Chapter 7

News on the Net

The Internet reconfigures space and time. Geographic space shrinks. Readers can access local papers from around the world, while companies can aggregate interested consumers across many different communities. Product space expands. No longer constrained by shelf space on newsstands or the costs of print, individuals and organizations can offer their take on events to millions by starting a website. Instant communication contracts the time news organizations have to do research before publishing and expands the demand for more news now. The changes in technology make novelty a constant, with the distribution of data over the Internet dubbed the "New New Thing" and the type of information conveyed called the "New News."[1] Though the form of information delivery on the Internet is new, this chapter shows that familiar theories explain how information markets function on the World Wide Web.

The economic fundamentals of information on the Internet are clear. The large fixed costs of production (e.g., printing presses, television studio equipment) disappear, as do the fixed costs of a delivery system and the marginal costs of paper and ink. Network effects become more important, since the value you derive from a site may depend on how many users it has and how these users contribute to its operation. Despite these changes in production and delivery, many of the properties of information delivered via newspapers or television also apply to the Internet. Content creation can still involve the fixed costs of investigating a story and learning what has transpired over the course of an hour or a day. Attention from consumers remains scarce.[2] With the large number of outlets contending for attention, the costs of becoming noticed and remembered remain important for Internet sites. Information on the Web retains the properties of a public good: one viewer's consumption does not prevent another from logging onto a website, and once information has been provided it is hard (in an age of emails, links, and attachments) to prevent another from consuming without paying for it. News on the Web is still an experience good; to know the good is to consume the good. This means brand names and reputations will be important as signals about the potential content of sites.

How these economic factors will interact to influence the provision of news remains an open question as the Internet evolves. Many of the theories about information on the World Wide Web take the form of bets or warnings about the likely paths of evolution.[3] Organizing these ideas and concerns around the five Ws shows again how the provision of news on the Internet raises questions that are common across media technologies.

Who cares about a particular piece of information? The large fixed costs of production and delivery have traditionally limited the number of papers or channels that can profitably survive

in a local or national market. The disappearance of these costs on the Internet have led some to envision a world where a news source on the Internet becomes the "Daily Me," a collection of stories tailored to the interests of each person. This glosses over the fixed costs of content creation. Even if transmission on the Web is costless, it will still matter how many "Me's" are interested enough in a topic to justify the creation of a story about it by a news agency. It is true that once stories are created, the technology of the Internet makes it easy to rearrange and bundle these stories for an individual based on his or her interests. The drop in fixed costs will also mean that more varieties of information outlets will appear on the Internet. These niche sites mean consumers may get closer to their ideal bundle of information than if they had to consume a mass appeal product. Two potential drawbacks are that the implied fragmentation of the market may reduce common experiences and lead to the polarization of views among groups.[4] The division of the audience into smaller groups may also reduce the revenues available to each outlet to produce a high quality product. This would mean that the topics covered were closer to the interests of consumers but the depth of coverage might be limited by the number of viewers.

What are they willing to pay to find it, or what are others willing to pay to reach them? An early analysis of intellectual property in the digital age proclaimed, "Information wants to be free"[5] While consumers may in the short run prefer a price of zero for information, in the long run the question arises of how free information would get produced. The search for revenue from information leads back to two familiar income sources, advertising and subscription. The use of advertising reintroduces the importance of who cares about particular pieces of information, since advertisers vary in their willingness to pay for particular demographic groups. News sites on the Web often require registration so that they can describe the zip code, age, and gender of their users to advertisers. The question of how much consumers are willing to pay for information on the Web leads back to the four types of information demand identified by Downs (1957). Consumers may be willing to pay for information related to entertainment demands, consumer purchases, or business decisions. For data that would help in making voting decisions, the logic of free riding implies that most will not express a demand for civics information and policy details. Note that these considerations of revenue may play less of a role on sites where content is created by consumers. If people get utility from self-expression on the Web, then content creation costs are reduced. If the ratings of information by others helps a viewer sort and screen data, again the need for revenues to cover editors and managers may be reduced.

Where can media outlets or advertisers reach these people? The Internet in part changes the meaning and importance of "where" a consumer is. The ability of a website to draw viewers from across the globe expands the varieties of news products available to a viewer. While fixed costs may limit the varieties of newspapers or channels available in a given area, a consumer with interests relatively rare in a particular locality may find the desired information combination in cyberspace. The low fixed costs of website operation and potential for aggregating like-minded individuals from many different areas or countries implies great

variety in news provision on the Internet. The prospect of niche outlets can also raise the specter of reduced common experiences and increased group polarization. If specialized sites do not expose consumers to opposite or unexpected views, learning or tolerance may be reduced.

Where the consumption of others takes them on the Net also influences the options available to a particular user. Internet sites are subject to network effects. The value of a site to an individual may grow as more people use the site. More users can increase content variety when sites are interactive, improve content targeting where the decisions of others provide information on likely tastes for stories, and provide more basis for discussion of common viewing experiences. A greater number of users for a site can translate into better quality and better recognition, which in turn reinforces greater use of the site. Where others have been on the Internet also matters in information cascades. When individuals use the choices of others as a short cut to reduce the costs of making their own decisions, then the Internet can give rise to cascades. News sites without significant reporting resources can contribute to cascades by simply repeating rumors or stories without independent investigation. Individuals can also propagate stories on the Net simply by repeating the statements of others.[6]

When is it profitable to provide the information? Though the technology of the Internet reduces delivery costs, creating content for a site can create significant costs. Traditional media firms, however, already pay the costs of hiring reporters and investing resources in developing stories for their newspapers, programs, or magazines. For these companies, putting stories they have already generated for other media on the Web gives them another way to earn revenues. This means that news on the Web may involve repackaging rather than original reporting. Established media companies also have an advantage in brand name recognition. Faced with limited time, viewers may forgo the costs of searching out new sites and rely on the reputation of established companies as a short cut to finding reliable news on the Web.

The existence of news sites on the Web does not imply profitability. In the late 1990s investors were willing to make large wagers on the potential profitability of many different types of Internet sites, including those providing news and public affairs information. Analysts predicted that Internet markets would behave like winner-take-all markets, which meant that gaining market share early on would translate into survival and then dominance down the road. The willingness to invest in sites promising political information took on the air of a speculative bubble. Investors believed the sites would generate revenue because other investors believed the sites would prosper. When the bubble burst for Internet companies, news and politics sites deflated as quickly as counterparts providing other services.[7] Sites with original political content do remain on the Web, but they may survive on the support of patrons, foundations, parties, or interest groups.[8]

Two traditional sources of revenue, subscription and advertising, have brought mixed success to Internet information sites. The *Wall Street Journal* has garnered more than 450,000 paid subscribers by charging annual online access fees of $29 for its newspaper subscribers and $59 for those who do not subscribe to the paper edition. Despite the large number of

online subscriptions, the site has failed to make money (except for one month) over the course of five years. Companies charging consumers for pornographic content on the Web earned an estimated $1 billion in revenues in 2000. Sites devoted to covering politics, society, or the media have generally failed in their attempts to sign up paying subscribers.[9] Though banner ads and pop-up displays attest to the willingness of advertisers to pay for the attention of Internet users, expenditures on Internet ads by companies with major advertising budgets remain minuscule. Firms apportioning advertising dollars across direct mail, television, newspapers, radio, and the Internet have many opportunities to reach consumers in larger numbers than those currently offered by most Internet sites.[10]

Why is this profitable? The Internet is often seen as a communications medium best left unregulated by the government. Yet the functioning of information markets on the Web already relies on property rights established by laws and inter-preted by regulators and judges. The ease of copying and accessibility of the Web have created disputes about how intellectual property rights should be defined and enforced on the Internet. Should sites allow individuals to post full copies of news stories from other outlets for discussion? Does a story pulled by a company from its electronic archive constitute a new publication that should generate additional revenues for its freelance author? Does the streaming of continuously updated scores from sporting events constitute an infringement on broadcast rights purchased by another media company? These are the types of disputes which have generated court cases about how information is transmitted in the Internet era.[11] Some analysts believe that much information should be defined as in the public domain, a move that would reduce the costs to consumers of accessing information. Others stress the need to protect the rights of producers in order to maintain long-term incentives for creating content. Disputes between consumers and producers about copying and conveying content on the Web will continue to affect how information is consumed and what types of information provision will be profitable.

This chapter does not try to predict the future of news on the Internet. Instead, I have taken snapshots of the Internet in 1999 and 2000 in order to examine questions about who uses the Web for news, how expression is concentrated or dispersed online, and what types of ideas fare well in information markets on the Internet. I use four different types of information: 1) survey data from the Pew Center; 2) information on the popularity online of the top one hundred daily papers in the United States; 3) measures of the relative incidence of authors, campaign contributions, and websites across a random sample of five hundred zip codes; and 4) statistics on Web searches for and advertiser support for soft news, hard news, and consumer or producer information. The results in general show that access to the Web is broader than for other forms of expression. Use of the Internet is correlated with income, but this correlation is lower than that for other forms of political expression. Concentration of attention is higher for the Internet than some other forms of expression, consistent with the impacts of network effects, brand names, and aggregation of interests across many localities. The relative demand expressed for information on the Web shows patterns familiar from other media. Soft news and “news you can use” often generate more interest than hard news.

Differences in information preferences by age and gender also appear in the types of news viewers seek on the Internet.

News Online and Online Newspapers

If information were to be had for the asking, what questions would you seek to answer? Would the data relate to daily life (What will the weather be like tomorrow?) and its diversions (Did my team win last night?)? Would the news relate more to society (Will a Patients' Bill of Rights pass?) or to the world (Does a recent study provide evidence of global warming?)? The Internet offers information on all of these questions, with most sites offering their data for free. Yet viewers still face the opportunity costs of attention. Time spent consuming one type of information is time that cannot be spent on another task. This means the relative interest of viewers still plays a large role in the types of news demanded by news consumers on the Internet. Given the choice between the mundane and the momentous, individuals on the Net will prefer news they can use in their daily life to reports about significant political policies and questions. Survey data from the Pew Research Center for the People and the Press (2000b) confirms that the hierarchy of news interests evident in Internet news consumption is similar to the patterns found in print and broadcast news consumption.

In the Pew survey on media consumption conducted in 2000, Internet use was significantly influenced by age and gender.[12] Table 7.1 shows that over 70% of males and females 18–34 went online to access the Internet, Web, or email. Among males 50+, only 39.9% went online. Women 50+ had the lowest rate of Internet use, just 27.8%. Of those who are online, men are much more likely to go online for news (a pattern similar to the greater use of newspapers by men than women). Among online users 18–34, 55.0% of women went online for news at least one to two times per week versus 70.0% of men. For those women online, a similar percentage in each age group went online for news. The same is true for men. In the online community, 70.0% of males 18–34 went searching for news online at least one to two days per week versus 67.2% of males 50+.

When they are online consuming news, men and women explore topics in a fashion predicted by Downs's theories of information used for consumption, production, entertainment, and voting. News one can use in everyday life tops the interests of online consumers. Table 7.1 indicates that among nine categories of online news, for women news about science and health has the highest percentage of users. Among females 18–34, 65.8% go online for science and health news. For women 50+ the figure is 70.6%. This mirrors the strong interest in health news by women documented in chapter 3. Among males 18–34 the top category of online news is technology news, which attracts 71.5% of online news consumers in that demographic. Technology news is also first among males 50+ (71.0% go online for this) and second among males 35–49 (72.2%). The strong preference for technology news among those online is not surprising, given that going online involves some interaction with technology. A popular category with many demographic groups among those going online for news is the weather. The

percentage of online news consumers seeking out this information ranges from 72.5% for males 35–49 to 63.4% among females 50+. In contrast, political news ranks near the bottom for each demographic group. Political news ranks eighth out of nine for all female demographics, seventh among males 35–49 or 50+, and eighth among males 18–34.

TABLE 7.1

Internet News Consumption

	% of Column Respondents						
	Total	*Females 18–34*	*Females 35–49*	*Females 50+*	*Males 18–34*	*Males 35–49*	*Males 50+*
Go online to access the Internet, web, or email	54.0	71.0	62.8	27.8	71.6	58.5	39.9
Of those that go online . . .							
Go online daily to get news	27.0	21.2	21.3	20.9	29.6	34.0	36.6
Go online at least 1–2 days per week to get news	60.7	55.0	51.0	51.4	70.0	67.7	67.2
Of those that go online for news, go online for . . .							
Political News	39.2	37.9	28.9	33.3	42.1	48.7	42.7
Sports News	42.2	31.0	22.5	25.8	68.1	51.9	46.0
International News	45.0	41.7	33.3	40.7	55.5	52.4	42.7
Science and Health News	63.2	65.8	66.9	70.6	59.8	58.3	60.0
Technology News	58.7	42.6	46.5	49.5	71.5	72.2	71.0
Weather News	66.0	63.7	63.9	63.4	65.2	72.5	67.4
Entertainment	43.6	57.2	40.6	34.9	53.6	34.5	26.5
Local News	36.7	41.3	37.3	36.3	37.6	36.3	27.4
Business News	52.5	38.0	45.5	46.5	54.6	65.3	70.8

Many of the gender and age differences in news interests in print and broadcast audiences are also evident among online news consumers. Women are more interested in entertainment news than are men. Among female online news consumers, entertainment news was the third most popular category among women 18–34 (57.2% of online news consumers viewing this category). Among males consuming news online, entertainment news ranked last among those 35–49 (34.5%) and 50+ (26.5%). Interest in entertainment news within each gender was highest among those 18–34 and lowest among those 50+. Sports news ranked last in use among women online. For males 18–34 this was the second most popular category, with 68.1% going online for this type of news. Within every age group men were much more likely to consume sports information online than women. Among those 35–49 who go online for news, 51.9% of males went online for sports news versus 22.5% of women.

Business news was consumed in higher percentages in each age group among men than among women. For males 50+ consuming online business news ranked second among news categories, with 70.8% consuming this category online. For females 18–34 consuming news online, this category of news ranked seventh (38%). In each age group online, men were much more interested in political news than women were. For those 35–49, 28.9% of female online users sought out political news versus 48.7% of men. International news ranked higher in five of the six demographic groups than political news. Among those 18–34 or 35–49, men reported consuming international news online in higher proportions than woman. Among those 50+ this

gender gap declines, with 42.7% of male online news consumers viewing international news versus 40.7% of women.

Though topics that Internet news consumers pursue are similar to those of print and broadcast consumers, the concentration of attention on the Internet is greater than in print markets. Prior to the evolution of the Internet, the relevant market for most local newspapers was local. The costs of physical distribution, remote printing, and content creation meant most papers were consumed locally. There were newspapers that had national distribution, including *USA Today*, *Wall Street Journal*, and *New York Times*. With the arrival of the Internet, however, consumers were free to search the Net and sample the offerings of papers without being limited by geography. If readers in a small or medium city in the United States did not find their interests served by a local paper, they could easily read the *New York Times* (or the *Washington Times)* on line. To investigate how this freedom affected consumption of news sites on the Internet, I examined how popular local newspapers fared on the Internet. I started with the one hundred largest newspapers measured by 1999 circulation, according to a list compiled by *Editor and Publisher*. I then entered the Web addresses of these newspapers into an Internet site that tracks how frequently other websites link to a particular address. For each newspaper I developed a link count, which I treat as a measure of relative popularity on the Internet.

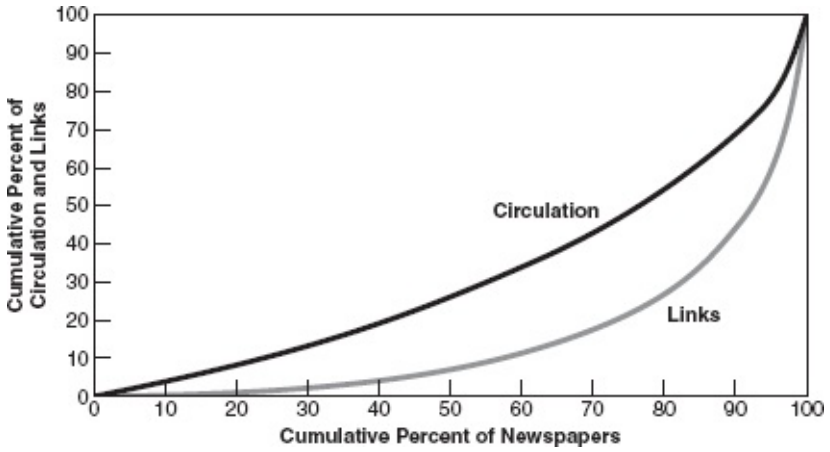

FIGURE 7.1 Relative concentration of circulation and links.

For the top one hundred newspapers in the United States, figure 7.1 traces out the cumulative percentage of newspapers that account for a cumulative percentage of circulation and links. The figure shows clearly that attention on the web is much more highly concentrated than the circulation of printed papers. The top five newspapers (*Wall Street Journal, USA Today, New York Times, Los Angeles Times, Washington Post)* accounted for 21.5% of the total circulation of the one hundred newspapers analyzed. The top five newspaper websites in the sample (*USA Today, New York Times, Detroit News, Washington Post, Seattle Times)* accounted for 41.4% of the total links generated by the top one hundred newspapers.[13] Many theories predict that attention on the Web should be more concentrated than attention in the print world. If the value of a website to a consumer increases with the use of others (e.g.,

through interactive features, polls, or better targeting of content by papers), then network effects will multiply the advantages of sites that get used early on. The variety of choices on the Internet enhances the advantages of brands established in the physical world. Internet users may go to *USA Today* or the *New York Times* in the virtual world because they already recognize their names and reputations as newspapers. The use of links can operate like an information cascade, with Internet users spreading a link through emails to friends who post the link again on their sites.

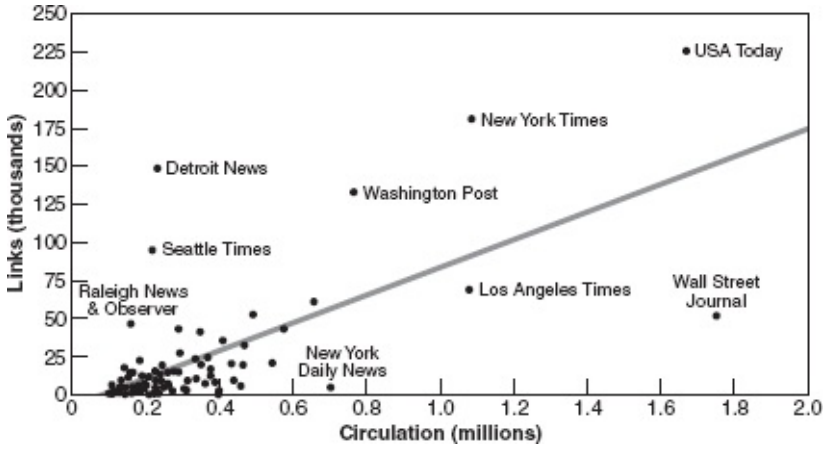

FIGURE 7.2. Newspaper popularity: circulation versus links and predicted links.

Figure 7.2 shows the early winners among newspapers on the web. The solid line represents the number of links for a newspaper website predicted by its circulation. Three of the top five newspapers in circulation fare extremely well in terms of links: *USA Today*, *New York Times*, and *Washington Post*. Their success shows that advantages in the physical world can be magnified on the Web through network and brand names effects. The link count for the *Wall Street Journal* vastly understates its relative popularity on the web for a simple reason—the *Journal* charges for access to most of its Internet content. This means that individuals are less likely to propagate links to *Journal* content, since accessing the information requires a subscription. Two of the outliers with greater than expected links are *Seattle Times* and *Raleigh News and Observer*. Both of these are published in areas with a high concentration of high-tech and Internet businesses, so it is not surprising that their readers/viewers are likely to spread links via the Internet.

Space and Speech

Speech on the Internet is cheap but not free. Talking with a friend about politics in the home or office involves opportunity costs, such as the time taken to develop your political views and the energy used to express them. Conveying your ideas on the Internet involves time spent learning, composing, and typing. Use of the Internet may also involve fixed costs such as

paying for space on a server or subscribing to a company that provides online access. These fees are still relatively small when compared to the costs of buying advertising space in a newspaper to convey your views or purchasing advertising time on television. Opening a full-fledged outlet on the Internet may involve up-front costs for servers and continuous costs for Web page design and content creation, but these still pale in comparison to the costs of setting up a newspaper or buying a radio or television station. For all of these reasons, conveying ideas on the Internet should be cheaper than in print or broadcast. The drop in costs should make Internet speech more accessible than other forms of communication.

Measuring and tracking political speech and expression at the individual level can be difficult, in part because of privacy concerns. More information, however, is available on expression aggregated at the zip code level. In this section I use the zip code as my unit of observation to study how use of the Internet compares to two other forms of speech. I start with a random sample of five hundred zip codes. To measure the use of the Internet, I entered each zip code and its state abbreviation into the search engine www.google.com. I counted the number of Web pages associated with a zip code as evidence of Internet expression. I also did a search using the zip code, its state abbreviation, and the term "news" to capture more directly Internet use related to the market for news. I used the Gale Contemporary Authors database to track another form of speech, publishing by authors. By searching this database of authors I was able to determine the number of authors who list a particular zip code in their contact information. Finally, I used the website www.opensecrets.org to develop a third indicator of speech, the amount of political contributions emanating from a zip code.[14] The Supreme Court in the 1976 decision *Buckley v. Valeo* recognized the relationship between political campaign contributions and expenditures and political expression. I used the contributions database to measure for the residents of each zip code the total political contributions to federal races in 1996 and 2000 (through July 1, 2000), including donations to the presidential campaigns of George W. Bush and Al Gore.

Table 7.2 shows that individuals from more communities participate in Internet speech than other forms of expression and that expression on the Internet is less geographically concentrated. The top five zip codes account for 6.4% of population and 18.7% of Web pages using the word "news" and a zip code in the sample. Authors and political contributions are much more highly concentrated. The top five zip codes contain nearly a third (31.3%) of all authors in the random sample of five hundred zip codes. The five zip codes with the heaviest donations accounted for 31.3% of political contributions to federal candidates in 2000, 37.7% of contributions to the George W. Bush campaign, and 39.6% of donations to the Al Gore campaign. The broader access to expression via the Internet is evident from the number of zip codes that had some expression recorded on the Internet. In the sample of five hundred zip codes, 486 turned up on Web pages and 482 turned up on at least one Web page with the term "news" In contrast, 156 zip codes had authors living in them according to the writers' database. While 345 out of five hundred zip codes registered some contributions in the 2000 federal elections data, only ninety-five had positive contributions to Al Gore's campaign. George W. Bush's campaign had broader support, generating contributions in 175 of the zip

codes.

TABLE 7.2

Measures of Expression by Zip Code (N = 500)

	Gini Coefficient	*Cumulative % in Top 5 Zip Codes*	*# Zip Codes with Non-zero/ Non-missing Data*	*Mean*	*Maximum*
Total Households, 1999	.63	5.8	500	3,763	24,967
Total Population, 1999	.63	6.4	500	9,973	81,047
Web Pages with Zip Code	.73	16.2	486	616	12,600
Web Pages with Zip Code and "News"	.76	18.7	482	129	4,258
Authors	.89	31.3	156	2	69
Political Contributions, 1996 ($)	.86	29.3	374	28,793	1,277,944
Political Contributions, 2000 ($)	.87	31.3	345	25,588	1,184,049
Contributions to George Bush, 2000 ($)	.91	37.7	175	3,001	264,179
Contributions to Al Gore, 2000 ($)	.94	39.6	95	1,001	65,250

The Gini coefficient represents another way to convey the relative concentration of expression. The calculation is meant to represent the degree of inequality in the distribution of a variable, with 0 representing equality of distribution and 1 representing total inequality of distribution. The higher the value the greater the inequality of distribution.[15] By this measure, total households and total population in the sample of five hundred zip codes had a Gini coefficient value of .63. The Gini for Web pages was .73, and for Web pages with the term "news" was .76. Authors and contributions were much more concentrated. Authors had a Gini coefficient of .89, compared to .91 for Bush contributions and .94 for Gore contributions. A final way to see the greater disparities across zip codes in authors and contributions is to compare mean values of expression with maximums. For the zip code sample, the mean number Web page listings for a zip code was 616 and the maximum was 12,600. For authors the mean number living in a zip code was 2 and the maximum was 69. While the mean number of political contributions to 2000 federal political campaigns in a zip code was $25,588, the maximum value for a zip code was $1,184,049. The ratio of maximum to mean values was thus 20.5 for Web pages, 34.5 for authors, and 46.3 for political contributions.

Expression on the Web is related in predictable ways to the demographics of zip code areas. Table 7.3 shows that Web page mentions of a zip code are positively correlated with zip population (.42 correlation). Authors and contributions are also positively correlated with population, though the relationship is less strong than for Web page counts. Median age is not correlated with author counts or contributions. Mentions of a zip code on Web pages and on pages with the term "news" are negatively correlated with median age. This is consistent with the Pew survey data showing that younger adults are more likely to log onto the Internet. Household income is positively correlated with Web pages mentions (.25), authors (.30), and

political contributions in 2000 (.35), although the strength of the relationship is smaller for Internet expression than for other forms of speech. The forms of expression are correlated with one another, so that zip codes higher in one form are likely to be higher in another. Author counts and contributions are more closely related to each other than Web page counts, again suggesting that expression on the Web is more accessible and dispersed. Political contributions in 2000 are correlated .45 with Web pages listing the term "news" and .63 with author counts. Contribution patterns remain relatively fixed over time, with contributions in 2000 having a correlation of .95 with contributions in 1996. The presidential contributions in 2000 indicate that author counts are more highly correlated with Gore contributions (.65) than Bush (.42). Note that zip codes participating in one party's fund-raising are likely to be contributing to the other party too. In 2000, contributions to Bush had a .76 correlation with contributions to Gore at the zip code level.

Table 7.4 shows that at the zip code level that population size and median household income drive the magnitude of Web expression. The larger the population in a zip code, the more Web pages that refer to the area and the greater the number of sites that refer to the zip code and the news. The higher the median household income, the larger the number of websites that refer to a zip code. Median age and percent black population do not have a statistically significant impact on these measures of Web expression. Overall, the analysis of speech at the zip code level shows that the Internet offers greater access to more areas than some other forms of expression, that this expression is less concentrated than authorship or political contributions, and that expression on the Internet is positively correlated with income but that this relationship is less strong than for other forms of speech.

TABLE 7.3

Correlations among Demographic Factors and Expression at the Zip Code Level (N = 500)

	Web Pages with Zip Code	*Web Pages with Zip Code and "News"*	*Authors*	*Political Contributions, 1996 ($)*	*Political Contributions, 2000 ($)*	*Contributions to George Bush, 2000 ($)*	*Contributions to Al Gore, 2000 ($)*
Total Population	.42***	.35***	.33***	.29***	.27***	.18***	.20***
Median Age	−.07*	−.10**	.00	.06	.05	.05	.07
Median Household Income ($)	.25***	.22***	.30***	.35***	.36***	.31***	.37***
% Black Population	.06	.08*	.07	.00	.00	-.02	.01
Web Pages with Zip Code and "News"	.87***						
Authors	.43***	.44***					
Political Contributions, 1996 ($)	.46***	.49***	.63***				
Political Contributions, 2000 ($)	.44***	.45***	.63***	.95***			
Contributions to George Bush, 2000 ($)	.30***	.28***	.42***	.83***	.85***		
Contributions to Al Gore, 2000 ($)	.38***	.39***	.65***	.87***	.87***	.76***	

Note: *** = statistically significant at the .01 level, ** = statistically significant at the .05 level, * = statistically significant at the .10 level.

TABLE 7.4

Determinants of Web Expression at Zip Code Level

	(1) Web Pages with Zip Code	(2) Web Pages with Zip Code and "News"
Total Population	0.04***	7.08e-3***
	(4.34e-3)	(1.10e-3)
Median Age	1.42	−1.83
	(11.06)	(2.79)
Median Household Income ($)	0.01***	2.76e-3***
	(3.92e-3)	(9.90e-4)
% Black Population	1.62	0.79
	(3.77)	(0.95)
Adjusted R^2	0.19	0.13
Number of Zip Codes	486	486

Note: Each specification includes an intercept. Standard errors are in parentheses. *** = statistically significant at the .01 level, ** = statistically significant at the .05 level, * = statistically significant at the .10 level.

Ideas on the Internet

Though the marketplace of ideas meets continuously on the Internet, the concept of an intellectual marketplace is only a metaphor unless one uses economic theory and data to explore its operation. The work of Anthony Downs (1957) provides a structure to examine the demand side of the market. According to Downs, people seek information to facilitate consumption decisions, improve business decisions, engage in entertainment, and make political choices. The first three types of demand lead people to seek out data, for if they do not make the effort to get the information they miss out on many of its benefits. The logic of collective action implies that most people will remain rationally ignorant about the details of politics and policy. Even if a policy might affect a person's livelihood deeply, the small probability that a person has of affecting an election, a legislator's vote, or a regulator's decision means that most individuals will not seek out data to improve their political decisions. This implies that individuals on the Net will search for information that helps them buy products, make work decisions, or entertains. The relative search for politically important and relevant information should remain small.

Exploring how these four types of information demands operate on the Net requires definitions and samples. In this section I use four sources to define information markets on the Net as they existed in May 2000. Hard news, the mix of current events reporting that emphasizes politics and government, is the type of civic information Downs predicts people will not seek out. To define hard news topics of interest in May 2000, I used the transcripts of the *News-Hour with Jim Lehrer.* At the end of this weekday evening news broadcast on PBS, the anchor provides a summary of the day's two or three top news stories. Collecting these for May 2000 yielded a sample of fifty-four different news topics. Soft news fits Downs's definition of information sought solely for entertainment, the pure pleasure of knowing. I used the people, programs, and topics that surfaced on the front page of the *USA Today* Life section in May 2000 as my guide to soft news topics. This resulted in 123 celebrities, shows, and

topics in my sample of soft news products. The GoTo website (www.goto.com) organizes searches by categories that capture consumer and producer questions: computing; education and career; finance; health; homelife; reference; shopping; travel. Using the search categories as examples of consumer or producer information yields a sample of ninety-six terms. The most popular search terms on the Internet are defined by many different sites. I chose the Lycos Top 50 as my source for popular search terms. This weekly list of the top fifty search terms on Lycos (www.lycos.com) yielded a total of sixty-four terms for May 2000.

For my snapshot of the Internet, in May 2000, I thus use four different types of information. The stories from the *Lehrer* program represent hard news topics. The people, groups, and entertainment products described on the *USA Today* Life section (front page) define soft news. The GoTo queries represent consumer and producer information demands. The Lycos data represent the most popular search terms on the Internet. Comparing how these different types of information fare in Internet markets requires rough measures of demand, supply, and pricing. To capture the relative demand for a piece of information, I used a function on the GoTo website that reported how many searches in the previous month had used the term on that search engine.[16] For each term in my four categories of information I used this function to find out how many searches had been done using the term. To measure the relative supply of the information, I used www.google.com. I entered each term in Google, which yielded a count of the number of Web pages cataloged by that engine that use the term.

To measure financial support for an idea on the Web, I used the GoTo website. GoTo allows advertisers to bid to support a search term. The advertiser with the highest bid is the first link that appears when the results of a search are reported from GoTo. The next highest bid gets the second listing in a search result on GoTo. When the list of sponsoring sites is exhausted GoTo then lists Web pages that use the term but have not paid an advertising fee. Since GoTo needs to provide advertisers with information on how much search terms are selling for, the engine lists the price paid for the link when results are reported from the search engine. Entering each of the terms in my sample into GoTo thus allowed me to explore the relative support by advertisers for different types of information on the Web. For each term I calculated the total advertisers for the term (i.e., number of links where sponsors paid for a listing in search results), the total of the advertiser bids in cents, the average advertiser bid, and the maximum of the advertiser bids (i.e., the price paid to be the top listing in a search term result).

Table 7.5 bears out the predictions that when people search for information on the Internet, they search for things that are entertaining or personally useful rather than information related to broader social or political decisions. The mean number of searches for soft news terms (11,522) was twice that for hard news topics (4,295) in May 2000. The mean number of queries (81,822) for information used in consumer or business decisions was nearly twenty times the mean for hard news topics. The differences in audience interest and value to the searcher of the information sought translate into different advertiser support for ideas on the Internet. For hard news topics, the mean number of total advertisers willing to pay to appear in search responses was 6.2. This contrasts with means of 17.4 total advertisers for soft news

topics, 36.4 advertisers for the popular terms on Lycos, and 94.4 for the consumer and business related terms defined by GoTo. Adding the bids for each term provides one way to characterize the relative advertiser support for different types of ideas. For hard news topics, the mean value of the total advertiser bids to support a term was 36.9¢. For soft news topics this figure was 93.9¢, and for Lycos terms the figure was 369.8i. The total advertiser bids for the GoTo terms was 1,901.8¢. The average and maximum bid patterns tell a similar tale. Hard news average bids had a mean of 2.5¢, versus 3.5¢ for soft news and 16.2¢ for GoTo terms. Companies that paid the maximum bid for a term earned the first spot in search term listings. The means of the maximum bid for each term were 8.0¢ for hard news, 13.0¢ for soft news, and 121.6¢ for GoTo terms.

Hard news information did dominate soft news and Lycos information on one dimension, the average number of Web pages with a search term. This at first appears counterintuitive, since it suggests that websites are more likely to offer information that is less sought after. Yet hard news topics are often covered by websites operated by entities that are not making decisions based on advertising. Government, academic, lobbying, and nonprofit websites may provide data that relate to hard news topics even though there would be little return in an economic marketplace for the information. These sites may offer data aimed at influencing votes or impressions rather than individual purchase decisions. Hard news topics had a mean of 390,467 websites using the hard news term. This is higher than the mean for soft news (174,629) and for Lycos terms (353,607). The GoTo terms have the highest number of mean websites using a term (1,092,596).

TABLE 7.5

The Market for Ideas on the Internet by Topic Area, May 2000

	Hard News Mean	*Soft News Mean*	*Lycos Mean*	*GoTo Mean*
Total Searches for Term	4,295	11,522	50,505	81,822
Number of Web Pages with Term	390,467	174,629	353,607	1,092,596
Total Advertisers for Term	6.2	17.4	36.4	94.4
Total of Advertiser Bids (Cents)	36.9	93.9	369.8	1,901.8
Average Advertiser Bid (Cents)	2.5	3.5	5.8	16.2
Max Advertiser Bid (Cents)	8.0	13.0	31.6	121.6
Ratio of Searches/Web Sites	0.01	0.11	0.95	0.10
N	54	123	64	96

	Difference of Means					
	Hard News-Soft News	*Hard News-Lycos*	*Hard News-GoTo*	*Soft News-Lycos*	*Soft News-GoTo*	*Lycos-GoTo*
Total Searches for Term	−7,227**	−46,210***	−77,527***	−38,983***	−70,300***	−31,317*
Number of Web Pages with Term	215,838***	36,860	−702,129***	−178,987***	−917,967***	−738,989***
Total Advertisers for Term	−11.2***	−30.3***	−88.2***	−19.0***	−77.0***	−57.9***
Total of Advertiser Bids (Cents)	−57.0***	−332.8***	−1,864.9***	−275.8***	−1,807.9***	−1,532.1***
Average Advertiser Bid (Cents)	−0.9**	−3.2***	−13.7***	−2.3***	−12.7***	−10.4***
Max Advertiser Bid (Cents)	−4.9**	−23.6***	−113.5***	−18.7***	−108.6***	−89.9***
Ratio of Searches/Web Sites	-0.10***	−0.94	−0.08**	−0.84	−0.01	−0.86

Note: *** = statistically significant at the .01 level, ** = statistically significant at the .05 level, * = statistically significant at the .10 level.

Dividing the number of searches for a term by the number of sites using the term puts the

demand and supply sides of the Internet together in one measure. The mean ratio of searches to websites for hard news topics is .01, which means that there was one query about the topic for every one hundred sites with information about it. The ratio is ten times greater for soft news stories (.11) and for GoTo topics (.10). For these terms, there was approximately one query about a person or topic for every ten websites using the term. The Lycos lists track search terms that are particularly popular in a given month, so it is not surprising that the mean ratio for these terms is .95. For every query about the term there was about one website using the term available. Overall, the figures in table 7.5 show that individuals search out entertaining or personally useful information more often than they look for hard news stories. Advertisers are more willing to support soft news or data used for consumer and business decisions. There are more websites that offer information related to hard news topics than soft news, in part because government and nonprofits may offer data that are not heavily demanded. Lack of discussion of hard news topics is not due to lack of supply. The ratio of searches to sites shows how audiences are much more likely to seek out data that entertain or help in purchasing decisions than information that aids in voting or other civic decisions.

Assessments of the operation of the Internet offer conflicting tales about the concentration and dispersion of attention. One strand of research emphasizes the concentration of viewership encouraged by network effects, media concentration, and information cascades. Another strand points out the diversity of expression and creation of niche markets on the Internet made possible by the low fixed costs of entry and the agglomeration of people with specialized interests across the Internet. Table 7.6 offers evidence on the relative concentration of attention to ideas, by calculating Gini coefficients for different categories of information. In terms of the questions that people ask on the Internet in search engines, soft news stories have the highest concentration of attention (measured by a Gini coefficient of .81). Searches for hard news topics (.74 coefficient) and consumer or business information (.71) are less concentrated. The supply of information is much more dispersed for information relating to purchasing decisions than for other types of data. The Gini coefficient for the number of Web pages using a term takes on a value of .35 for the GoTo terms. This contrasts with .54 for the Lycos terms and .56 for hard news topics. Soft news items had the highest concentration of offerings, with a coefficient of .64. For hard news, soft news, and GoTo information the pattern holds that audience demand for information is more concentrated than the supply of information by websites. For each of these categories of information, searches are more concentrated (i.e., have a higher Gini coefficient) than websites offering a particular type of information.

While table 7.5 shows that hard news topics attracted fewer advertisers and lower total or average bids, table 7.6 indicates that the advertising that hard news does manage to attract is more concentrated than the support provided for other types of information. The Gini coefficient for the total advertisers supporting a term is .66 for hard news, compared to .55 for soft news and .47 for GoTo. The total advertiser bids are also more concentrated among the hard news terms. The Gini coefficient for total advertiser bids was .78 for hard news stories, versus .68 for soft news topics and .60 for GoTo. These results indicate that overall soft news or information that helps consumers and producers make decisions enjoy greater advertising

support than information that aids in social or political decisions. The willingness of advertisers to sponsor links for hard news is less robust than for soft news, whether the measure used is total advertisers, average bids, or dispersion of support across topics within a news category.

TABLE 7.6

The Relative Concentration of Attention to Ideas on the Internet by Topic Area, May 2000

	Gini Coefficients by Sample			
	Hard News	*Soft News*	*Lycos*	*GoTo*
Total Searches for Term	0.74	0.81	0.48	0.71
Number of Web Pages with Term	0.56	0.64	0.54	0.35
Total Advertisers for Term	0.66	0.55	0.63	0.47
Total of Advertiser Bids (Cents)	0.78	0.68	0.79	0.60

Table 7.7 examines the markets for different types of information by looking at the top five terms for each of the measures of Internet activity. For total searches conducted using a specific term, the top item in the soft news category was pop star Britney Spears (who generated 357,818 searches in May 2000 on GoTo). The top item in the Lycos sample was Pokemon, the trading card cartoon figures favored by children (387,147). The GoTo term with the largest number of searches was computers (736,113), which is not surprising given that the searches were done by people using computers. The top five stories in the soft news category predictably relate to entertainment (e.g., a singer, band, movie). The top five hard news stories include terms that individuals might use in business or consumer searches, such as “Cuba” or “computer virus.” Those searching for “Social Security” may be more interested in pursuing payments than debating policy. The Lycos top five tend toward the interests of young consumers (e.g., trading cards, two singers, and tattoos); the exception might be “flowers,” which was a popular term in the weeks surrounding Mother’s Day. The GoTo terms related to consumer or business decisions, including those involving computers, travel, health, software, and Web hosting. Note that the top five searches in the GoTo category are at least twenty times greater than the corresponding figures for hard news stories. “Computer virus” earned the fifth place on the hard news list with 14,877 searches, while Web hosting garnered the fifth place in the GoTo set with 426,597 searches.

TABLE 7.7

Relative Popularity of Terms on the Internet, May 2000

	Top 5 Terms by Sample			
	Soft News	*Hard News*	*Lycos*	*GoTo*
Total Searches				
1.	Britney Spears (357,818)	Social Security (31,557)	Pokemon (387,147)	Computers (736,113)
2.	EBAY (117,214)	Memorial Day (18,549)	Britney Spears (357,818)	Travel (671,149)
3.	Napster (100,785)	Cuba (18,039)	Flowers (185,820)	Health (434,322)
4.	Oasis (76,791)	Crime (16,322)	Tattoos (154,760)	Software (434,321)
5.	*Gladiator* (49,941)	Computer Virus (14,877)	Eminem (104,582)	Web Hosting (426,597)
Number of Web Pages				
1.	*X Files* (1,689,999)	May Day (3,180,000)	Golf (1,556,999)	Family (2,920,000)
2.	*South Park* (1,190,000)	Crime (1,150,000)	The Big Game (1,440,000)	News (2,439,991)
3.	*Party of Five* (1,069,995)	President Clinton (979,997)	Las Vegas (1,260,000)	History (2,370,000)
4.	Prince (1,060,000)	Social Security (967,993)	*South Park* (1,190,000)	Books (2,339,999)
5.	James Brown (985,000)	Supreme Court (818,000)	Flowers (1,180,000)	Computers (2,309,998)
Total Advertisers				
1.	*Star Wars* (113)	Cuba (40)	Flowers (238)	Books (242)
2.	Britney Spears (90)	Computer Virus/Crime (27)	Golf (237)	Gifts/Health/ Marketing/ Shopping/ Software/ Wedding (240)
3.	Vitamin C (80)		Baseball (154)	
4.	N Sync (64)	George W. Bush/ Social Security (23)	Mother's Day (129)	
5.	*Harry Potter* (57)		*Star Wars* (113)	

Total of Advertiser Bids (Cents)				
1.	Vitamin C (831)	Crime (369)	Flowers (4,402)	Web Hosting (13,262)
2.	*Star Wars* (718)	Cuba (290)	Golf (3,867)	Gifts (12,306)
3.	Britney Spears (664)	Computer Virus (233)	Baseball (2,175)	Marketing (8,058)
4.	*Harry Potter* (531)	Social Security (163)	Pokemon (1,370)	Website Design (7,544)
5.	N Sync (525)	George W. Bush (107)	Mother's Day (1,298)	Shopping (6,851)
Average Bid (Cents)				
1.	Vitamin C (10.39)	Crime (13.67)	Flowers (18.50)	Web Hosting (55.96)
2.	Metallica (9.50)	Computer Virus (8.63)	Golf (16.32)	Gifts (51.28)
3.	*Harry Potter* (9.32)	China Trade (7.42)	Baseball (14.12)	Debt (48.80)
4.	N Sync (8.20)	Cuba (7.25)	Las Vegas (14.09)	Health Insurance (41.17)
5.	Faith Hill (8.10)	Korean War (7.13)	Pokemon (13.70)	Home Finance (40.26)
Maximum Bids (Cents)				
1.	Vitamin C (79)	Social Security (64)	Flowers (215)	Web Hosting (590)
2.	Metallica (68)	Crime (61)	Golf (205)	Small Business (554)
3.	Britney Spears (51)	Computer Virus (41)	Baseball (121)	Insurance (436)
4.	*Star Wars* (48)	Cuba (33)	Mother's Day (117)	Marketing (433)
5.	Ricky Martin (43)	Supreme Court (21)	Backstreet Boys (87)	Debt (415)

Note: Counts of dimension ranked are in parentheses (e.g., Social Security (31,557) in the total searches ranking indicates there were 31,557 searches for that term).

On the supply side, the top items in hard news are generally smaller than for other categories (with the exception of May Day, which garnered 3,180,000 mentions on Web pages). The top sites in soft news relate to television shows and music performers. Web pages referring to the cartoon program *South Park* (1,190,000) outnumber those that discuss crime (1,150,000). References to the teen soap opera *Party of Five* (1,069,995) are more frequent than to President Clinton (979,997). The soft news figure with the fifth highest count of Web page references (James Brown, 985,000 sites) generates more citations than the fifth item in the hard news category (Supreme Court, 818,000 sites). The figures for soft news, hard news, and Lycos terms are generally much smaller than the top terms in the GoTo section (family, news, history, books, and computers).

The desire of advertisers to support terms used by those making purchase decisions is evident in the top five terms for total advertisers or bids. The GoTo terms that attracted the most advertisers were "books" (242 advertisers willing to pay for links), and "gifts," "health," "marketing," "shopping," "software," and "wedding" (each of which had 240 advertisers). The terms that attracted the most advertisers in soft news related to movies *(Star Wars*, 113 advertisers), music (Britney Spears, 90; N Sync, 64), books *(Harry Potter*, 57), and health

(Vitamin C, 80).[17] The advertisers sponsoring hard news topics were more likely to be companies trying to sell products related to these topics rather than interest groups or parties trying to influence opinions or votes. The hard news topic with the most advertisers was Cuba. The forty sponsored links generally were from companies selling travel packages to Cuba or books about Cuba. The links for Social Security related to service companies offering to help a person obtain benefits or detective agencies offering to trace individuals via social security numbers. Links to the hard news topics were often related to the purchase of products rather than the discussion of public policies.

Advertisers on GoTo pay their bid price when someone clicks through to their sites after viewing the results of a GoTo search. The prices companies are willing to pay to be associated with different terms on this search engine reveal how strong the returns are for providing entertaining or personally useful information compared to the returns for providing analysis of hard news. Table 7.7 offers three ways to analyze prices paid by Internet advertisers. If one sums the total ad prices from sponsored links for each term, the top terms for total potential ad revenue are similar to the rankings for total number of advertisers. The soft news term "Vitamin C" had a total of 831¢ in advertising bids, compared to 369¢ for the top hard news, crime. The support for the top Lycos search term was 4,402¢ for "flowers" The advertising totals for the GoTo terms swamp all these figures, with 13,262 for "web hosting," 12,306 for "gifts," 8,058 for "marketing," 7,544 for "website design," and 6,851 for "shopping." The potential to influence a purchasing decision on the Web elicits large total bids for information helpful to consumer and producer decisions. This pattern is also evident in the average bids. The average bids for the top terms in each category were 55.96¢ for "web hosting" (GoTo), 18.50 for "flowers" (Lycos), 13.67 for "crime" (hard news), and 10.39 for "Vitamin C" (soft news). The sponsors for the "crime" term were primarily lawyers seeking clients and companies selling protection services, which yielded a higher average price than that paid by companies selling Vitamin C. For the second through fifth terms, the soft news average prices are higher than those for hard news (e.g., the heavy metal band Metallica had an average bid of 9.50 versus 8.63 for the hard news topic computer virus). The prospect of influencing purchasing decisions generated relatively high average bids in the GoTo category, yielding 51.28 for "gifts," 48.80 for "debt," 41.17 for "health insurance," and 40.26 for "home finance"

The GoTo site instructions for advertisers estimate that the top listing in a search term receives three times more clicks by viewers than the fifth term. The maximum prices listed in table 7.7 reflect the bid paid to be listed first in the search term list. The disparities in advertiser support across information categories are again apparent using this measure. "Web hosting" was the top term in the GoTo category of producer and consumer information, with a maximum bid of 590¢. This compares to top maximums of 215 for "flowers," 64 for "social security," and 79 for "Vitamin C." At the fifth spot the dominance of personally useful information in attracting advertiser support is still clear. The GoTo term "debt" attracted a high bid of 415¢, versus 87 for the "Backstreet Boys" (Lycos), 43 for "Ricky Martin" (soft news), and 21 for the "Supreme Court" (hard news).

Advertisers value attention that can be translated into sales. Table 7.8 shows that the

attention of those seeking out information related to consumer and producer choices generates the highest bidding activity from advertisers. Advertisers' total bids, average bids, and maximum bids all increase with total searches for a term. Advertisers are willing to pay more for clicks for popular terms. The transaction costs of assembling audiences may play a role here. It may be cheaper for a company to pay a higher price to generate traffic with a few terms than to try and assemble traffic through small click throughs from less popular (and cheaper) terms. Total bids and maximum bids, controlling for other factors, increase as the number of Web pages mentioning a term grows. This may indicate that with many sites talking about a topic, companies are more willing to spend to generate traffic for their particular sites. The indicator variables in the equation measure whether there is a special impact on prices for a category of information relative to the hard news terms. In the total bids equation, only the GoTo indicator variable is statistically significant. The coefficient indicates that relative to a hard news topic, a topic relating to consumer/producer decisions would generate a bid total that was 1,012¢ higher. In the average bid equation, Lycos terms generate higher bids (by 2.27¢) and GoTo terms receive higher bids (by 11.42¢). Even after one controls in 3 the maximum bid equation for audience demand (measured by total searches) and supply side factors (represented by the number of Web pages using a term), the GoTo terms generate a higher maximum bid (by 78.29¢) relative to hard news topics. In each equation the soft news term was not statistically significant, indicating that once one controls for the magnitude of traffic or interest related to a topic that whether it is a hard or soft news term does not affect the willingness of advertisers to bid for a link. Overall, these results show the strong advertiser support for information related to consumer and purchasing decisions and the relatively weaker advertising support provided for hard news topics.

TABLE 7.8

Advertisers' Willingness to Pay to Support Search Terms (cents), May 2000

	(1) *Total of Advertiser Bids*	*(2)* *Average Bid*	*(3)* *Maximum Bid*
Total Searches for Term	7.29e-3***	1.62e-5***	2.20e-4***
	(8.82e-4)	(4.58e-6)	(4.95e-5)
Number of Web Pages with Term	3.26e-4**	1.24e-6	2.18e-5***
	(1.48e-4)	(7.71e-7)	(8.34e-6)
Soft News Term	28.63	0.73	4.27
	(178.83)	(0.93)	(10.04)
Lycos Term	−52.39	2.27**	10.65
	(199.31)	(1.04)	(11.19)
GoTo Term	1012.11***	11.42***	78.29***
	(220.70)	(1.15)	(12.40)
Adjusted R^2	0.44	0.49	0.40
Number of Terms	318	318	318
Mean of Bid Term	672.63	7.47	47.73

Note: Dependent variable in the OLS regression in 1 is (in cents) the total amount of advertiser bids, in 2 the average bid, and in 3 the maximum bid. Each specification also included an intercept. Standard errors are in parentheses. *** = statistically significant at the .01 level, ** = statistically significant at the .05 level, * = statistically significant at the .10 level.

Conclusions

What will news markets look like when information is costly to create but nearly costless to disseminate? The Internet offers many possible answers to that question. Current guesses about the development of news on the World Wide Web often emphasize the change on the supply side brought about by technology and ignore the demand side composed of people's interests in different types of news. The results in this chapter confirm that the low costs of information distribution on the Net make it possible for more voices to be heard through this medium. At the zip code level, expression is more dispersed on the Internet than the relative concentration of authors or political contributions. Topics in hard news are discussed on a greater number of Web pages than topics in soft news or consumer/producer information. In terms of information demanded by individuals, however, the dominance of entertaining and personally useful information predicted by Downs does hold true on the Internet. People are much more likely to search the Net for information about entertainment figures than political issues. Decisions about purchasing a consumer item or making a business purchase attract more attention (and advertiser support) than voting decisions.

In a world of many outlets and scarce attention, consumers on the Internet are likely to go with familiar media brands. The analysis here shows that online use of newspapers is much more concentrated, as measured by links, than newspaper circulation. The pressures to stand out in a world of many viewing options extends beyond company brand names to the use of individual brand names to attract attention. In print and broadcast markets, the personal fame of anchors, reporters, pundits, and columnists may draw consumers to a particular media outlet. The following chapter explores the extent to which journalists are becoming part of the products marketed as news.