regular users and over 50 percent of the population online (Nielsen, 1999). Moreover, not only are there more users, but these people are spending increasing amounts of time browsing the Web (Media Metrix, 1999a). Given the two types of services on the system – computer-mediated communication and information retrieval – the multitude of services the network allows for is unprecedented. To assist users in dealing with the abundance of information available on the Web, sites started offering search services and categorization of online content.

The increase in users has been complemented by an increase in services and materials available on the Web. To address the complexity of the system, search sites were created soon after people started using Web browsers to navigate online content (e.g. Yahoo! and Excite in 1994, Lycos in 1995). Initially, these sites worked in two ways: they either featured a search engine to find sites (e.g. Excite, Lycos) or offered a list of category directories where staff from the search site's company manually compiled the list of featured selections (e.g. Yahoo!).

The most popular Web sites grew out of earlier search sites and browser home pages. Browser home pages are the default pages of Netscape and Internet Explorer. Initially, these home pages did not offer much beyond software upgrades, but eventually they started to adopt the multiplicity of services that search sites started to feature. Today, Netscape's Netcenter and Microsoft's MSN.com are both much more than a resource for browser software upgrades. They both offer categorization of Web content in addition to many other online services.

Eventually, these sites grew and started offering new services to attract and keep site visitors. As one company's self-description puts it: "The Company [Lycos] seeks to draw a large number of viewers to its Websites by providing a one-stop destination for identifying, selecting and accessing resources, services, content and information on the Web" (Lycos, 1998). By augmenting their services and continuing to be the most visited sites, these pages increasingly became known as the point-of-entry sites or 'portals' on the Web. According to one report (WebSideStory, 1999), Yahoo accounts for over fifty percent of all search engine page referrals. Given that 20–50 percent of all Web users visit each of the most popular Web sites (Media Metrix, 1999b), these sites play a potentially immense role in allocating user attention. Therefore, a look at the organization of content on these sites is warranted. Information on the specific activities of users at these sites is proprietary and kept highly confidential by navigation service companies. Consequently, it is hard to know what proportion of visitors to such sites use category directories versus search engines to locate information. Information from an interview with a site content manager reveals that there is a considerable amount of browsing through category directories. Such user behavior has led the company to concentrate increasing efforts on the improvement and placement of such categories.

Navigational sites function as points-of-entry to the content of the World Wide Web by offering various services (e.g. search engines, category directories, special content, etc.) that enable users to find their way through the enormous amount of

material on the Web. (See the Appendix for a detailed description of the services offered on these sites.) The most popular of these became known as megaportals.

3. Locating content on the Web

Because this article is concerned with understanding to what extent consumers have access to the whole range of content produced on the Web, it is important to delineate the processes by which consumers locate content. Two processes are important to consider here. On the one hand, there are the mechanisms consumers use to arrive at a site. On the other hand, there are mechanisms that portal sites use to decide what content to feature on their pages. This section looks at how users may arrive at a site. A later section will explore the mechanisms portal companies use to organize content on their sites.

The following is a list of ways a user may find his or her way to a Web page:

3.1. Browser's default page

The most straightforward way of getting to a page is by having it as the default page on one's browser. Although the user may change the original default page, it is often a page specified by the browser's manufacturer, the user's Internet service provider, or the institution where the machine is operated (e.g. school or work).

3.2. Previous knowledge of a page

A similarly direct way of getting to a page is by typing its URL (Uniform Resource Locator, i.e. Web address) in the location bar of the browser. A user may know about a specific URL by having used it in the past. In such a case, he can access it through a bookmark he may have set or simply by typing the address in directly on the location bar. He may also know about a page by having been exposed to it through another medium, e.g. a recommendation from a friend or colleague, or an advertisement in a newspaper, on radio, TV, or a billboard.

3.3. No previous knowledge of a page

If a user does not have a specific site in mind (or does but does not have the actual URL on hand), but she is interested in finding something relatively specific or even somewhat general, she can use one or a combination of the following mechanisms to arrive at a page.

(1) *Address guesses*

A user may attempt to use a specific URL even if he does not have previous knowledge of a page. If the user is interested in locating the Web page of a known company, brand of product, organization, or individual, he can try to

guess the URL of the desired location. It is because of the traffic gained from such guessing that domain names have become such hot commodities. Beyond intellectual property claims (e.g. Umbro International's claim to umbro.com (Kaplan, 1999)), domain names that are derived from existing words are important because people use them to find sources. Users who have a general topic in mind can type in a generic URL along the lines of www.topic.com (e.g. www.cars.com, www.games.com) which will probably lead to an actual site, although perhaps not of particular relevance. In some versions of browsers, it is enough to type 'topic' in the location bar since www.x.com has been designated as the default and thus 'topic' will automatically lead one to www.topic.com.

Instead of typing in a URL in the location bar, a user may go to a portal site and use one of several options provided by the portal to find a page. One of the oldest forms of locating information is through the use of a search engine.

(2) *Open search*

The user types a word or phrase she is interested in in a search text bar and gets a list of results. She can then choose from that list or run another search. The search will result in a list of links that the user can then click on and be linked to sites that supposedly have some connection to the search criteria. Some of the possible links lead to directories that include information and links categorized according to topic.

Although it may seem as though a search engine is the mechanism that leaves least room for gatekeepers to influence outcomes, this is not necessarily the case. First of all, search engines only index 10–35 percent of all Web pages leaving a very large portion of the Web inaccessible through the use of this search mechanism (Lawrence and Giles, 1999). Moreover, as Nissenbaum and Introna (1999) have shown, search engines systematically exclude (in some cases by design and in some accidentally) certain sites in favor of others. Thus, it is important to realize that even through the use of what seems like a neutral mechanism, users' actions are influenced and their attention channeled toward certain content over other material.

Instead of typing in search terms in a search bar, the user may click on one of a myriad of links on the portal, links such as directory categories or advertisements.

(3) *Directory categories*

The user may start browsing the categories presented prominently on portals. These are category directories compiled either by staff or a program. They reflect conscious decisions on someone's part about what to include – and what to exclude. On Lycos, for example, such directories – in order of presentation – include a list of subcategories, a list of related content pages on the Lycos Network, a list of actual Websites, a list of related newsgroups, and a list of related Lycos Web directories.

(4) *Advertisements and sponsorship*

Portal sites – and many other Web pages – include banner advertisements or other strategically positioned links that may or may not look like ads, but are in fact the result of partnerships with content providers or other businesses. By clicking on such links, users may arrive at various pages.

(5) *Browser software add-ons*

One more option is for the user to rely on various browser software add-ons for finding previously unknown pages. These programs (e.g. NeoPlanet and Alexa) add to the functionality of the browser software by offering site recommendations depending on the users' browsing habits. However, these programs are platform specific and may require specific software. Moreover, users must have some technical expertise in downloading and installing special software in addition to the initiative required for such action.

The list demonstrates that there are numerous mechanisms by which a user may arrive at a Web site. In several of these mechanisms, what the user finds depends on decisions made by portal site personnel on what to feature on their navigation sites, where to feature them, and how they should be presented (e.g. a simple text link or a more elaborate graphical logo). To complement the discussion in this section on mechanisms by which Web users locate content, the following sections discuss how portal companies decide what content to include in the directory categories, advertisements, and sponsorships on their sites.

4. Informed advertisers, uninformed public

The goal of any online commercial venture – in this case portal companies – is to make a profit. The question of how this may be done has been the topic of debate and speculation ever since commercial interests appeared on the network in the early 1990s. Government support for media content is rare in the United States.[4] This leaves the burden of financing to other potential sources such as individuals (e.g. users paying subscription fees for services), private foundations, or corporate sponsorship. So far, the trend has been toward the latter. Most online services are funded through advertisements, by venture capitalists, or through corporate cross-subsidization where the profitable division of a company covers the costs of the online undertaking. In order to legitimate funding, Web sites must attract and keep visitors and encourage them to stay and revisit frequently.

The rise in the importance of visitor information has led to a system of viewer ratings analogous to the Nielsen ratings system used for assessing television viewership. The portal companies themselves and third-party agencies (e.g. Media Metrix

---

[4] Although the Web is an international medium, given that the most popular portal sites are based in the United States and the largest proportion of users are still from this one country, this discussion concentrates on the American aspects of the network.

and Nielsen Net Ratings) collect information on use of the Web site such as the number of unique visitors to Web sites, how much time users spend on sites and how many unique sites users visit in one browsing session.

Due to the interactive nature of the Net, online advertising is allowing for much more intelligent and informed methods of promotion than previous technologies. It is not only easier to target specific consumers, but the possibilities for tracking effectiveness of various ad campaigns are also more sophisticated. By gathering information on user activity and preferences, sites have information on individual visitors and can target advertisements accordingly. A site can obtain information about a user through various methods. By using cookie technology that sends information about a user's computer and their online actions to the site, portals can derive information about the visitor by analyzing what activities they engage in online (e.g. what type of pages they visit, what type of links they usually click on). Alternatively, through personalized services such as the 'My' options of various portals (described in more detail in Appendix I), portals have information about the self-identified interests of customers. A user who decides to have information about a sports team displayed on their customized welcome page (e.g. My Yahoo! or My Lycos) is probably interested in that sport and in that team and thus may be a good target for advertisements on products relating to either.

Similarly, by having exact information on how a user got to a page (e.g. what type of advertisement (banner, text, etc.) the user clicked on to get there) the advertiser gains knowledge about the relative effectiveness of various marketing tactics. More traditional advertising techniques do not allow for such detailed information about user response to marketing strategies. Although the use of 1-800 numbers can inform companies on which particular ad may have worked best, such methods are unable to provide details of user habits surrounding a response to an ad. The specifics of online advertising make the Web an attractive medium for marketing purposes. Portals can greatly benefit from detailed marketing information as the most visited online destinations.

As companies refine their advertising strategies in response to information they gather through the interactive nature of online marketing campaigns, they devise strategies that are often hard to assess from the user's perspective. Advertisements blend in with site content to such an extent that it can be extremely difficult for users to distinguish between the two. It is possible to purchase placement on search result lists (see GoTo.com for a completely fee-based example) or special graphical or text references prominently displayed next to search results. There are also many commercial links on various category directory pages, mostly to products that correspond to the topic of the directory.

The business motivations behind featured links are not always clear to users. Note, for example, the following example from the welcome page of Lycos on May 17th, 1999. 'Lycos the dog' is the mascot of the Lycos portal. The dog is usually prominently displayed somewhere on the welcome page and is always present on advertisements for the company in other media, e.g. print magazine ads. On the May 17th, 1999 welcome page, the dog was positioned left of the search engine box on the upper part of the page. Note that the date is just two days before the release date

of the Star Wars movie Episode I, the Phantom Menace. In this image, the dog is holding a Star Wars light saber and the link above the image reads "Shop here for Episode I products". A press release from April 6, 1999 (Lycos, 1999) reveals that Lycos had made "an exclusive agreement to sell Star Wars merchandise in a new co-branded Web site, The SCI FI Store". Although clicking on the image does not lead to the same shopping site as the text link, it does take the user to the Lycos Star Wars page that features several local (i.e. Lycos-based) commerce links.

5. Portal strategies

The multitude of services offered by megaportals such as Yahoo!, Lycos and Netscape's Netcenter illustrates that they have had to redefine their role over the years in light of increasing industry competition and increasing complexity in information and services available on the Web. In their first years, the forerunners of today's portals – search engine sites and browser homepages – offered few services beyond their narrow foci (i.e. search engine results and software upgrades, respectively). Over time, the sites have adopted new strategies to achieve their goals of high revenues. Sites have implemented changes in an attempt to attract an increasing number of visitors and to keep them on the site as long as possible. By doing so, they increase their value to advertisers and can increase their revenues.

How do portal sites determine what services and information to present on their pages and how do they decide how to organize this information? I explore the processes used by these sites in deciding what information to present to their audience and how they organize their content. There are two main sources of revenue for megaportals: income through advertisements placed on their pages and commissions collected from e-commerce destinations of their users. Portals strike deals with companies that sell services and products online. In exchange for offering links to the pages of such companies, portals derive a percentage of the income generated by the customers who find the e-commerce site through the portal. Branding and direct sponsorship of content is also a possibility. Increasingly, portals also offer products directly on their sites most often by allowing businesses to sell on their online property. The portal company also gets a percentage of such income, thereby augmenting its own revenues.

In order to attract advertisers, portal sites must 'deliver eyeballs' for the ads, i.e. they need as many visitors as possible. Moreover, they need these visitors to stay on their sites as long as possible generating additional page views as they surf throughout the pages of the portal site. However, advertisers are looking for users who click through their advertisements and visit their pages as opposed to being stuck on the portal's pages. Whereas stickiness is good for portals for proving their popularity and benefiting their CPM (cost per thousand) revenues,[5] stickiness is not necessarily

---

[5]   CPM is a term used to describe the cost of banner ads per one thousand impressions where an impression occurs each time a banner ad is called up on a page that is visited by a user.

favorable to the advertisers who are trying to lure Web users to their own sites. This is why, for example, community sites are increasingly seen as bad for direct return on investment due to the stickiness of their services (Allen et al., 1999). Community sites offer services such as chat rooms and discussion forums. Users visit them to relax and interact with other users instead of surfing to other content and product sites. Portals continue to offer these services because they help build brand recognition and loyalty to the portal. However, in an attempt to lure users away to the advertisers' sites, commercial content infiltrates increasing parts of the services blending in as part of the forum. One example of this placement is inserting ad messages between entries made by participants in a forum. This can add to users' misunderstanding of whether they are looking at commercial content or content that is part of the ongoing forum discussion.

'Co-branding' is a common theme in many deals struck between portal companies and content-generating companies (whether online or offline ventures). I interviewed the taxonomist of one portal who reported that the company is concerned with presenting recognizable branded information. This informant made it clear that offering 'legitimate' information was a key concern of the company. This company's solution to dealing with the myriad of online information – whose legitimacy is often hard to confirm – is to create alliances with content providers that are established organizations in their fields and may already have gained legitimacy in the pre-Web era. By partnering with encyclopedia firms and corporations that provide content on sport and entertainment stars, they are reassured that the information they present is of acceptable quality. Such content keeps visitors coming to the site, while sponsorship of these pages by other companies keeps revenue flowing to the portal. A page on entertainment may contain direct links to agencies selling tickets to upcoming related events. The portal has a special deal with such ticket agencies either through payments that have already been made to the portal for prominent placement on the entertainment page or through getting a percentage of the ticket sales revenue (or both).

It is important to note that such issues as economies of scale probably also play a role in the decisions of portals to partner with other large established companies. By making a deal to offer co-branded material from a big content producer, a portal gains immediate access to thousands of records and updated information encompassing an entire topic area instead of portal company personnel facing the burden of finding information individually on all related subtopics.

In addition to co-branding and deals, there have also been a number of mergers and acquisitions in the field. It is beyond the scope of this paper to explore the mergers and acquisitions of the major portals with prominent corporations of more traditional media. Nonetheless, it is important to realize that large established media firms are increasingly present in the online market suggesting increasing market concentration and convergence of the different mass media industries (Evans, 2000). This point is relevant to the later discussion of media content diversity. Such strategies are especially common when a portal tries to expand into offering more services to users. By acquiring MailCity, Lycos was able to offer free email accounts to its users. Yahoo significantly increased its traffic by making the most popular free Webpage provider, Geocities, part of its network.

6. Diversity in media content

Of ongoing interest to sociologists of culture is the relationship between market concentration and homogeneity of cultural products. This question may be of particular interest to people in positions to formulate public policy toward arts and culture. In so far as cultural diversity is to be encouraged, seeing whether the quickest spreading communication medium of our times is supportive of a diverse cultural environment is essential in understanding the future possibilities for diverse viewpoints in the online world.

A number of theories address the question of media content diversity. The pure economic model suggests that in the absence of government intervention – as is the current case in the online marketplace of ideas – the free market will meet the demands of users. If there is demand for diversity, then market mechanisms will make sure that the suppliers of content provide the desired diversity.

Theorists who have argued that there is a general tendency toward homogeneity in markets challenge such assertions. Harold Hotelling's work (1929) suggests that new entrants in a market are more likely to present products that resemble existing players than to deviate significantly from them. The motivation underlying this tendency toward sameness is that all producers of commodities (in our case directories of cultural products online) are interested in capturing as large an audience as possible. Their chances of doing so are increased if they do not deviate too much from existing successful products and focus on a common denominator instead.[6] For our purposes, this theory suggests that we will not find too much diversity in online cultural content across portals.

Peter Steiner (1952) looked at how competition in radio broadcasting influences diversity in that medium's content. Steiner suggested that the number of competitors in a market and the relative size of audience interests determine whether the producers of content will meet the variety of interests. Although the assumptions of his model are somewhat stringent, they are helpful in understanding the basic idea behind his centrism model. Assuming that each portal can only provide one type of content and each user is only interested in one type of content, the model suggests the following. By trying to attract the largest possible audience, portals will try to satisfy the interests of the majority. They will not bother catering to minority views because potential gains from attracting the smaller audience segment are outweighed by capturing even just half of the majority. Steiner's model also rests on the assumption that content producers will be interested in attracting the greatest possible audience. For the reasons outlined in the previous footnote, these assumptions are acceptable in this case.

---

[6] Of course, this approach assumes that all players are concerned with profit (or in this case exposure) maximization. The earlier discussion of portal strategies suggests that these assumptions are reasonable in the portal market given that these companies are all commercial ventures with goals of generating profits. As things stand, profits depend on advertising and revenue generated through e-commerce both of which seem to be correlates of the number of viewers suggesting a need for as many visitors as possible.

In a more refined discussion, Russell Neuman (1991) explores what the new media will imply for a pluralist versus mass society with respect to political communication. His main concern is with the pressures on communications homogeneity. The ideas are also useful for understanding diversity in online cultural content. Neuman presents a model that considers the interplay between volume of communication and homogeneity in content. He argues that the factors influencing level of homogeneity can be grouped into three categories: technology, culture/psychology, and public policy/economics. The overall argument is that there is a complex interplay among these variables that influences the final outcome. Neuman argues that judging from the use and implementation of past communication technologies, coupled with the economics of the communication industry in the United States, we are not likely to see a move toward much more pluralism on the Web than is already present in other media. Neuman's model is more refined than either Hotelling's or Steiner's because it considers users' psychological needs and the cultural context of the new technologies. Nevertheless, his predictions are similar with respect to the final outcome: he expects to see some level of homogeneity in the new media content market.

In contrast to these views of homogenization, Joseph Turow (1997) has pointed out that advertising strategies of the past decades suggest a move toward hypersegmentation in American society. Advertisers prefer to see customers separated into demographic, psychographic and lifestyle categories, because that allows for more targeted marketing. Although Turow's argument includes all media, he specifically mentions the ease of such fragmentation thanks to detailed information provided about consumers by interactive technologies, e.g. the Web. This approach suggests that portal sites will pick particular audiences to concentrate on by specifically catering their services to various segmented groups. Instead of featuring a broad range of content, they will concentrate on particular interests and styles only. All major portals offer customization providing the opportunity for the 'daily me' (Negroponte, 1995) such as 'My Yahoo' or 'My Lycos' with content tailored to one's interests (see Appendix for details).

Glenn Carroll (1985) offers yet another perspective by considering how oligopolies leave room for entrants catering to marginalized interests. In his study of the newspaper industry, Carroll demonstrated that increased concentration in the local newspaper market leaves room for and leads to an upsurge in specialized publications. This theory implies that although main portals may not offer much diversity in content, they will be complemented by smaller categorization services that cater to special interests. This approach may be seen as a combination of the two others where a few big portals with relatively homogenous content are supplemented by portal sites catering to specialized interests.

Innovation and diversity in content has also been an object of inquiry in the production of culture literature among sociologists. Diversity in media content has been a topic for discussion regarding cultural goods such as popular music records (Peterson and Berger, 1975; Lopes, 1992), regional theatre performances (DiMaggio and Stenberg, 1985a,b), and opera performances (Martorella, 1977). Work on the popular music market by Peterson and Berger (1975) suggests that high market concen-

tration leads to homogeneity while low market concentration encourages diversity. Lopes (1992) followed up this study and concluded that even oligopolistic markets with high concentration can produce innovation and diversity in popular music because companies are concerned with reinvigorating the selections they offer in order to minimize the risk of losing consumers with interests not met by market products. This finding is in opposition to the homogenization models presented earlier in this section and suggests that diversity may be expected even in oligopolistic markets.

DiMaggio and Stenberg looked at diversity (1985a) and innovation (1985b) among American resident art theatres. The authors emphasize that the factors influencing the resident theatre market are different from those pertaining to commercial culture production. Nonetheless, given the novelty of the portal market and the recent entry of visible commercial interests into the industry, it is useful to see what their findings suggest for portal site content diversity. They find that theatres' innovation declines as their reliance on market contributions and dependence on earned income goes up. Under such circumstances, theatre groups tend to innovate less and choose more conventional pieces that have proven their popularity and are more likely to guarantee large audiences. Regarding online content diversity, these findings suggest that as portal companies become completely dependent on outside advertisers and retailers for revenues, they will favor content that caters to the mainstream over very specialized material to guarantee the widest possible popularity with Web users. In addition to pointing out the influence of market factors on cultural diversity, this study is especially useful in guiding our attempts to measure cultural diversity online. The following section presents some ways in which we can operationalize diversity in online content and thereby test the propositions brought forth in the theories discussed above.

7. Online content diversity

As mentioned earlier, 20 to 50 percent of Web users visit the most popular portal sites (Media Metrix, 1999). This gives portals a potentially significant role in allocating user attention to online content. We know that individual search engines index no more than one fourth of Web content (usually even less than that) and even collectively the largest ones only account for a combined coverage of about 42 percent of all Web pages (Lawrence and Giles, 1999). Other services on portal sites – such as directory categories – may alleviate this problem by offering direct links to Web sites. With respect to Web content diversity, the question is whether the possible diversity of information present in the aggregate among over one billion Web pages on the Internet is reflected on the most popular navigational sites?

To understand these sites' role in encouraging and fostering a diversity of viewpoints (or not), it is important to be aware of the mechanisms by which users are exposed to the existing content available on the Web. By having outlined the ways in which users find a page and by having explored what motivates the type of content portal sites display, I have presented the mechanisms that mediate the accessi-

bility of information to Web users. The next step is to examine empirically the diversity of content present on portal sites and thus test the relevance of media theories presented in the previous section with respect to Web content. It is beyond the scope of this paper to engage in such detailed content analysis. Some methods of inquiry are suggested for future undertaking, however.

One possible method of testing diversity is to pick some topics and see how easily – and if at all – pages featuring such content can be found on portals. Possible site topics include the Web page of an art gallery featuring works by Native Americans, art pages including nudity in art, a child's collection of her drawings, music by unknown groups, poetry about specific topics, or a site featuring film critiques of opposing views. Protocols can be designed to simulate different levels of expertise regarding search technique. Testers would be briefed on these differing techniques and would use them to see what type of content they can locate with their use. A similar, but somewhat different approach would be to select specific Web pages and see whether it is possible to arrive at them with the use of the various navigational sites, again, using the different levels of search expertise. However, this method poses the problem of selection bias. How else can we know to include a page in our survey without having found it via a portal site and thus guaranteeing the page is reachable via at least one of the megaportals? One way of sidestepping this problem would be to use very skilled search techniques excluding megaportals to find the chosen pages. Also, the person selecting the sites to be included in the study should not be among those who test whether the site is reachable through megaportals.

Instead of looking for specifics, content analysis of portal sites can be employed to see how diverse a selection they offer. One method would be to create a list of all sites in the category directories of portals. This list would include subcategories until depth *n* where *n* is the number of links that were followed through down the category hierarchy in order to reach a subdirectory. In addition to coding subcategories, specific site recommendations in the various categories should also be recorded. Because every category is split into numerous subcategories, this analysis will lead to quite a high number of subcategories. It may be necessary to randomly select some subcategories for detailed analysis.

It will also be interesting and revealing to see whether the subcategories and specific site recommendations are identical across portals. A comparison across navigation services will offer two ways of understanding content diversity on portals. On the one hand, we will be able to analyze whether the sites presented by a portal represent a diversity of styles and viewpoints in and of themselves. On the other hand, we will also see whether there is variation across portals. If the different navigation sites feature similar Web pages, it suggests that the overall coverage of Web content on portal sites is even smaller than expected. In so far as even all portals taken together can only be capable of offering links to a small percentage of the over one billion available Web pages, lack of diversity across portals would suggest an even smaller coverage.

It should also be noted that whether we find diversity in content also depends on how we measure it. Studies of content diversity in newspapers have come to contrasting conclusions largely due to their different ways of operationalizing the con-

tent variable (Lacy and Simon, 1993). It will be important to be sensitive to this issue in order to sidestep such problems. I suggest three ways in which content diversity may be operationalized.

One way to compare diversity in content is to numerically compare the accessible options for products within cultural industries. With respect to Web sites featuring painters' artworks, for example, we may look at how many links to painters' works are featured on the various portals. This analysis is mainly suited for comparison across portals. An option that allows for measuring diversity even within one portal looks at the size and type of the sponsor/publisher/creator of particular Web pages that are accessible via the portal. Is a group or individual responsible for the Web site and is it sponsored by a for-profit corporation, a non-profit entity, or an individual?

An alternative is to employ information on the universe of options in a given category. If we were to look at the diversity of musical content accessible via portals, for example, we could use existing musical genre categories to guide our analysis. We would start out with a long list of musical genres and see (1) how many of them are reflected on various portals; and (2) what percentage the various genres account for among the referenced pages. Analyses of other cultural industries have looked at the level of duplication to determine diversity (Steiner, 1952; DiMaggio and Stenberg, 1985a). The idea here is that the level of duplication can be defined by the difference between the number of portals and the site types presented on them. Duplication exists if the same types of sites ('types' understood as the different categories suggested by the defined universe) are present or most common on two or more portals. Alternatively, the duplication of specific sites instead of genres can also be tested. In this case, duplication is present if two or more portals feature the same site. In another example, regarding museum and gallery pages, we could take a list of all existing museums and galleries and see how many of them are easily accessible through portal sites. The limitations of this approach lie in the fact that we have to go with preexisting notions of content possibilities and a preexisting group of sponsor establishments. Online Web museums without 'real-world' equivalents would not be included in the universe and thus may be excluded from coding. This would prevent them from counting toward diversity even though they may easily serve as outlets for more artistic representation and thus be important measures of content diversity. This example shows that the details involved with measuring online content diversity set up new challenges in understanding the production of culture and gatekeeping in this new media landscape.

8. Policy implications

The quotation in the beginning of this paper highlights the importance of attention scarcity that is especially relevant to a communication medium with as much information as the World Wide Web. Portal sites define their roles as services that help users navigate the Web. For example, Excite Inc.'s company report states that the service is "[d]esigned to help consumers navigate the Web ..." (Excite Inc., 1999: 3). Portals see it as their strength and asset to deliver content to the user in an orga-

nized and assimilable manner. Given the abundance of information coupled with attention scarcity, it is understandable that users turn to such sites for help in locating content. Accepting that it may be necessary for users to rely on categorization services, and recognizing that commercial interests underlie the most popular navigation services, we must ask the implications of such commercial presence on these important Web sites.

This paper has presented the ways in which commercial interests influence what content portals decide to feature on their pages. Content produced by entities with large enough budgets can attain prominent placement on widely visited portal sites. This privileges such content over material by smaller and less financially-endowed creators. Depending on the diversity of information accessible through these sites, including to what degree they give voice to minority viewpoints, it may become necessary to create a large-scale non-profit directory system that is aggressively promoted to users. Although the Web may host information on every imaginable topic making a diverse set of content *available*, this does not automatically mean *accessibility* to all such content. If users have no way of getting to these diverse types of content, how valuable is the mere existence of such diversity?

It may become necessary to create a publicly or philanthropically funded portal site to insure quality 'programming' with sufficient educational materials and limited commercial input. Public broadcasting proved to be an important complement to commercial broadcasting on radio and television. Similarly, public Web organization may also prove to be an essential component in fostering the egalitarian and democratic nature of the World Wide Web that the medium was heralded for in its earliest years.[7] Empirical studies will lead to a better understanding of how the most popular portal sites organize the content of the Web. Such data will highlight what types of information are most accessible to the majority of Web users and thus whether there may be an eventual need for a strong non-profit presence in the portal-site industry. An important caveat here is that in so far as commercial portal sites are the primary avenue through which most users enter and explore the Web, it is likely that non-commercial sites will be harder to find than commercial ones. This suggests that the mere creation of a non-profit portal is not enough but has to be complemented by an aggressive campaign to let users know of its existence.

Another policy aspect concerns users' knowledge of online resources. Informed users will better understand the commercial nature of site content and thus can better judge for themselves whether what they are presented with is the type of information they were looking for. Moreover, by educating Web users about search strategies, their reliance on central classification systems may be reduced. A 1998 study analyzing almost one billion queries on the Altavista search engine showed that in 85 percent of the cases users only viewed the first screen of results and 77 percent of the sessions only contained one query (Silverstein et al., 1998). This sug-

---

[7] Such attempts at organizing Web content do exist. The Open Directory Project (www.dmoz.org) is one example where volunteers compile the list of sites featured in the directories of the navigation service. However, it is owned by America Online and is not nearly as popular as any of the top commercial sites, presumably at least in part because people do not know about it.

gests that users heavily rely on portal sites for presenting them with information rather than using sophisticated search strategies to fine-tune their queries. This implies that information prominently displayed on portal sites – whether selected because of high content value or for commercial reasons – has a good chance of being the destination of visitors.

There is much debate on what the Internet can do for education. However, there has been little discussion of what our existing educational institutions can do for training students to be more conscious Internet users. Merely insisting on wiring all classrooms is not enough in achieving equal access to online information. It is important to insure equally skilled access to diverse content so the Internet can be an empowering and educational resource. In addition to school training, encouraging sites to display search strategy options more prominently can also help to alleviate the problem of uninformed users. A more informed customer-base is not necessarily in opposition with the goals of commercial interests. By better understanding what exactly customers are looking for, commercial sites would have more refined data on their audience. However, more sophisticated search techniques would also benefit users by allowing them to have more say in what content they locate and view on the network. This would better allow them to take advantage of the diversity that the vast content resources of the Web have to offer instead of relying on the selection of sites picked for them based on commercial interests.

9. Conclusion

The goal of this paper has been to call attention to the tension present between the abundance of information available on the World Wide Web on the one hand, and the attention scarcity on the part of users on the other. Moreover, the implications of this opposition for the diversity in online content were discussed. By comparing this communication medium to other cultural industries, I have argued that portal sites are the gatekeepers between creators and consumers of content. A frequently quoted phrase by online media experts emphasizes that if on the Internet "content is king, then distribution is King Kong" (Green, 1998). This shows that distributors of online content are aware of their role in allocating user attention to Web content. Although production may be more within the reach of many users, distributing information to a large public and thereby calling people's attention to content is key in gaining an audience.

Features of the new medium – the low costs of storing, replicating and distributing cultural goods – have changed the locus of the gatekeeping activity in the production of online cultural products. The change in the location of the main gatekeeping activity, i.e. the ability to allocate user attention to products, requires that we reevaluate how we analyze this cultural industry in comparison to others. I have suggested specific methods for empirically testing the propositions brought forth regarding diversity in online cultural content. The paper has set the framework and justification for future detailed analyses on the cultural content presented on portals. Systematic empirical follow-up is necessary to understand the actual extent to which online gatekeepers mediate the level of content diversity accessible to Web users.

Appendix A: Services offered on portal sites

1. Web-based email and personal Web page
Portals offer users free personal Web-based email accounts on the site with addresses refer-
ring to the site name, e.g. eszter@mailcity.com (MailCity is the email site of the Lycos Net-
work. It used to be a separate service that was acquired by Lycos in 1998). Portals also allow
users to set up their own personal homepages, e.g. http://eszter.tripod.com. Lycos acquired
two of the most popular such services in 1998: Tripod and Angelfire. Yahoo! acquired the
most popular such service, GeoCities, in 1999. Users get several megabytes of space to put
up their own content on the Web.

2. Personalized welcome page
The user is given the option of personalizing their Welcome page to the site. The 'My Lycos'
feature allows the user to enter their zip code and get geographically specific information
(e.g. weather forecasts, local events). It is also possible to select specific news items (e.g.
international, business, technology, sports, etc.), get information about specific stocks and lot-
tery, set up one's own calendar and have a list of favorite links. Similar 'My' versions are
offered on other portals and usually have the same name, e.g. My Yahoo! and My Excite.

3. Online communities
Portals offer several services to facilitate the creation of online communities. They are aimed
at increasing the quality of users' Web experiences and developing a sense of belonging that
– sites hope – results in increasing visits and an affiliation with the site that leads visitors to
return to the site. Portals offer chat services, message boards and clubs organized around a
multiplicity of topics. It is also possible to create one's own interactive services.

4. News and entertainment
Over the years, portals have added explicit content to their welcome pages. News items are
most common and are usually presented as a list of links on the side. Portals are also increas-
ing the amount of entertainment they offer directly on their sites. There are many options of
online interactive games (e.g. board games, card games, arcade games, etc.) as well as direc-
tories of music sites and lists of movie trailers. Some sites also feature live video coverage of
sports and entertainment events.

5. Shopping services
In addition to being a guide to online content, portal sites also function as a guide to online
shopping. Electronic commerce is an increasingly visible component of the Web with annual
revenues in the billions of dollars. The multitude of auction sites, classifieds sites, online
malls, and other such services can be just as overwhelming to navigate as any other part of
the Web and portals are taking advantage of being the navigational guide in this arena too.

6. Local variations
In addition to local content presented through the 'My portal' feature, local versions of por-
tals also exist for several countries in their native languages.[8] These sites are usually not quite
as evolved as their English counterparts, but they are a clear indication of the portal compa-
nies' desire to cater to an audience beyond the United States. According to one report, portals

---

[8] Some portals, such as Yahoo!, even offer local variations for some American cities.

are equally prominent in other countries such as Australia, Canada, France, Germany and the United Kingdom (Media Metrix, 2000).

References

Allen, L, C. Charron, S. Shindler, T. Grimsditch and A. Shepard, 1999. Cashing in on community. The Forrester Report, September.

Carroll, Glenn, 1985. Concentration and specialization: Dynamics of niche width in populations of organizations. American Journal of Sociology 90(6), 1262–1283.

Censorware, 2000. Size of the web: A dynamic essay for a dynamic medium.
http://censorware.org/web_size/ (last viewed February 3, 2000)

DiMaggio, P. and K. Stenberg, 1985a. Conformity and diversity in American resident theaters. In: J.H. Balfe and M.J. Wyszomirski (eds.), Art, ideology, and politics, 116–139. New York: Praeger.

DiMaggio, P. and K. Stenberg, 1985b. Why do some theaters innovate more than others? An empirical analysis. Poetics 14, 107–122.

Evans, Jim, 2000. The web's gatekeepers. The Industry Standard, January 17.

Excite Inc., 1999. Annual report pursuant to section 13 or 15(d) of the securities and exchange act of 1934 for the fiscal year ended december 31, 1998.

Gieber, Walter, 1964. News is what newspapermen make it. In: L. Dexter and D. White (eds.), People, society, and mass communications, 172–182. Glencoe, IL: Free Press.

Goldhaber, M.H., 1997. The attention economy and the net. First Monday.

Green, H., 1998. The skinny on niche portals. Business Week, October 15.
http://www.businessweek.com/1998/43/b3601141.html

Hargittai, Eszter, 1999. Weaving the Western Web: Explaining differences in internet connectivity among OECD countries. Telecommunications Policy 23(10/11), 701–718.

Hirsch, Paul M., 1972. Processing fads and fashions: An organization-set analysis of cultural industry systems. American Journal of Sociology 77: 639–659.

Hotelling, Harold, 1929. Stability in competition. Economic Journal 34(March): 41–57.

ITU (International Telecommunications Union), 1999. Challenges to the network: Internet for development 1999. Geneva.

Kaplan, C., 1999. Ruling against domain name speculator could set precedent. The New York Times CyberTimes, April 9.

Lacy, S. and T.F. Simon, 1993. The economics and regulation of United States newspapers. Norwood, NJ: Ablex.

Lawrence, S. and C.L. Giles, 1999. Accessibility of information on the Web. Nature 400, 107–109.

Lopes, P.D., 1992. Innovation and diversity in the popular music industry, 1969 to 1990. American Sociological Review 57 (February), 56–71.

Lycos Inc., 1998. Annual report pursuant to section 13 or 15(d) of the securities exchange act of 1934. Commission file number 0-27830.

Lycos Inc., 1999. Sci Fi and Lycos announce joint marketing agreement to sell star wars merchandise online. April 6. http://www.lycos.com/press/scifistore.html

Martorella, R., 1977. The relationship between box office and repertoire: A case study of opera. The Sociological Quarterly 18, 354–366.

Media Metrix, 1999a. At-work Internet users spend more time and view more content online than at-home users. Press release, April 8.
http://www.mediametrix.com/PressRoom/Press_Releases/04_08_99.html

Media Metrix, 1999b. Media metrix releases the top 50 digital at home and at work digital media and Web audience ratings for August 1999
http://www.mediametrix.com/PressRoom/Press_Releases/09_20_99.html

Media Metrix, 2000. Media metrix releases worldwide Internet measurement results for Australia, Canada, France, Germany and United Kingdom. Press release, New York, February 4.

Negroponte, Nicholas, 1995. Being digital. New York: Knopf.

Neuman, Russell W., 1991. The future of the mass audience. New York: Cambridge University Press.

Nielsen NetRatings, 1999. Internet usage statistics for the week ending October 24, 1999.

Nissenbaum, H. and L. Introna, 1999. Sustaining the public good vision of the Internet: The politics of search engines, Center for the Arts and Cultural Policy Studies. Working paper 9, Princeton University, Princeton, NJ.

NTIA (National Telecommunications and Information Administration), 1999. Falling through the net III: Defining the digital divide. Washington, DC: U.S. Department of Commerce.

Peterson, R.A. and D.G. Berger, 1975. Cycles in symbol production: The case of popular music. American Sociological Review 40 (April), 158–173.

Powell, Walter W., 1985. Getting into print. Chicago, IL: The University of Chicago Press

Silverstein, C., M. Henzinger, H. Marais and M. Moricz, 1998. Analysis of a very large Altavista query log. SRC Technical Note 1998-014, October 26.

Simon, Herbert, 1971. Designing organizations for an information-rich world. In: Martin Greenberger (ed.), Computers, communications, and the public interest, 37–53. Baltimore. MD: Johns Hopkins Press.

Steiner, P.O., 1952. Program patterns and preferences and the workability of competition in radio broadcasting. Quarterly Journal of Economics (May).

Tuchman, Gaye, 1978. Making news. New York: The Free Press.

Turow, Joseph, 1997. Breaking up America. Chicago, IL: University of Chicago Press

WebSideStory, 1999. Yahoo! Surges in popularity. Press release, December 21.

White, David, 1950. The 'gate keeper': A case study in the selection of news. Journalism Quarterly 27, 383–390.

Eszter Hargittai is a Ph.D. Student in the Sociology Department at Princeton University. Her focus is on the social and policy implications of information technologies. Her dissertation looks at the production and distribution of online content complemented by an empirical analysis of how users find information on the Web. She is also interested in the use of social network methods in the empirical study of globalization, with particular emphasis on integrating developments in international communications into a broader vision of globalization.