WEB SITE USE
AND NEWS TOPIC AND TYPE

By H. Denis Wu and Arati Bechtel

This study investigates the relationship between types of news events and daily traffic at the New York Times on the Web. CNN and ABC newscasts were content analyzed to represent each day's news coverage and compared with Web site usage data made available by the Times. The results indicate that level of disruptiveness and episodicity were positively correlated with online traffic. Also, several news topics—international politics, education, and science and technology—were positively correlated with online news usage. During the period examined, dominated by the Clinton-Lewinsky scandal, domestic politics, weather, and accident and disaster news were negatively correlated with Web site usage.

The World Wide Web is a great source of news because it is available twenty-four hours a day and Web users do not have to wait for a newspaper or newscast to find the story that interests them. Online news users can choose directly and immediately from a great variety of content and not rely on the necessarily more narrow selections of a newspaper editor or television news director. Two major news events in the United States, John F. Kennedy Jr.'s plane crash in 1999 and the presidential election of 2000, illustrated the Web's clout in the news marketplace. The week beginning 18 July 1999 saw a record high number of pages visited of the New York Times on the Web as millions of Web users checked on developments in the disappearance, recovery, and funeral of Kennedy and his plane's passengers.1 In November 2000, news Web sites also reported substantially higher traffic as people sought news on the resolution of the controversial presidential race,2 and sites saw even greater jumps in use immediately following the terrorist attacks of 11 September 2001.3

The response of Web users to these stories suggests that audiences may actively seek out online news sites for ongoing, breaking news—stories that command people's attention. But if online newspapers are to serve their audiences effectively, what topics of news stories should they select and what qualities should the stories have? Will users be more likely to seek out updates of breaking events or stories that provide context and background on issues? Using multiple data sets, this study seeks to investigate the types of stories and topics that attract audiences.

H. Denis Wu is an assistant professor in the Manship School of Mass Communication at Louisiana State University. Arati Bechtel is a former assistant professor in the Manship School of Mass Communication and now writes about technology for the News & Observer of Raleigh, North Carolina.

J&M Quarterly
Vol. 79, No. 1
Spring 2002
73-86
©2002 AEJMC
This study relies on the uses and gratifications approach, which maintains that "individuals differentially select and use mass media to gratify or satisfy felt needs." The uses and gratifications approach to studying media assumes that audience members are active and make motivated choices based on previous exposure to the media. It also assumes that media use is only one way among others of satisfying needs that audience members experience in daily life. These characteristics make uses and gratifications an appropriate theoretical basis for an investigation of the content that leads users to news Web sites, a conclusion supported by Morris and Ogan and by Heeter.

Researchers have already begun to address numerous questions relating to online journalism. They have described visual aspects of online journalistic content, compared the acquisition of knowledge in online and traditional newspapers, observed the increase in the number and popularity of online newspapers, and testified to the skyrocketing rise of traffic at a news Web site. In addition, media scholars have analyzed factors, such as credibility, by which news audiences judge online news; used eye-tracking technology to study how users interact with online news; examined the impact of multimedia elements on the processing and perception of online news; described dimensions of interactivity in online newspapers; and compared the user demographics of the Internet to those of traditional media. Except for the latter study by Stempel and his colleagues, little research has attempted to connect the news landscape provided by television to audiences' use of a news Web site, which is the aim of this research. This article explores four factors that might be related to audiences' use of the New York Times on the Web: topics of news; disruptiveness of news; episodic or thematic quality of news; and expectedness of news.

News Topics. Studies on preferences in news topics have varied findings. A study of the readers of a regional daily newspaper found that local news—followed by national, international, and university news—was the favorite topic in 1985 as well as in 1994. Studies of online news have yielded contradictory answers to the question of whether readers prefer local news. More respondents to a poll conducted by the Pew Center in 1998 went to the Web for news on science, health, technology, finance, and weather than for international, political, sports, or local news. But an online poll conducted by Editor & Publisher found local news the most popular content of online newspapers. It should be noted, however, that people's self-reported preferences may not accurately represent their news consumption. The literature suggests the following tentative hypothesis:

H1: National, international, science, and technology news will result in greater traffic at a news Web site than other topics.
This study is an exploration of a new news medium. We cannot assume that the same kinds of topics that attract traditional-media users also would draw online users. Still, it is reasonable to guess that consumers of online news resemble older-media audiences in their preferences and that the online surveys offer clues about these tastes. This study, unlike those surveys, goes beyond the broad labels of news sections, such as national, international, and local, to find out more specifically the types of topics that appeal to online news consumers.

**Disruptiveness of News.** Although the presence of certain traditional news values is believed to be a significant factor in news consumers' interest in stories, few empirical studies have demonstrated this specifically. Straughan found that readers rated news stories that included conflict and proximity news values as significantly more interesting than stories that lacked those values. An annual survey by the Pew Center asks Americans to identify the stories that attracted their interest the most. In 1999, the survey found that Americans were most interested in the shootings at Columbine High School. In fact, the top ten stories for 1999 in terms of audience interest were mostly about violence in the United States, conflicts involving U.S. military forces, and weather-related disasters. In many annual surveys of the past, the Pew Research Center has found that stories with such a disruptive quality commanded the interest of news audiences. Additionally, anecdotal evidence indicates that viewers swarmed to news Web sites for updated information about the development of the September 11th attacks in 2001. For example, MSNBC reported Web traffic ten times the normal volume, reaching over 20 million a day.

H2: The disruptiveness of the day's news is positively related to traffic at a news Web site.

Disruptiveness in this study is defined as the timely, urgent, breaking characteristic of news. Highly disruptive news would interrupt an audience member's everyday activities to command attention, whereas a nondisruptive story would go unnoticed by a Web user. Disruptiveness is at the heart of newsworthiness. Therefore, we hypothesize that a more disruptive day of news would result in more traffic to a news Web site.

**Episodic vs. Thematic.** Iyengar distinguished between news that is episodic or thematic. He found that television routinely used "episodic framing" to present news as specific events or particular cases. On the other hand, "thematic framing" of news places issues and events in some context. "Episodic framing depicts concrete events that illustrate issues, while thematic framing presents collective or general evidence," Iyengar explained. Iyengar's studies and other studies have found that episodic news coverage is much more common than thematic. For example, a study of local television news coverage of youth violence found that episodic coverage of violence was more than five times more common than thematic coverage.

A survey of news audiences' preferences suggests that people prefer episodic news. The 1999 year-end survey by the Pew Center
showed that stories that lent themselves to episodic news coverage, such as the destruction wrought by Hurricane Floyd or the military clashes between the United States and Iraq, captured the attention of news audiences.29

H3: Days dominated by episodic news result in more traffic to a news Web site.

Because there are more episodic news stories than thematic ones, which may reflect audience preference, episodic news coverage will result in greater Web traffic. This study’s design provides a suitable natural setting to test the relationship between these types of news events and traffic.

Expectedness of News. Not surprisingly, a desire to know what is going on in the world, which is among the major functions served by mass media,30 is a strong explanation for news media use. In a study of college students’ media use, Vincent and Basil found that increasing surveillance needs resulted in greater use of all kinds of news media.31 In addition, Wenner reported that interest in current events drives news media use across various media.32 It is, therefore, reasonable to predict that audiences will actively seek information about upcoming and prescheduled events to satisfy their needs.

H4: Expected news is positively related to traffic at a news Web site.

The researchers suspect that if audiences know in advance that a particular event or proceeding will take place on a certain date, such as a political ceremony or a sports contest, it is likely that users would tap Web sites for coverage of that specific, expected event.

The literature does not provide explanatory or conceptual bases for information consumption following the advent of the Internet. In addition, a great deal of public and private investment in Internet-based news services is attributable to anecdotal cases and, perhaps, has occurred purely because of cyberhype. This study is designed to add to the literature by investigating online news usage and providing conceptually constructive answers as to why news consumers may visit a news Web site.

Method

This study first employed content analysis to capture the attributes and elements of news that were covered during the time frame. As it is impossible to trace the news items that were published on the Web, and given the fact that the paper version of the New York Times is different from its online counterpart and could not offer as timely coverage as broadcast news (also a feature of Web news), the researchers used broadcast news—from ABC and CNN—to represent “the news of the day.” The findings of the content analysis of ABC and CNN newscasts were then analyzed along with the Web site’s traffic data. With these two sets of data, the researchers were able to detect whether a correlation
exists between the types of news events covered on a particular day and people's use of the *New York Times on the Web*.

The researchers coded twelve months of evening news abstracts of CNN and ABC from 1 July 1998 through 30 June 1999. The news abstracts were downloaded and printed out from the Web site of Vanderbilt University's Television News Archive. Trained students coded the top five stories of ABC and CNN nightly newscasts for each day. Each story was coded for primary and secondary topics of the story, whether the story was episodic or thematic, the level of disruptiveness of the story, and whether the story was about a planned news event. The news topic includes nineteen categories ranging from domestic politics, business news, international relations, sports, weather, to human interest.

Intercoder reliability overall was satisfactory, with minimal judgment variables (date, broadcaster name, story identification, top story, length of broadcast, and story venue) receiving perfect intercoder agreement as measured by Holsti's formula. The judgment variables were topic (85%), episodicity (87%), disruptiveness (77%), and expectedness (82%).

The researchers then created an index for each of the following attributes for each day's news: disruptiveness, episodicity (as opposed to thematic), and expectedness. The disruptiveness index for each news day, for example, results from averaging the sum of each story's disruptiveness, derived by multiplying its level of disruptiveness by broadcast time. Therefore, the formula of the index takes both the direction of the attribute (extremely disruptive, disruptive, or nondisruptive) and time (in terms of seconds) into account. Both indexes of episodicity and expectedness used similar calculation except that dummy coding was used to quantify the binary nature of coding (e.g., episodic story was coded as 1 while thematic 0). The 19 news topics were also dummy-coded (1 was entered for a given topic, 0 for none) and aggregated for each news day to test whether any of the news topics was associated with Web site traffic.

The researchers correlated the various topics and the three indices with daily site traffic data provided by the research department of the *New York Times on the Web*. The data, which cover the period of 1 July 1998 through 30 June 1999, include measures of the amount of time users spent at the site and the number of visits, unique viewers, and pageviews. The distribution of the measures of Web traffic (see Appendix 1 for details) indicates the Web site's viewership varied dramatically from day to day. The measures of time, visits, and viewers seem appropriate to represent Web use. Based on our initial examination, the three variables are highly correlated. For the sake of parsimony, a “Web-use” factor was extracted by a Varimax method to represent them. The factor accounted for 98.6% of variance from the three variables. In addition, because the researchers were interested in the potential differences in relationship due to the time order, the Web-use factor was also transformed into both “lead” and “lag” variables. In other words, the correlation between Web use and news types/topics was examined in three distinct situations: (1) Web use led broadcast news by one day; (2) Web use and broadcast news occurred on the same day; (3) Web use occurred after the broadcast news.
Background of the Web site. In 1996 the *New York Times* launched its Web site. During the period covered in this study, the Web site's operations were as follows: Each night the Web staff put every story that appeared in the following day's newspaper online. During the day, the staff updated the site's content with new material from the newspaper's reporters, wire services, and other partners with whom the site has contracted. Most stories put up on the site were not accessible, without payment of a fee, past the date of publication. The content on the site's homepage and on section fronts (such as Business or International) changed throughout the day as stories broke and progressed. The top seven or eight stories of the moment were placed at the top of the homepage.

During the period of this study, *Times* user data were generated using Open AdStream tracking software by Real Media. Open AdStream allowed collection of data about users by first requiring users to register with the Web site, in exchange for free access to the Web site's content. To register, users had to report personal information, including their name, ZIP code, age, sex, education, and household income, and had to select a logon I.D. and password. In order to access stories on the Web site, users had to log on to the site and set their Web browsers to accept "cookies," small text files that a Web site can place on a hard drive to track individual usage and viewing throughout the many pages of the site.

The combination of the registration of each user, logon requirement, and use of cookies permitted the site to collect data on precisely which stories, sections, images, and advertisements each individual user downloaded to his or her computer. In other words, Open AdStream enabled the site to build a database consisting not only of demographic information about its users, but also traffic data (number of pageviews, number of unique visitors, time spent on the site, etc.). The researchers received partial data, as the *Times* research department declined to allow access to data about the demographics of site users, click-through rates for banner ads, and traffic data for specific pages and sections of the site. As of September 2001, the site claimed to have a total of 350 million pageviews by 9 million unique visitors, who had an average household income of almost $85,000. About 58% of these were male, and about 85% lived in the United States.

The major reason the researchers were interested in this news site is that the *New York Times* has long been considered the top elite newspaper in the country. Its online site, carrying the same trustworthy brand name, might well exert as much influence as does the paper version. The second reason why the site was chosen, however, was accessibility and practicality. The authors were fortunate to be offered the user data on the Web site, which is closely guarded and prohibited from public release.

All of the data—content analysis results and the Web site's daily traffic data—were managed with a spreadsheet program. A total of 3,288 news stories from ABC and CNN were coded (on some days, the transcripts were not available from the archive). Calculation and compilation of the indexes for each day during the time frame and correlation analysis were executed using SPSS. For the study period (1 July 1998 to
The content analysis part of the project resulted in 3,288 news stories coded—roughly half of the stories came from ABC and the other half from CNN. During the study period, 28% of the stories dealt with international politics, 13% were about domestic politics, 14% covered legal proceedings (the Lewinsky scandal occupied much of the air time), and 10% were about the domestic economy. The following news topics are also relatively prominent: crime-violence (7%), accident/disaster (6%), health/medical information, public issue/policy, and weather (3% for each).

Fifty-two percent of the stories included in the sample took place in the United States, 34% in foreign countries, and roughly 14% in both the United States and at least one foreign country. The predominant majority of the news was episodic (91%); only about one in ten stories was thematic. Only 4% of the stories were coded “extremely disruptive,” 69% “moderately disruptive,” and 27% “nondisruptive.” Only 22% of the news was about preplanned events.

Table 1 shows the relationships among the four factors examined in the newscast content—topics of news, disruptiveness of news, episodic or thematic quality of news, and expectedness of news—and Web site traffic. To examine any lag between the time the evening news is broadcast and the time users check into the New York Times on the Web, we considered and individually tested three different models. The first model was that, after getting news from the Web site, the user turns to television, possibly for live images and/or better soundtrack. Stempel and his colleagues, however, speculated that this scenario is unlikely to take place because it does not fit an information-seeking model. Second, an audience member might view television nightly news on the same day she checks the New York Times on the Web, although it is technically impossible to know which behavior precedes the other, based on these data. Third, an audience member might view a newscast prior to her exposure to the Internet news source—the audience member may be interested in getting more information about certain news after exposure to a television newscast.

Among the news topics coded, it is apparent that newscast coverage of international politics is positively related to use of the New York Times on the Web. This correlation is statistically significant at the .01 level for all three scenarios. Stories about science and technologies, as well as education topics, are also associated with substantial Web visits. Other significant, although negative, correlations were found for Web site usage with domestic politics, weather, and accident/natural disaster news stories. Explanations could include audience fatigue with the Clinton-Lewinsky scandal, which dominated domestic politics during the time period studied, and the possibility that audiences do not turn to the New York Times on the Web for weather or disaster coverage, preferring other sites, such as Weather.com or CNN.com. However, it is worth pointing out that international coverage is negatively related to all of
TABLE 1
Correlation between Web Use and News Topic and Type

<table>
<thead>
<tr>
<th>News Topics</th>
<th>Web-1day</th>
<th>Same day</th>
<th>Web+1day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic Politics</td>
<td>-.179</td>
<td>-.148</td>
<td>-.144</td>
</tr>
<tr>
<td>International Politics</td>
<td>.316</td>
<td>.292</td>
<td>.335</td>
</tr>
<tr>
<td>Legal Proceedings</td>
<td>-.118</td>
<td>-.097</td>
<td>-.113</td>
</tr>
<tr>
<td>Domestic Economic/Business</td>
<td>-.039</td>
<td>-.022</td>
<td>-.124</td>
</tr>
<tr>
<td>International Trade</td>
<td>-.011</td>
<td>-.008</td>
<td>-.065</td>
</tr>
<tr>
<td>Celebrity/Personality</td>
<td>.001</td>
<td>-.010</td>
<td>.038</td>
</tr>
<tr>
<td>Culture/Arts/Film/Performance</td>
<td>.111</td>
<td>.000</td>
<td>.001</td>
</tr>
<tr>
<td>Sports</td>
<td>.023</td>
<td>-.023</td>
<td>-.052</td>
</tr>
<tr>
<td>Environment/Energy/Conservation</td>
<td>-.007</td>
<td>-.008</td>
<td>.005</td>
</tr>
<tr>
<td>Weather</td>
<td>-.122</td>
<td>-.124</td>
<td>-.154</td>
</tr>
<tr>
<td>Health/Medical News</td>
<td>.063</td>
<td>-.037</td>
<td>-.084</td>
</tr>
<tr>
<td>Accident/Natural Disaster</td>
<td>-.126</td>
<td>-.196</td>
<td>-.152</td>
</tr>
<tr>
<td>Social Issues and Policies</td>
<td>.038</td>
<td>.024</td>
<td>.018</td>
</tr>
<tr>
<td>Education</td>
<td>.101</td>
<td>.136</td>
<td>.145</td>
</tr>
<tr>
<td>Religion</td>
<td>.006</td>
<td>.033</td>
<td>.063</td>
</tr>
<tr>
<td>Science and Technology</td>
<td>.094</td>
<td>.138</td>
<td>.119</td>
</tr>
<tr>
<td>Feature/Human Interest</td>
<td>.007</td>
<td>-.015</td>
<td>.075</td>
</tr>
<tr>
<td>Crime/Human Violence</td>
<td>-.081</td>
<td>-.015</td>
<td>.010</td>
</tr>
</tbody>
</table>

Expectedness                     | -.077    | -.070    | .082     |
Episodicity                       | .253     | .283     | .323     |
Disruptiveness                    | .249     | .244     | .268     |

**Bold:** significant at 0.01 level (2-tailed).
**Underline:** significant at 0.05 level (2-tailed).
N=362

these topics, which suggests that there may be an invisible balancing mechanism in each day's news menu—when international politics news increases, coverage of other subjects must be reduced due to the limitation of space. In addition, because the New York Times is known for its international coverage, its Web site might benefit from this reputation and draw users. These findings provide only partial support for H1, because international, science, and technology coverage was positively correlated with Web site traffic, but domestic politics was not.

The results also indicate that disruptiveness of a day's news coverage correlates positively with online news use in all three of the scenarios of audience behavior. In other words, when the news landscape of the day is disruptive, there is greater Web traffic. Therefore, H2 is supported. And as Table 1 shows, episodicity of news on a particular day is significantly correlated with Web site traffic. This finding supports H3. Another index, "expectedness," which described the level of planned-event coverage in each day, appears to be unrelated.
to traffic at the Web site in all three of the scenarios. H4, therefore, is not supported.

The different correlations across the three different time models seem to suggest that New York Times on the Web users may have learned the news covered by television or other media first and then tapped into the Web site for additional information. This interpretation is consistent with anecdotal evidence in the wake of the September 11th attacks when people checked the Web for updated information about the incident.38

The findings lead us to conclude that news Web sites probably benefit from other media coverage of unexpected events and incidents rather than coverage of planned events. We also speculate that coverage of certain news topics leads to news Web site visits. Audiences with specific interests will visit Web sites for further coverage after learning from other sources that certain topics are in the news. Therefore, news Web sites probably should target special segments of audiences to meet their needs and news tastes rather than seeking to appeal to a broad, mass audience.

The findings indicate that traffic and usage of the New York Times on the Web had a positive relationship with the other media coverage of international politics, international relations, and issues that involve foreign countries. In addition, some of the dominant, traditional news categories, such as natural disaster, domestic politics, and legal coverage, were negatively related to online usage. This finding suggests that each news Web site may be perceived differently by online news users, and therefore be accessed according to a user’s perception of the strength or features of a particular site. The New York Times on the Web may be perceived as authoritative and comprehensive on international news, but not an especially strong source for information on weather or crime—news categories that may be better covered by other sites, such as Weather.com or local news media sites.

The findings also suggest that the users of the New York Times on the Web, unlike many news consumers and contrary to what many news professionals assume about audiences, are keenly interested in issues pertaining to the international community. Traffic at the site increases when international politics stories are in the news landscape. International news is often considered a less important news category by many news editors throughout the United States, but the finding here shows that at least some Web users do not share this belief. International news could very well be the principal attraction of the New York Times on the Web.

Another important finding is that these users are interested in disruptive and episode-oriented news stories, even though they may demand high news quality and probably appreciate background and contextual information in addition to the basic facts. Scholars’ recommendation that news media should focus more on thematic stories and limit episodic coverage will remain a balancing act—between breaking, episodic event coverage, and coverage that provides a “bigger picture.” We found many episodic stories from CNN that do provide substantial

Discussion
background information. Yet whether this hybrid style of news coverage is practiced by news Web sites and appreciated by Web users is unclear. More studies on this issue are needed.

One limitation of this study is that the users of the New York Times on the Web are probably not representative of all users on news Web sites. Although definitive statistics about newspaper Web site users are not available, Editor & Publisher's July 1999 survey offers some information: the average income of a newspaper Web site user was $59,317, which is about $21,000 less than the average income of a New York Times on the Web user; and about 42% of news Web site users are women, as compared with 41% at the New York Times site.39 The researchers suspect that the users of the New York Times Web site are probably slightly more affluent, educated, and socio-politically active than average online users, which might affect the generalizability of this study's results.

Another limitation is that the content we analyzed was from network and cable television newscasts, which do not offer necessarily the same content as the New York Times Web site. In addition, this sampled cable and network television news coverage does not necessarily represent each day's "news universe." However, choosing ABC over other networks may be relatively unproblematic since network nightly news in general is fairly similar.40

Understanding people's use of online news services is important for mass communication theory, for informing media industries, for analyzing society at large, and for examining the democratic system of the country. Unfortunately, a great deal of the data generated by news Web sites in the last decade is treated as confidential information because of corporate interests. Our study would have been substantially enriched by other, inaccessible data from the Web site, such as the users' demographics and user interest in specific sections of the site. With these data, it would be possible for us to know more about the predictors of Web use and also to examine empirically or update uses and gratifications theories in today's "cyber circumstances." In addition, the research findings generated from Web traffic and demographic data would be beneficial to the new technology industries, which could base their Web site design on the knowledge to better serve their audiences. This article, however, only offers a starting point for more explanatory research on the use of news Web sites.

Appendix 1 and Notes follow.
### APPENDIX 1

*Web Traffic Statistics from the New York Times on the Web*  
(from 1 July 1998 to 30 June 1999, N=362)

<table>
<thead>
<tr>
<th>Hours</th>
<th>Visits</th>
<th>Viewers</th>
<th>Pageview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>85,703.83</td>
<td>336,171.15</td>
<td>222,299.38</td>
</tr>
<tr>
<td>Median</td>
<td>88,694.26</td>
<td>353,821.00</td>
<td>232,801.50</td>
</tr>
<tr>
<td>S.D.</td>
<td>34,750.06</td>
<td>117,204.15</td>
<td>70,597.23</td>
</tr>
<tr>
<td>Minimum</td>
<td>3,847</td>
<td>73,658</td>
<td>61,567</td>
</tr>
<tr>
<td>Maximum</td>
<td>166,912</td>
<td>597,233</td>
<td>384,150</td>
</tr>
</tbody>
</table>
NOTES


33. The coding instruction and sheet are available upon request from
the authors.

34. Roger D. Wimmer and Joseph R. Dominick, *Mass Media Research* (Belmont, CA: Wadsworth, 2000): 151-52. Reliability=2M/(N1+N2), where M is the number of coding decisions on which two coders agree, while N1 and N2 are the total number of coding decisions by the first and second coder, respectively.

35. The mean and standard deviation values for the three indexes are as follows: expectedness (.6553, .6517); disruptiveness (5.39, 2.26); episodicity (2.46, .85).

36. “Unique viewer” refers to a single registered user of the Times site. Each user must select a logon name and password at the time of registration and use it to gain access to the site. In this way, the Times is able to get a reasonably reliable grasp of the number of users, how long each user stays at the site, how many pages a user looks at, etc. “Pageview” refers to the loading of a particular HTML page onto a Web site viewer’s browser.


