The Effectiveness of Internet Content Filters

Philip B. Stark*
Department of Statistics
University of California, Berkeley
November 10, 2007

Abstract. *As part of its defense of the Child Online Protection Act, which seeks to prevent minors from viewing commercially published harmful-to-minors material on the World Wide Web, the U.S. Department of Justice commissioned a study of the prevalence of "adult" materials and the effectiveness of Internet content filters in blocking them. As of 2005–2006, about 1.1% of webpages indexed by Google and MSN were adult—hundreds of millions of pages. About 6% of a set of 1.3 billion searches executed on AOL, MSN and Yahoo! in summer 2005 retrieved at least one adult webpage among the first ten results, and about 1.7% of those results are adult webpages. These estimates are based on both simple random samples of webpages indexed by search engines and on a stratified random sample of searches. Webpages with sexually explicit content intended for adult entertainment (i.e., not in an educational, medical or artistic context) were used to test a variety of Internet content filters for underblocking—failing to block webpages that they are intended to block. A random sample of "clean" webpages with no sexual content or reference to sex was used to test the filters for overblocking—blocking webpages they are not intended to block. Webpages retrieved by the most popular searches according to Wordtracker were also categorized and used to test filters. Generally, filters with lower rates of underblocking had higher rates of overblocking. If the filter most effective at blocking adult materials were applied to search indexes, typical query results, or the results of popular queries, the number of clean pages blocked in error would exceed the number of adult pages blocked correctly.*

* This work was supported by the United States Department of Justice. I testified on behalf of the United States at trial in *Am. Civil Liberties Union v. Gonzales*, 478 F. Supp. 2d 775 (E.D. Pa. 2007), and submitted expert declarations in the related matter of *Gonzales v. Google*, 234 F.R.D. 674, 688 (N.D. Ca. 2006). Much of the data collection was performed by a group led by Paul Mewett at CRA International. I am grateful to David Freedman, Raphael Gomez, Theodore Hirt, Joel McElvain and an anonymous referee for helpful conversations or comments on an earlier draft.

Introduction

Although the Child Online Protection Act (COPA) was enacted by Congress in 1998,[1] it has yet to be enforced. COPA provides:

> Whoever knowingly and with knowledge of the character of the material, in interstate or foreign commerce by means of the World Wide Web, makes any communication for commercial purposes that is available to any minor and that includes any material that is harmful to minors shall be fined not more than $50,000, imprisoned not more than 6 months, or both.[2]

COPA does not restrict material with literary, artistic or educational value for minors;[3] the statute does contain exemptions for Internet Service Providers (ISPs) and search engine companies.[4] The law allows an affirmative defense: if the publisher uses an "age screen" to keep minors away from the material, he is not liable.[5] Requiring users to provide a valid credit card number also suffices to satisfy the "age screen" requirement.[6]

COPA is the second law Congress has passed to protect children on the Internet.[7] The first was the Communications Decency Act of 1996, which the Supreme Court found unconstitutional, in part, because it was overly broad.[8] COPA has a narrower scope. It focuses on the World Wide Web rather than on all interactive computer communication and only restricts commercial publishers.[9]

---

1 *Child Online Protection Act*, Public Law 105–277, *U.S. Statutes at Large* 112 (1998): 2681–736.

2 Ibid., § 1403, 2681–736.

3 Ibid., 2681–739.

4 Ibid., 2681–737.

5 Ibid.

6 Ibid.

7 Ashcroft v. Am. Civil Liberties Union (*Ashcroft II)*, 542 U.S. 656, 661 (2004).

8 *Communications Decency Act of 1996*, Pub. L. No. 104-104, §§ 501–561, 110 Stat. 133-43 (codified in scattered sections of 47 U.S.C. and 18 U.S.C. (2000)); Communications Decency Act of 1996, Public Law 104-104, *U.S. Statutes at Large* 110 (1996): 133-43, codified at *U.S. Code* 47 and 18 (2000), scattered sections; *Ashcroft II*, 542 U.S. at 661 (citing Reno v. Am. Civil Liberties Union, 521 U.S. 844 (1997)).

9 *Ashcroft II*, 542 U.S. at 661, 779.

The American Civil Liberties Union (ACLU) and other plaintiffs filed suit in district court challenging COPA on constitutional grounds on October 22, 1998 the day after it was signed.[10] Among other things, plaintiffs argued that Internet content filtering is a less restrictive means of achieving the government's (legitimate) goal of protecting minors.[11] The government was enjoined from enforcing COPA as passed.[12] The Department of Justice appealed the injunction. The matter has now been considered by the Supreme Court twice.[13] In the second case, the Supreme Court upheld the injunction preventing the enforcement of COPA pending adjudication of the constitutionality of the Act.[14] Citing the rapid pace of change in technology and the web as well as "a serious gap in the evidence as to the effectiveness of filtering software," the Supreme Court remanded the case to the District court in 2004 so that the parties could "update and supplement the factual record to reflect current technological realities."[15]

Studies have recommended the use of Internet content filters as one of many tools to help protect children.[16] Previous empirical studies of the effectiveness of Internet content filters used to block objectionable material on the World Wide Web relied on small samples of convenience, rather than random samples from some larger universe of webpages.[17] As a consequence, the findings of

10 Ibid., at 663.

11 Ibid.

12 Ibid.

13 Ibid. at 656; Ashcroft v. Am. Civil Liberties Union (*Ashcroft I*), 535 U.S. 564 (2002).

14 *Ashcroft II*, 542 U.S. at 672–73.

15 Ibid., 671–72.

16 Commission on Online Child Protection, "Report to Congress – October 20, 2000", 19–22, http://www.copacommission.org/report/COPAreport.pdf; and Dick Thornburgh and Herbert S. Lin, eds., *Youth, Pornography, and the Internet* (Washington, D.C.: National Academy Press, 2002), 267–303.

17 E.g., eTesting Labs, "Corporate Content Filtering Performance and Effectiveness Testing," http://web.archive.org/web/20030406232751/www.websense.com/whyqualitymatters/etestinglabs-fullreport.pdf, 14–19; eTesting Labs, "U.S. Department of Justice: Updated Web Content Filtering Software Comparison," http://www.surfonthesafeside.com/aboutcic/usdoj.pdf), 6–7; Paul Greenfield, Peter Rickwood, and Huu Cuong Tran, "Effectiveness of Internet Filtering Software Products," CSIRO, 23,

those studies cannot be extrapolated reliably beyond the samples they tested.[18] Moreover, the independent tests of filters that had been conducted in the previous few years were consumer reviews, not scientific studies.[19]

To fill the gap in the evidence regarding filtering software, the Department of Justice commissioned a study of the performance of filters. The study was also designed to determine whether "adult" webpages were hosted in the U.S. or abroad because the presence of commercial ties between foreign-hosted commercial providers of pornography and the U.S. affects methods of enforcement available. I was involved in the design of the study, data collection and analysis. Ultimately, I testified about the study at trial. This paper presents both the results of the study and some of the technical details.

The case was tried on its merits between October 23 and November 20, 2006 in the Eastern District of Pennsylvania by Senior Judge Lowell A. Reed, Jr.[20] Judge Reed issued a permanent injunction enjoining the government from enforcing COPA in a decision announced on March 22, 2007.[21] Among the reasons for his decision, Judge Reed included: "defendant has failed to meet his burden of showing that COPA is the least restrictive and most effective alternative in achieving the compelling interest [of protecting children from sexually explicit material on the Web]."[22] The

---

http://www.acma.gov.au/webwr/aba/newspubs/documents/filtereffectiveness.pdf (accessed March 19, 2008).

18 See, e.g., William G. Cochran, *Sampling Techniques*, 3rd ed. (New York: John Wiley & Sons, 1977), 10; Shari Seidman Diamond, "Reference Guide on Survey Research," in *Reference Manual on Scientific Research*, 2nd ed. (Washington, D.C.: Federal Judicial Center, 2000), 242–44; Leslie Kish, *Survey Sampling* (New York: John Wiley & Sons, 1965), 19.

19 E.g., "Filtering Software: Better, But Still Fallible," *ConsumerReports.org*, June 2005, http://www.consumerreports.org/cro/electronics-computers/resource-center/internet-filtering-software-605/overview/index.htm (accessed March 19, 2008); Jay Munro, "Editor's Choice: Cybersitter 9.0," *PCMag.com*, August 3, 2004, http://www.pcmag.com/article2/0%2C2704%2C1618830%2C00.asp (accessed March 19, 2008).

20 Am. Civil Liberties Union v. Gonzales, 478 F. Supp. 2d 775 (E.D. Pa. 2007).

21 Ibid., 478 F. Supp. 2d at 821.

22 Ibid.

Department of Justice appealed this decision on September 17, 2007.[23]

How effective are Internet content filters at blocking adult materials on the World Wide Web?[24] To answer that question quantitatively requires a precise definition of "effective." Internet content filters make two kinds of errors: 1) blocking a page that should not be blocked, called *overblocking*, and 2) failing to block a page that should be blocked, called *underblocking*I.[25] Measuring underblocking and overblocking separately is important to understanding how well filters work. Single-number summaries of accuracy ("the filter is 99.9% accurate") are virtually meaningless. A filter that does nothing at all has 99.9% accuracy on a test set that contains 1,000 webpages, 999 of which should not be blocked. But it misses 100% of the pages that should be blocked.

Errors are more likely for some pages than for others: overblocking and underblocking rates depend on the webpages used for testing. What is a reasonable set of webpages for testing filters? For the purpose of protecting children, it would help to know what webpages children attempt to view. In principle, one could monitor the web browsing behavior of a large random sample of children in various age groups to see what webpages they come across, and then test filters on those pages. To the best of my knowledge, such data do not exist and would be extremely expensive—if not impossible—to collect.[26] To test the effectiveness of Internet content filters, the study used several complementary

23 Brief for the Appellant, Am. Civil Liberties Union v. Gonzales, No. 07-2539 (3rd Cir. Sept., 2007).

24 Internet content filters may also restrict access to other content and services on the Internet—the focus here is on browsing the World Wide Web, the scope of COPA. One should be skeptical of manufacturers' claims about the performance of Internet content filters: "Filter vendors sometimes provide estimates of overblock and underblock rates, but without knowing the methodology underlying these estimates, the cautious user must be concerned that the methodology is selected to minimize these rates." Thornburgh and Lin, *Youth, Pornography, and the Internet*, 277, *supra* note 16.

25 Plaintiffs and their experts argued that only underblocking matters. *Gonzales* at 794. Whatever merit this might have as a legal argument, it makes little sense scientifically. If all that matters is underblocking, one can simply turn off the computer. No objectionable material will be displayed: the underblocking rate is zero. But, the overblocking is draconian.

26 The government obtained data from a company that studies individual web browsing behavior, but the quality of the data was poor and the sample size was small. The test sets used in the present study attempt to "bracket" the collection of webpages minors are likely to encounter. The sets contain the webpages minors are most likely to encounter frequently, although possibly in different proportions. There would need to be something quite peculiar about the mix of webpages minors encounter for filters to perform well on that mix but poorly on the test sets. Regardless, the test sets in this study provide complementary slices through the indexed portion of the World Wide Web, reasonable sets on which to test filters.

sets of webpages derived from search providers and a company that reports search behavior. Individuals' experience of the World Wide Web is largely mediated by search engines: according to Wordtracker reports, "google" and "yahoo" are among the most common search queries.[27] This indicates that people even use search engines to find search engines by name, instead of typing the address of the search engine into the web browser.

When a search engine receives a query, it retrieves results from a catalog of webpages called an *index*. Each search engine has its own index, and the major search providers add and remove webpages from their indexes regularly to keep them up to date. Providers also record the queries that they receive.[28] Search providers dedicate staggering amounts of resources to creating and maintaining their indexes; this includes figuring out which webpages users want to see. Only a fraction of the World Wide Web is indexed by search engines, but it is the most accessible part, comprising tens of billions of pages, and it is the entry point for most people. The set of webpages indexed by search engines is both the most accessed and accessible and thus comprises a reasonable set of webpages to test filters.

However, just because a webpage is in a search index does not guarantee that the webpage will be seen by anyone. The webpages returned when actual queries are entered into a search engine are a better measure of what people see and how often they see it. Of those, the webpages that appear on the first page of results are those most frequently viewed. For the Google, MSN and Yahoo! search engines, the first page shows approximately the first ten results of each search. The first ten results of actual queries typed into search engines are a second reasonable set of webpages to test filters.

---

27 Wordtracker markets lists of the most popular search terms, collected from Dogpile.com and MetaCrawler, "meta search" services that send users' queries to a variety of search engines—including MSN, Yahoo!, Google and Ask.com. Lists of the Wordtracker Top 500 Search Engine Keywords of the Week can be found at http://www.searchengineguide.com/wt/.

28 The queries can include some information that could identify the user, such as the user's IP address and "cookies." Different search providers retain different pieces of information for different periods of time. For this study, the Department of Justice requested that the search providers not supply any information that could identify the individual who ran the search.

Many queries are rare. The webpages they return might not be viewed by many users. In contrast, the first ten webpages returned by extremely popular queries comprise a third test set: webpages many users see. One would expect filters to perform best for this set, because overblocking or underblocking those webpages would quickly generate a large number of customer complaints.

These three sets of webpages—those indexed by search engines, the first ten results of searches, and the first ten results of the most popular searches—are different slices of the World Wide Web. They provide complementary universes on which to test Internet content filters. If filters performed well on any of them, one might wonder whether filters perform well on the particular mix of webpages children tend to encounter. We shall see that filters do not perform very well on any of these sets, and, hence, not on mixtures of the three. As a result, it is highly implausible that filters protect minors effectively.

Data

The study used data collected in 2005 and 2006. Through subpoena, the government obtained a random sample of 50,000 webpages from the Google search index and a random sample of 1 million webpages from the MSN search index.[29] I drew random sub-samples of 50,000 webpages from the MSN index and 11,000 webpages from the Google index.

The government also obtained a week of search queries from AOL, MSN and Yahoo! by subpoena.[30] There were approximately 1.3 billion searches in all. Google, Yahoo!, MSN and AOL

29 The MSN sample was drawn in November 2005 and the Google sample was drawn in March 2006. I worked with Microsoft and Google to find mutually acceptable methods to draw the samples and I generated the random numbers used to draw the samples. Yahoo! also provided 1 million webpages from its index, but those data were not reliable enough to use: two domains—www.cracks.me.uk and the anesthesiology department at the University of Washington—comprised about 5% of the Yahoo! sample. At the time, www.cracks.me.uk had a sexually explicit banner ad. (A domain is the "root" of a web address. For example, statistics.berkeley.edu is a domain; statistics.berkeley.edu/~stark/index.html is one of the pages in that domain.)

30 See footnote 28 above. The queries were from 2005: AOL from July 22–28, MSN from July 17–23, and Yahoo! from August 18–24. The AOL and MSN queries had weights, which appear to be the number of times each query was run each day. Protective orders prevent me from listing the number of queries each search provider produced. Problems

have the bulk of the search market, although AOL does not have its own search engine.[31] I drew a stratified random sample of the queries comprising 7,541 queries with total weight 15,461.

The random samples of webpages and queries were sent to CRA International, along with a list of 685 of the most popular searches from November 12, 2005 through February 20, 2006, according to Wordtracker.[32] A team at CRA International led by Paul Mewett attempted to view and classify 39,999 random webpages from the MSN index, 11,000 random webpages from the Google index, the first ten results of each of the randomly selected searches and the first ten results of the 685 Wordtracker searches.[33]

The classification scheme was nuanced so that alternative definitions of "harmful to minors" could be applied. For the present paper, only two categories matter: category 1a ("clean") and category 5f ("adult"). Category 1a webpages contain no reference to sex and no nudity. Category 5f webpages are "adult entertainment." The two categories are disjoint but not exhaustive. For example, a webpage that shows genitalia in an artistic or educational context would be in some other category.[34] In all, CRA International attempted to categorize 68,150 webpages. Of those, 5,045 were no longer working. Of the 63,105 webpages that were working at the time of testing, 60,833 were category 1a and 1,382 were category 5f. The remaining 890 were in other categories.

---

decompressing the AOL data resulted in the exclusion of 0.6% of the searches; those data would not have a material effect on the estimates. To protect the data providers, separate estimates for the three providers were not reported. The government issued a subpoena to Google for queries, but Google refused. The issue was litigated; Google was compelled to provide a random sample of URLs from its index, but no search queries. Gonzales v. Google, 234 F.R.D. 674, 688 (N.D. Cal. 2006).

31 It is my understanding that at the time, Google processed searches for AOL, and that the same query executed directly on Google or through AOL produced very similar results.

32 See note 27 above.

33 MSN queries were run on the MSN search engine, Yahoo! queries on the Yahoo! search engine, AOL queries on the Google search engine, and Wordtracker queries on Metacrawler. If a query retrieved fewer than ten results, all results were used.

34 COPA does not restrict such materials. See note 3 above.

CRA International determined the country in which each adult webpage was hosted. For foreign adult webpages that did not require payment, CRA International looked for commercial ties to the U.S. such as advertisements or links to commercial websites hosted in the U.S. Finally, CRA International tested whether a variety of Internet content filters blocked each of the adult webpages and a random sample of the clean webpages.[35] CRA International tested 15 combinations of filters and filter settings. The tests, run in the spring and summer of 2006, used the (then) latest version of each filter, with all updates included. The settings were either the default, settings intended for teenagers, or the settings the manufacturers' instructions indicated were tailored most narrowly to block pornography. Some of the filters were client-side programs; three (AOL, MSN, and Verizon) were services provided by ISPs.

CRA International sent me a database that classified webpages by content. The database showed which webpages were used to test each filter and whether those webpages were blocked. It also gave the country of origin of the category 5f webpages, and, for foreign "free" adult webpages, whether the page had commercial ties to the U.S. That database is the basis of all the estimates in this paper.[36]

Throughout this paper, "adult" means material CRA International would put in category 5f, and "clean" means material CRA International would put in category 1a. A webpage is "domestic" if CRA International would identify its host country to be the U.S. A foreign adult webpage has commercial ties to the U.S. if CRA International would say it does.

Results

The results are presented in tables 1 through 9. Table 1 lists the sizes of the populations and

35 They sent me a list of all category 1a webpages; I drew random samples and sent them back to CRA International.

36 I checked the internal consistency of the data, but beyond that, I did not verify the work CRA International performed.

samples in the study. Table 2 gives estimates of the percentage of adult (category 5f) and clean (category 1a) webpages in the universes of study. Table 3 gives conservative lower 95% confidence bounds on the percentage of adult webpages in the Google and MSN search indexes, and the percentage of webpages in those indexes that are adult webpages hosted in the U.S. Table 4 gives estimates of overblocking and underblocking of webpages in the Google and MSN search indexes. Table 5 gives lower 95% confidence limits for the entries in Table 4. Table 6 gives estimates of the percentage of adult webpages in search engine indexes that filters do not block that are hosted in the U.S.

Table 7 gives estimates of overblocking and underblocking for AOL, MSN and Yahoo! searches, by result and by search, and lower confidence limits on the percentage of searches that return at least one adult webpage that is not blocked. Table 8 reports underblocking and estimated overblocking for Wordtracker searches, by result and by search. Finally, Table 9 gives estimates of the percentage of nominally free adult webpages hosted abroad that have commercial ties to the U.S.

Technical Details

The estimates of the prevalence of adult webpages and of the rates at which filters fail to block adult webpages and block clean webpages have downward biases. For example, searches that did not retrieve any working webpages were included in the denominator of estimates of the prevalence of adult material. Category 5f is very restrictive: there must be sexual content that is clearly adult entertainment, and that content must be visible without clicking anything, including the “play” button of a video. Category 1a is also quite restrictive: there can be no nudity or sexual content whatsoever, not even in a medical, educational or artistic context. Those restrictions make it easier for filters to classify the webpages correctly. Months passed between collecting the webpages and testing the filters

which allowed time for the filter companies to classify the pages correctly.[37]

The Wordtracker queries are a population of queries: the measured prevalence of adult materials and the overblocking and underblocking rates for Wordtracker search results are essentially parameters rather than estimates.[38] In contrast, the samples of webpages from the Google and MSN search indexes and the samples of queries from AOL, MSN and Yahoo! are random samples from the corresponding populations of queries. I used sample percentages to estimate the percentage of adult webpages in the Google and MSN indexes, the percentage of adult webpages that originate in the United States, and overblocking and underblocking rates for webpages. The sample percentages are unbiased.

Because the sampling fractions are minute, the number of items in the sample with a given property has essentially a binomial distribution — the exact distribution is hypergeometric. The binomial approximation yields conservative confidence limits. I used the binomial distribution to find the lower 95% confidence limits in Tables 3, 5, and 7.[39]

I used weighted sample percentages to estimate the percentage of AOL, MSN and Yahoo! searches that had various properties. There were two sets of weights. AOL and MSN each provided weights that were apparently the daily frequencies of each query. Those weights were used to make estimates for the two services' searches. The overall estimates for the population of 1.3 billion searches were a weighted average of estimates for the vendors separately. The weights in the average were the fractions of searches the services contributed to the total pool. The resulting weighted sample

37 See notes 29 and 30 above. Filters were tested in spring and summer 2006.

38 Overblocking was estimated from a random sample of search results to save labor. A query run on a single search engine at two different times does not necessarily retrieve identical results; this could contribute a small amount of variability to the blocking rates.

39 The parameter n in the binomial distribution is equal to the sample size and the parameter p is equal to the proportion of items in the parent population that have the property. The lower confidence limit is the smallest population proportion p such that the chance of observing a sample proportion at least as large as was observed is at least 5%. The limits are conditional on the number of working URLs in each sample. See, e.g., E. L. Lehmann and Joseph P. Romano, *Testing Statistical Hypotheses*, 3rd ed. (New York: Springer, 2005), 75. Lehmann and Romano compute analogous upper confidence limits

percentages are ratio estimators, which may be biased, although I expect the bias to be small because the sample sizes are moderate.[40]

The exact probability distribution of the weighted sample percentage is not computable in this problem. To have conservative—rather than approximate—lower confidence bounds required basing the bounds on something other than weighted sample percentages. The Wordtracker, AOL and MSN query data show that searches that retrieve adult webpages have higher than average weight.[41] Therefore, treating every query from a given search provider as having the same weight introduces a negative bias in estimates of the prevalence of adult materials. For each vendor, a conservative lower confidence bound based on the sample percentage without weights would, therefore, be lower than a bound that used the weights. I inverted binomial tests to find lower confidence limits (at level $(0.95)^{1/3}$) for each set of searches, then formed a weighted average of those confidence limits using as weights the fractions of searches the vendors contributed to the pool. Since the samples are independent, combining the confidence bounds this way is conservative: the overall confidence level is higher than 95%.

Conclusions

This study reports on webpages drawn from the Google and MSN indexes, AOL, MSN and Yahoo! search results, and results of 685 of the most popular searches according to Wordtracker. Data were collected in 2005 and 2006. Fifteen combinations of Internet content filters and filter settings

40 For example, for each vendor, the estimated percentage of queries that return at least one adult webpage was estimated as (total weight of queries in the sample that return at least one adult webpage)/(total weight of queries in the sample)x100%. Both the numerator and denominator are random. The expected value of a ratio is not generally equal to the ratio of the expected values, which introduces a bias of order $1/n$, where n is the sample size. The variance is of order $1/n^{½}$, so for large samples the variance dominates the bias. See, e.g., Cochran, *Sampling Techniques*, *supra* note 18, at 160–62; Kish, *Survey Sampling*, *supra* note 18 at 186.

41 I computed the average weights using the query data. The average weight was larger for the queries that retrieved adult pages.

were tested on the webpages in 2006.

About 1.1% of the webpages in the MSN and Google search indexes are "adult entertainment" (CRA International category 5f). Since the indexed portion of the web contains tens of billions of pages, 1.1% amounts to hundreds of millions of adult webpages. About 1.7% of AOL, MSN and Yahoo! search results were adult, as were about 14% of the results of the Wordtracker searches. A substantial percentage of adult webpages are hosted in the U.S.: about 44% of those in the Google index, 56% of those in the MSN index, 88% of those in the sample of search results, and 87% of those in the Wordtracker search results. About 6% of AOL, MSN and Yahoo! searches and 37% of the Wordtracker searches retrieve at least one adult webpage among the first ten results.

Filters vary widely in their performance, and there is a tradeoff between failing to block adult materials ("underblocking") and erroneously blocking clean materials ("overblocking"). Filters that block a large percentage of adult webpages also block a sizable percentage of clean webpages in error. For example, the most restrictive filter blocked about 91% of the adult webpages in the Google and MSN search indexes, but also blocked about 23–24% of the clean webpages in the indexes. On average, if that filter were applied to every webpage in the Google search index, the filter would erroneously block about 22.1 clean webpages for each adult page it blocks correctly.[42] For the MSN search index, it would block about 23.1 clean webpages erroneously for each adult webpage it blocked correctly. Less restrictive filters blocked as little as 40% of the adult webpages in the indexes. Those filters blocked fewer clean pages in error.

The performance of filters on search results is qualitatively similar. The most restrictive filter blocked about 94% of the adult webpages among search results, but erroneously blocked about 13% of

42 An estimated 98.1% of the webpages in the Google index are clean and 1.1% are adult (Table 2). The MSN filter at the Mature Teen setting blocks an estimated 22.6% of clean webpages and (100% - 8.9%) = 91.1% of adult webpages (Table 4). The estimated percentage of webpages in the Google index that are clean and blocked in error is (98.1% x 22.6%) = 22.171%, and the estimated percentage of webpages in the Google index that are adult and blocked correctly is (1.1% x 91.1%) = 1.002%. The ratio is 22.171%/1.002% = 22.1. The calculations for the MSN index and the results of AOL, MSN, Yahoo!, and Wordtracker searches are analogous.

clean search results. On average, it would block about 7.6 clean search results in error for every adult search result it blocks correctly. For the most popular searches, the most restrictive filter blocked over 98% of adult results, but also blocked nearly 20% of clean results. On average, the filter blocked about 1.1 clean results of popular searches erroneously for each adult result it blocked correctly.

These figures have biases in favor of filters: underblocking was measured using webpages that were unambiguously "adult entertainment" pages; overblocking was measured using webpages that had no nudity or sexual reference whatsoever**,** and a time lag of months between collecting the webpages and testing the filters gave the filters an advantage.

The amount of adult material on the World Wide Web is vast: there are hundreds of millions of webpages in the indexed web alone. Filters can reduce the amount of adult material children come in contact with, but they are far from perfect. Millions of adult webpages slip through even the most restrictive filters, which block an even greater number of webpages that they should not block.

	Google index	MSN index	AOL, MSN, and Yahoo! searches	Wordtracker searches
Webpages in sample	11,100	39,999	22,405	206 million
Working webpages in sample	10,009	36,557	21,870	195 million
Searches in population			1.3 billion	20.6 million
Searches in sample			2,345	20.6 million

Table 1. Sizes of populations and samples in the study. Numbers for searches are weighted by the frequency of the searches.

Source	Google index	MSN index	AOL, MSN, and Yahoo! searches	Wordtracker searches
Adult webpages	1.1%	1.1%	1.7%	14.1%
Domestic adult webpages	44.2%	56.7%	88.4%	87.4%
Searches with adult results			6.0%	37.1%
Searches with domestic adult results			5.7%	37.0%
Clean webpages	98.1%	98.3%	97.5%	76.5%

Table 2. Estimated prevalence of "adult" (category 5f) and "clean" (category 1a) webpages. Column 2: estimates for the Google search index. Column 3: estimates for the MSN search index. Column 4: estimates for the first ten results of searches on AOL, MSN, and Yahoo!. Column 5: estimates for the first ten results of Wordtracker's most popular searches. The first row contains the estimated percentage of webpages that are adult. The second row contains, of the adult webpages, the estimated percentage that are domestic. The third row contains the estimated percentage of searches that return at least one adult webpage among the first ten results. The fourth row contains the estimated percentage of searches that return at least one domestic adult webpage among the first ten results. Estimates for the Google and MSN indexes and the AOL, MSN and Yahoo! searches are based on random samples. Figures for Wordtracker are based on 685 of the most popular searches for November 12, 2005 through February 20, 2006.

	Google index	MSN index	AOL, MSN, and Yahoo! searches
Adult	1.0%	1.0%	2.5%
Domestic adult	0.4%	0.5%	2.2%

Table 3. Conservative 95% lower confidence limits for the prevalence of "adult" (category 5f) webpages. The second and third columns contain lower confidence limits for the percentage of adult webpages and of domestic adult webpages among all webpages in the Google and MSN indexes. The fourth column contains lower confidence limits for the percentage of the AOL, MSN and Yahoo! searches that return at least one adult webpage and for the percentage of AOL, MSN and Yahoo! searches that return at least one domestic adult webpage.

Filter	Underblocking		Overblocking	
	Google	MSN	Google	MSN
AOL Mature Teen	8.9%	8.6%	22.6%	23.6%
MSN Pornography	16.8%	18.7%	19.6%	10.3%
MSN Teen	17.7%	20.5%	21.9%	18.9%
ContentProtect Default	38.3%	45.4%	2.8%	3.0%
ContentProtect Custom	28.3%	46.7%	1.4%	0.7%
CyberPatrol Custom	31.0%	33.5%	1.4%	0.9%
CyberSitter Default	12.7%	16.5%	3.6%	4.1%
CyberSitter Custom	12.4%	18.9%	4.0%	3.7%
McAfee Young Teen	16.1%	26.0%	12.4%	13.2%
Net Nanny Level 2	44.0%	46.1%	3.3%	2.2%
Norton Default	60.2%	54.9%	1.4%	0.7%
Norton Custom	58.4%	54.2%	0.9%	0.4%
Verizon	41.8%	40.3%	9.4%	5.7%
8e6	18.3%	23.0%	9.4%	7.5%
SafeEyes	16.2%	15.2%	3.3%	3.2%

Table 4. Estimated underblocking and overblocking of webpages in the Google and MSN indexes. Among adult (category 5f) webpages, the percentage that are not blocked by a filter is the rate of underblocking. Among clean (category 1a) webpages, the percentage that are blocked by a filter is the rate of overblocking.

Filter	Underblocking		Overblocking	
	Google	MSN	Google	MSN
AOL Mature Teen	5.6%	6.5%	18.4%	21.0%
MSN Pornography	12.1%	15.7%	15.8%	8.5%
MSN Teen	12.8%	17.4%	17.8%	16.6%
ContentProtect Default	31.3%	41.3%	1.5%	2.1%
ContentProtect Custom	22.2%	42.6%	0.6%	0.4%
CyberPatrol Custom	24.6%	29.7%	0.6%	0.5%
CyberSitter Default	8.6%	13.6%	2.1%	3.1%
CyberSitter Custom	8.4%	15.9%	2.4%	2.7%
McAfee Young Teen	11.4%	22.5%	9.3%	11.3%
Net Nanny Level 2	36.8%	41.9%	1.9%	1.5%
Norton Default	52.9%	50.7%	0.6%	0.4%
Norton Custom	51.1%	50.1%	0.4%	0.2%
Verizon	34.7%	36.2%	6.7%	4.4%
8e6	13.1%	19.6%	6.7%	6.0%
SafeEyes	11.4%	12.3%	1.9%	2.3%

Table 5. Conservative 95% lower confidence limits for underblocking and overblocking of webpages in search

indexes. For illustration, at 95% confidence, the MSN filter at the "Teen" setting fails to block at least 12.8% of the adult (category 5f) webpages in the Google index. Similarly, at 95% confidence, the MSN Teen filter blocks at least 16.6% of the clean (category 1a) webpages in the MSN index.

Filter	Estimated Domestic Underblocking	
	Google	MSN
AOL Mature Teen	40.0%	40.6%
MSN Pornography	31.6%	42.9%
MSN Teen	40.0%	37.7%
ContentProtect Default	39.0%	45.8%
ContentProtect Custom	40.6%	47.1%
CyberPatrol Custom	48.6%	44.0%
CyberSitter Default	50.0%	32.8%
CyberSitter Custom	57.1%	36.2%
McAfee Young Teen	44.4%	37.5%
Net Nanny Level 2	41.7%	48.1%
Norton Default	35.3%	49.3%
Norton Custom	36.4%	49.7%
Verizon	37.0%	42.4%
8e6	42.1%	46.8%
SafeEyes	35.3%	40.4%

Table 6. Domestic underblocking. Of the adult (category 5f) webpages in the Google and MSN indexes that filters do not block, the estimated percentage that are domestic webpages.

Filter	Underblocking for results	Overblocking for results	Domestic Underblocking	Underblocking for searches	95% confidence limit
AOL Mature Teen	6.2%	12.5%	57.0%	15.6%	5.3%
MSN Pornography	21.4%	4.4%	86.1%	32.3%	20.9%
MSN Teen	20.8%	5.8%	91.9%	28.1%	18.8%
ContentProtect Default	18.4%	6.4%	70.1%	46.2%	10.0%
ContentProtect Custom	20.4%	0.0%	62.1%	42.2%	25.4%
CyberPatrol Custom	34.6%	0.4%	94.9%	65.6%	24.4%
CyberSitter Default	11.2%	4.6%	33.8%	23.2%	11.2%
CyberSitter Custom	10.0%	5.3%	44.1%	20.1%	8.1%
McAfee Young Teen	14.2%	20.7%	80.7%	30.9%	10.4%
Net Nanny Level 2	28.1%	3.7%	79.4%	36.6%	20.8%
Norton Default	42.1%	0.8%	85.3%	51.6%	49.3%
Norton Custom	43.4%	0.0%	85.6%	56.1%	54.3%
Verizon	23.1%	1.3%	80.9%	41.6%	31.4%
8e6	7.3%	7.5%	78.0%	23.4%	11.7%
SafeEyes	13.7%	1.9%	87.8%	29.8%	14.9%

Table 7. Estimated underblocking and overblocking of the results of AOL, MSN, and Yahoo! searches. "Underblocking for results" is the percentage of adult (category 5f) search results that are not blocked. "Overblocking for results" is the percentage of clean (category 1a) search results that are blocked. "Domestic underblocking" is the percentage of domestic webpages among adult search results that the filters do not block. "Underblocking for searches" is, among searches that retrieve any adult webpages, the percentage that retrieve at least one adult webpage that is not blocked. "95% confidence limit" is a conservative lower 95% confidence limit for underblocking for searches.

Filter	Underblocking for results	Overblocking for results	Domestic Underblocking	Underblocking for searches
AOL Mature Teen	1.3%	19.6%	69.2%	4.3%
MSN Pornography	2.7%	13.3%	86.1%	8.2%
MSN Teen	2.6%	13.7%	83.1%	8.3%
ContentProtect Default	7.5%	12.4%	84.1%	23.1%
ContentProtect Custom	8.1%	7.8%	84.9%	25.3%
CyberPatrol Custom	3.9%	9.2%	86.4%	10.1%
CyberSitter Default	1.4%	19.9%	69.3%	5.1%
CyberSitter Custom	2.9%	18.2%	84.0%	9.4%
McAfee Young Teen	2.8%	32.8%	70.7%	9.3%
Net Nanny Level 2	12.6%	9.5%	82.9%	34.4%
Norton Default	9.9%	4.8%	79.4%	25.2%
Norton Custom	10.2%	2.9%	79.4%	25.9%
Verizon	4.4%	16.1%	67.9%	15.0%
8e6	3.4%	25.1%	93.0%	10.3%
SafeEyes	2.0%	16.5%	96.6%	6.4%

Table 8. Underblocking and estimated overblocking for the results of Wordtracker searches. “Underblocking for results” is the percentage of adult (category 5f) search results that are not blocked. “Overblocking for results” is the percentage of clean (category 1a) search results the filter blocks. “Domestic underblocking” is the percentage of domestic webpages among the adult search results that the filters do not block. “Underblocking for searches” is, among the searches that retrieve any adult webpages, the percentage that retrieve at least one adult webpage that is not blocked. Overblocking was estimated from a random sample of clean search results. Underblocking was determined from all the adult search results.

Data Source	“free” adult webpages with commercial ties to the U.S.
Google index	90.3%
MSN index	89.8%
AOL, MSN, and Yahoo! search results	88.2%
Wordtracker search results	95.9%

Table 9. Estimated percentage of nominally free adult entertainment (category 5f) foreign webpages that have commercial ties to the United States.