International Journal of Communication 6 (2012), 980–1000 1932–8036/20120980

Information Capacity
IJoC

Measuring Consumer Information

ROGER BOHN
JAMES SHORT
University of California, San Diego

How much media information, of what kinds, and delivered on what devices, do Americans consume? We measure each consumer information stream using 3 different measures of what is consumed—hours, words, and bytes—and sum across each recipient. We estimate that, in 2008, Americans consumed about 1.3 trillion hours of information outside of work, an average of almost 12 hours per person per day. Media consumption totaled 3.6 zettabytes and 1,080 trillion words, corresponding to 100,500 words and 34 gigabytes for the average person on an average day. We measure information flows, not stocks, and find that information consumption measured in bytes grew at an annual rate of 5.4% from 1980 to 2008, only a few percentage points greater than GDP growth over this period. We report our findings for different media types, including television, the Internet and computer games, and discuss the utility of analyzing contrasting measures of information consumption in totaling how much media information Americans consume.

Introduction

How much information, of what kinds, and delivered on what devices do Americans consume? This turns out to be a tricky question. There are no agreed upon standards for most of the key words: the definitions of *information*, *media consumption*, *kinds*, and *how much* all affect the results. With the definitions we use, we estimate that, in 2008, Americans consumed about 1.3 trillion hours of information outside of work, an average of almost 12 hours per person per day. Media consumption totaled 3.6 zettabytes and 1,080 trillion words, corresponding to 100,500 words and 34 gigabytes for the average person on an average day. A zettabyte is 10 to the 21st power bytes or 1 million million gigabytes. One contribution of this article is that we use consistent methodologies and definitions across a wide variety of technologies and information types, from the very old (newspapers and books) to the very new (mobile devices, portable computer games, satellite radio, and Internet video). We count multiple formats and

---

Roger Bohn: rbohn@ucsd.edu
James Short: jshort@ucsd.edu
Date submitted: 2012–02–17

delivery channels, including print and electronic; voice, text, pictures, and video; analog and digital; and broadcast, person-to-person, and self-generated.

A second contribution of this article is that we measure each information stream using three different measures of what is consumed: bytes, words, and hours, summing across each recipient. The relative importance of each kind of information varies significantly across the three different measures of consumption. Surprisingly, different communities of scholars have very different reactions to these three measures. Those from an information technology background understand bytes, but don't see value in measuring words, while those from a traditional "media studies" background understand hours, but don't see any value in measuring bytes. One colleague in Communication Studies opined that a football game on TV was "the same amount of information," whether shown in high definition or in traditional analog format; the visible sweat droplets and blades of grass in one format over the other were not important. For ourselves, we conclude that all three measures give different and useful insights, although arguably none of them precisely captures the "value" to consumers of different kinds of information.

A third contribution is that we include nontraditional information sources that have not previously been counted, and we distinguish different bandwidths (bits per second) for all media. The most significant addition is computer games, which turn out to be heavy contributors to bytes consumed. We distinguish different resolutions for both television and computer games.

A final contribution is that we distinguish flows of information from stored information, and in this article we report only on flows. Some high-level measurements of information attempt to measure only stored information, while others measure some combination of storage and flows. Furthermore, we measure information as it is consumed, not when it is created. For example a phone call is consumed by two people but it is usually not stored. On the other hand, a home DVR may record hours of television programs each day that are never consumed and therefore are not counted.

Our analysis is heavily influenced by two previous research reports—*How Much Information 2000?* and *How Much Information 2003?*—both conducted by Professors Peter Lyman and Hal Varian (2000, 2003). However, there are some important differences in approach. Lyman and Varian measured "original" information, that is, the first instance of new information being created, such as a voice telephone call or raw film footage. They added up the quantity of original content created: how many original hours of radio broadcasting were produced worldwide, added to how many original book titles were published, and so on for 20-odd media types. In contrast, we defined "information" as all of the flows of data delivered to people for their consumption. We do not distinguish between original and replicated content, so *N* viewers of a TV program are counted as *N* separate streams. Our measures contrast with other approaches taken in this special issue: Hilbert and Lopez (2012, this Special Section) measure the changing technological capacity to digitally communicate information over long distances in optimally compressed bits, independent of the amount actually consumed by an individual and independent of whether the intended destination is people or machines; and Neuman, Park, and Panek (2012, this Special Section) report, among other things, media consumption in minutes per household, which is very similar to one of our measures reported in Table 1.

Defining Media and Information Metrics

We define three measures of media and information consumption: the number of *hours* that a consumer watched or listened to media, the number of *words* contained in that media, and the number of *bytes* required to present it, as shown in Table 1.

***Table 1. Three Measures of Information*.**

What is measured	Variable name	U.S. 2008 Consumption
Hours of consumption	$INFO_H$	1,273 billion hours
Words consumed	$INFO_W$	10,845 trillion words
Compressed bytes consumed	$INFO_C$	3,645 exabytes

We analyzed U.S. media consumers aged 2 and above, measuring the sum of fixed and mobile device non-office usage for all consumers.[1] Measuring bytes is bound to be controversial, as it appears to emphasize types of information that stream at very high rates (such as computer games), yet accounts for only a fraction of the words received or the hours spent consuming media information each day. We comment on this later.

We distinguish between data and information in measuring the volume of media actually delivered for consumption. We define *data* as artificial signals intended to convey meaning and *information* as data that is actually received by a person for use. The term *artificial* is used because we consider only data created by machines, such as microphones, cameras, environmental sensors, barcode readers, or computer keyboards. Streams of data from sensors are extensively transformed by a series of machines, such as cable routers (location change), storage devices (time shift), and computers (symbol and meaning change). These transformations, in turn, create new data. Most high-level studies of data have included only two kinds of data: that which gets stored, and data that is transmitted over long distances, such as over the Internet backbone. In our analysis, we include all data that is used inside the home, even if it was never stored and never transmitted outside the home, such as the display from a computer game.

For looking at consumers, we measure information that is delivered to people outside their place of employment, whether for entertainment such as TV, for communication (e.g., voice telephone or e-mail) or for any other reason. Our measurement location is a hypothetical bubble around each individual's head—"information to the eyeballs." Therefore, if three people are watching television in a household, we

[1] There are different estimates for the U.S. "digital population" for different age groups and households. For our purposes, we used data from the U.S. Census for 2008, combining data compiled for 18+, 12+, and 2+ age groups where available and appropriate.

measure it as three streams of information. If one of the TV viewers is also answering e-mail on a notebook computer, we count both activities.[2]

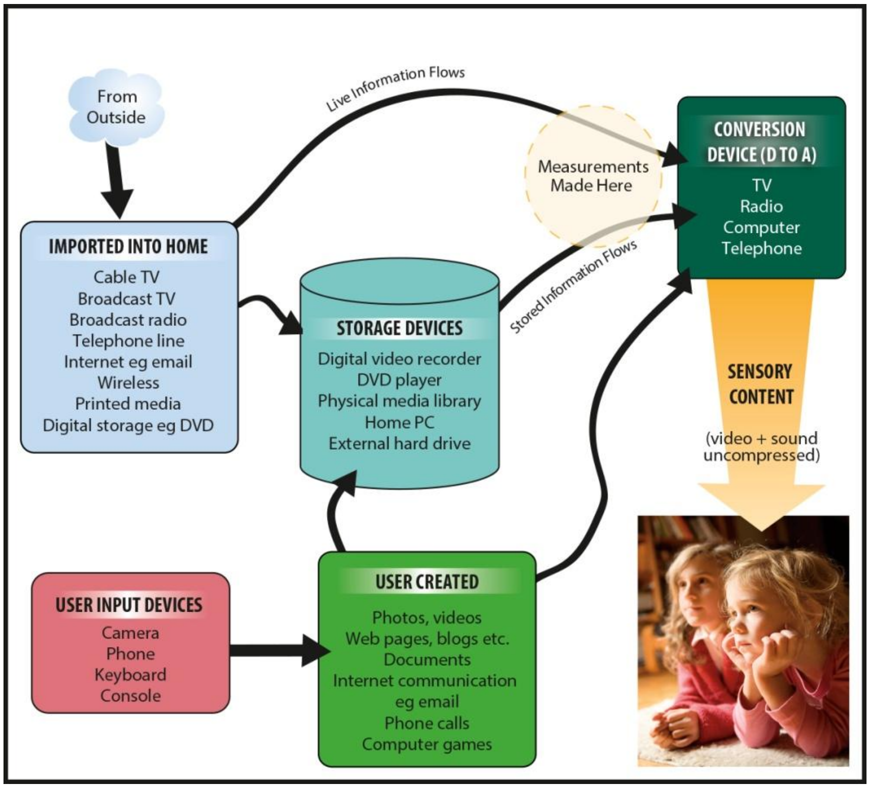

Figure 1. Information flows in a home.

Figure 1 illustrates some of the data flows around a home. The data displayed directly to consumers, shown by the wide arrow, is the information we are interested in. Data that enters a home but is never viewed, such as an unread newspaper, is not counted. See Neuman et al. (2012, this Special Section) for some estimates of its magnitude. Once information "reaches the eyeballs," it is counted; we do not attempt to estimate whether the viewer is paying attention.

[2] In other words, we do not adjust for double counting in our analysis. If someone is watching TV and using the computer at the same time for one hour, our data sources record this as two hours of consumption. This method is consistent with that of most other researchers. Note, however, that this means there are theoretically more than 24 hours in an information day. The use of multiple simultaneous sources of information and how much people really use the media is analyzed extensively in the Middleton Media Studies (Papper, Holmes, & Popovich, 2004).

There are a wide variety of types of media consumed daily such as:

• Text on a printed page or cell phone display
• Moving pictures on a TV, in a movie theater, or on a computer screen
• MP3 audio track received through earphones or speakers
• Electronic spreadsheet.

Our calculations for measuring the information used by consumers start by breaking media down into 26 high-level categories as illustrated in Table 2.

Table 2. Partial Breakdown of Delivery Methods Analyzed.

Television	Cable TV – SD (Standard Definition)
	Over air TV - SD
	DVD
	Cable TV – HD (High Definition)
	Over air TV - HD
	Satellite - HD
	Satellite - SD
	Mobile TV
	Other TV (Delayed View)
	Internet video
Print Media	Newspapers
	Magazines
	Books
Radio	Satellite Radio
	AM Radio
	FM Radio
Phone	Fixed Line Voice
	Cellular Voice
Computer	High-end Computer gaming
	Computer gaming
	Console gaming
	Handheld gaming
	Internet including email
	Offline programs
Movies	Movies in theaters
Music	Recorded Music

For each medium, we estimate the number of people who use it and the average number of hours of use per user per year. The data on numbers and hours was compiled from 200+ government and industry sources, including the U.S. Census, the Bureau of Labor Statistics and other government sources, Nielsen, ComScore, PQ Media and other industry sources, and a variety of academic, industry, and government studies on special topics. These sources, in turn, used a variety of surveys and observation methods. A complete list of sources is contained in our (Bohn & Short) 2009 report.

A pioneering media study by Pool (1983) in the early 1980s reduced measurements for television and some other media to compensate for partial attention by recipients. Many consumers engage in other behaviors while receiving information streams, such as driving, eating, and socializing. Television, for example, may be "on in the background" while doing chores. Consumers also receive multiple information streams simultaneously. Such "multitasking" behavior has probably increased considerably since the 1970s, but we have no basis for measuring it. Therefore, unlike Pool, we conceptually count every word that reaches consumers' eyes and ears, regardless of whether it "reaches their brains."

Summary of 2008 Consumption

Next, we report on three measures of consumer use of media information: hours, words, and bytes. To get total consumption, we sum over the various media as shown in Figure 2.

Total information for a year from technology *Z*, population segment *M*

= Average daily hours of *Z* use per person in segment *M*

x Total number of people in *M* who use *Z*

x 365 days per year

x 3,600 seconds per hour

x Information per second for *Z* (bytes or words)

Total for technology *Z* = Sum over all population segments *M*

Figure 2. Formula for size calculations.

All of our numbers are estimates built from published information about American's media consumption and related topics. We converted between annual and daily per-capita averages using an estimated digital population of 295 million (ages 2 and up).

How Many Hours?

In considering U.S. households, a natural question is how much time Americans spend with different sources of media information. Our time statistics for U.S. households in 2008, including use of mobile phones and movie-going, are tabulated in Figure 3.

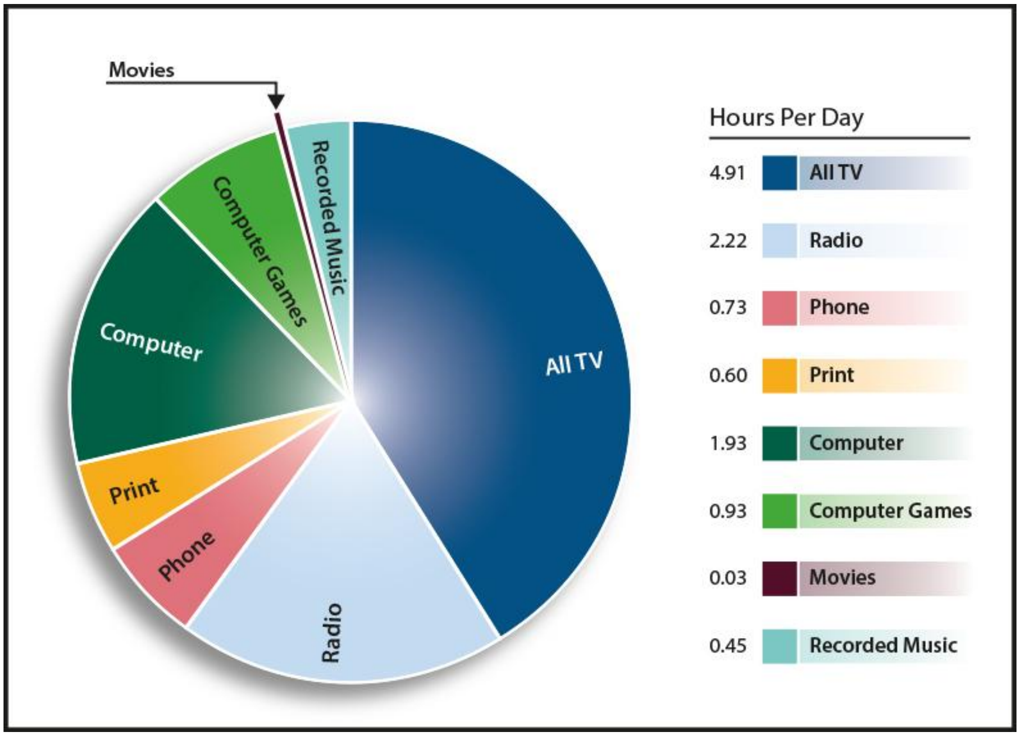

Figure 3. $INFO_H$ hourly information consumption.

We define $INFO_H$, one of our three measures of information, as "hours spent receiving information." Like our other measures, we estimate it for the total U.S. population and the entire 2008 year, and then convert it to per capita and per day.

Our hourly statistics confirm that a large chunk of the average American's day is devoted to receiving information. We estimate that an average American on an average day receives 11.8 hours of information. Considering that on average we work for almost three hours a day and sleep for seven, this means that three-quarters of our waking time in the home is spent receiving information, most of it electronic.[3]

[3] Because of the large number of children, retirees, and other non-employed persons in the population, only about half of the population works—on average about 250 days per year and 7.6 hours per day as reported by the most recent *American Time Use Survey* (annual) from the Bureau of Labor Statistics. We include all information at home, so a small amount of our totals is information consumed at home for work purposes. Americans average about 0.3 hours per capita per day working at home. Most byte-intensive activities are entertainment, so this adjustment would have almost no effect on compressed bytes consumed (INFOc).

We estimate that on average, 41% of information time is watching TV (including DVDs, recorded TV and real-time watching). An additional 19% involves listening to the radio, even though this activity is increasingly relegated to our daily commute. In other words, traditional media still dominated U.S. households in 2008, based on how much time ($INFO_H$) we spent consuming information: More than seven hours spent watching TV and listening to the radio accounted for more than 60% of total information hours. By comparison, computers accounted for 24% of $INFO_H$ time (including browsing the Internet, playing computer games, texting, watching videos on the PC, and so on).

How do we compare with Americans in the past? Not surprisingly, $INFO_H$ has gone up. The per capita time spent consuming information has risen nearly 60% from 1960 levels—from 7.4 hours per day in 1960 to 11.8 in 2008, a growth rate of only 1% per year. Neuman et al. (2012, this Special Section) report roughly similar totals, with some differences: Their data is per household, and it does not include use of computers or handheld/mobile devices except when they are being used to view the Internet.

Of course, our hypothetical "average American on an average day" is a composite of many different people. For example, although adults frequently complain about how much time children spend watching TV, American teenagers watch less than four hours per day while the largest amount—more than seven hours per day—is watched by Americans 60 to 65.

How Many Words?

In the 1960s, the concept that we now know as *bytes* barely existed. Therefore, early efforts to size up the information economy used words as the best barometer for understanding consumption of information. Using words as his only metric, Pool (1983) estimated that 4,500 trillion words were **"**consumed**"** in 1980. Using Pool's metric, we calculate that words consumed grew to 10,845 trillion words in 2008, which works out to about 100,000 words per American per day. "Words consumed" is our second metric, which we label $INFO_W$. We calculate it by multiplying the amount of consumption time $INFO_H$, by the rate of consumption in words per minute for that medium.

How Many Bytes?

While the statistics based on hours and words are useful, especially when analyzing trends, we now live in a digital age when most of the information we consume comes in digital formats. Music is consumed via MP3 devices, "newspapers" can be read online, and virtually all electronic devices are now based on digital integrated circuits. Therefore, an appropriate way to measure information is by the number of bytes consumed. We call this measure $INFO_C$, where the *C* stands for "compressed bytes."

Our formula for measuring bytes of information starts from $INFO_H$, the measure of hours. For each media type, such as high definition TV, we estimate the rate at which information is delivered (called the "bandwidth"), traditionally measured in bits per second. Multiplying the bandwidth by the number of hours and adjusting the units gives bytes per hour. Determining the correct bandwidth to use, however, is quite literally "tricky." The reason is that communications engineers use a variety of techniques to transmit information as rapidly and economically as possible. The definition we use for $INFO_C$ is the rate at which compressed information is transmitted over the link between the originator and the consumer. This rate is sometimes only 1% of the uncompressed rate, especially for so-called high-definition TV.

When information moves through a communication system to a viewer, it is converted into different formats multiple times. The appropriate measure used for $INFO_H$ is the lowest bandwidth along the transmission path. For example, an original analog movie on film being shown on television is first digitized onto a hard drive. The bytes on the hard drive are heavily compressed and sent digitally over the cable network to a home. There, they are decompressed and displayed on a high-definition TV screen. The screen converts them to analog light signals that are registered on the viewers' retinas. After the retina, the signals pass through four successive stages in the central nervous system as the viewer "sees" them. The bandwidth of the human retina is higher than that of the original film. But the measure we used was the slowest part of the system, namely, the digital bandwidth of the cable TV channel—4 megabits per second.

Not all information is transmitted in the usual meaning of the term. For example, newspapers and movies are mostly still delivered physically on analog media, and computer games are created on the fly and viewed on a home screen. In these cases, we developed measures of bandwidth that would be needed to transmit the same information over a digital link in compressed form, with no loss of resolution. That is, when no traditional transmission channel was used, we estimated the bandwidth that would have been needed for lossless compression. Computers are discussed in section 4.2. For printed media, such as books and newspapers, we used the bandwidth needed to transmit the text at the speed it is read, plus a slight amount of photographs, totaling little over 1,000 bits per second (bps) for books and 18,000 bits per second (bps) for newspapers.[4]

Whatever the precise definitions used for measuring $INFO_C$, one fact stands out: Moving pictures (video) of various kinds, when measured by bytes, dominate all other types of consumer information. Even photographs are tiny by comparison. A high-resolution digital picture in 2008 might be 10 megabytes, but this is equivalent to only 20 seconds of a standard TV picture. This led to a big surprise: Only three activities— television, computer games, and movies in theaters—contribute a significant amount of information based on $INFO_C$, as Figure 4 illustrates. Everything else adds up to less than 1%.

[4] This approach is conceptually similar to the way Hilbert and Lopez (2012, this Special Section) measure communication capacity. They emphasize changes in compression effectiveness over time, which we do not cover. Their criteria for what they refer to as "optimal compression rates" are that it be undetectable to viewers. We follow essentially this approach for computer games, but we measure what is actually displayed, which is a small fraction of the theoretical capacity.

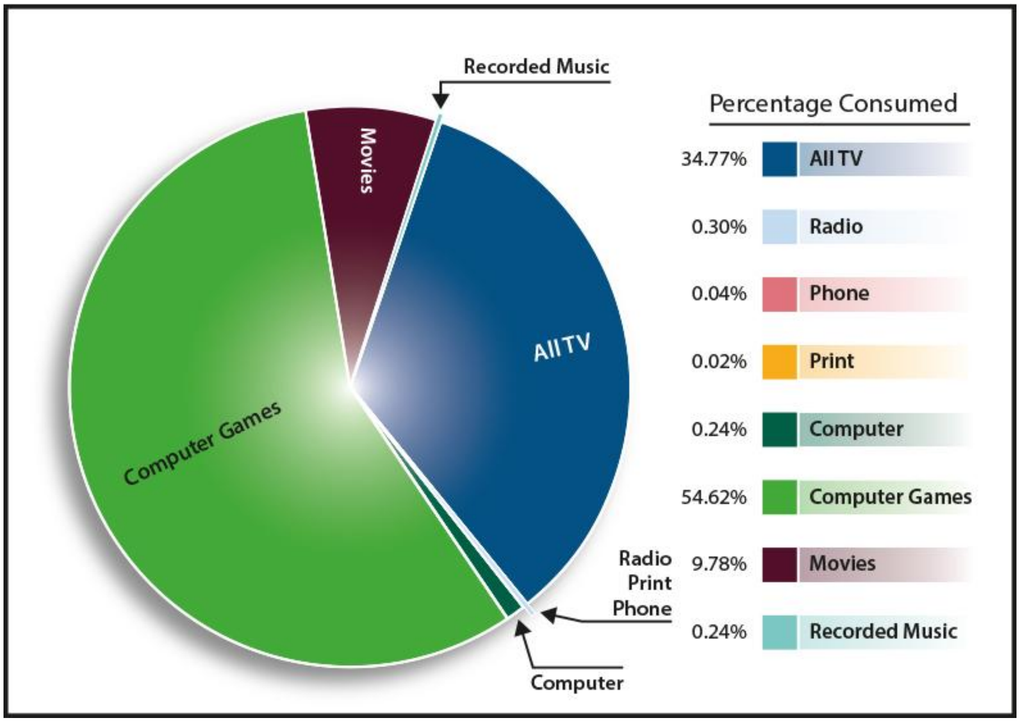

Figure 4. INFO$_C$ consumption in compressed bytes.

In total, we estimate that an average American consumed about 34 gigabytes of digital information per day in 2008. That number of gigabytes would fit on about 7 DVD disks, or 1.5 Blu-ray disks, or about one-fifth of an average notebook computer's hard drive, depending on when you last purchased the computer. About 35% of the digital information consumed was from television, 10% from movies, and 55% from computer games. Compared to the 140% increase in total words consumed from 1980 to 2008, there was a 350% increase in the number of bytes consumed—to 3.6 zettabytes. The higher growth of bytes than words reflects the faster growth in visual media (TV and computers) than in audio and textual media (radio and print), as well as the increasing bytes per hour for those visual media as the technologies improved.

Special Topics in Media Measurement

We now consider six special topics prompted by our assumptions and results: stored data, computer games, media quality, Internet delivery, interactive media, and information value.

Information Flows Versus Storage

The majority of consumer information use is transient: Information is streamed into the home and quickly thrown away without ever being stored. More precisely, it is created (for example, in a computer game) or received from a remote site (TV), "stored" for a few milliseconds while it moves through the display electronics, then is consumed and discarded. An example is a video game, where each frame is "stored" for only 33 milliseconds in the frame buffer of a graphics card. Frame buffers are made out of dynamic random access memory (DRAM), a common type of semiconductor chip; digital TVs use similar technology. DRAM and other volatile memory are so fast that they are not generally classified as storage.

Stored data and flows of data must be measured in different units. Storage is normally measured in bytes at a snapshot in time, while flows must be measured in bytes over time—typically, one second at the micro level and one year at the aggregate level. Physically, the way this works is that each GB of storage can be used many times over the course of one year. We measure information not as it sits in storage, but as it is consumed by a person over time. Stored data does not necessarily become information, and information is generally not stored at the point of use.

The primary consumer storage media include books, DVDs, CDs, MP3 players, computer hard drives, and increasingly, hard drives in digital video recorders (DVRs). The total amount of nonvolatile storage worldwide at the end of 2008 was roughly 200 exabytes. American consumers owned approximately 10% of that. In other words, the 20 exabytes of home digital storage, if it were all used, would be enough to hold only about two days of consumer information flow. Equivalently, the storage could act as a two-day buffer on incoming information. For example, digital video recorders work this way, as they automatically overwrite old TV programs when needed to make space for new ones.

Overall, we estimate that about 87% of bytes were streamed and thrown away; 12% were recorded briefly, used once, and thrown away (newspapers, magazines, DVRs, movie theaters), and 1% were consumed from "permanent" household media (books, purchased DVDs, e-mail, music on MP3s or CDs, and some Internet use). Using words or hours as the measure, however, approximately doubles the importance of stored media. Print media, Internet, music, and home computer use are about 37% of words and 25% of hours.

Measuring Computer Games

Data on computer use in the home is tracked much less than is data on TV use, but we attempted to analyze computer use as much as feasible. A large, but generally unmeasured category of computer use turned out to be computer games. We view them as "movies" or "stories" that are generated on the fly by the player, then immediately thrown away. Social games are somewhat more complex, but the amount of information that flows over the Internet in a social game is less than 1% of what the players see. In either case, by our definitions, they are displaying information just as a TV program does.

We then had to estimate bytes and words per hour of game play. It is easy to measure the theoretical raw bytes from games, determined by image resolution. For example an Xbox 360 game console connected to a 720p digital TV generates up 30 frames per second, each 1,280 by 720 pixels. This corresponds to 559,872,000 bits per second, or 0.6 Gbps. Television programs deal with this by sending compressed signals that are then expanded in the TV by ratios of 50:1 or more. We decided to use the equivalent compression ratio for a computer game: If the game had actually been transmitted using lossless compression, how many bits per second would be needed? We estimated this by running compression tests on a variety of computer games, from simple casual games to high-resolution strategy games. Depending on the type of game and type of device it was played on, our estimated information rates ranged from 2.7 Mbps for a casual game on a good cell phone to 180 Mbps for a complex game on a high-end video card.

For words per minute, we had a student examine various games of different genres. The variance in words, both between and within games, is very large. In a few games, the only words are at the start of the game; the actual content is entirely visual and musical. We settled on 20 words per minute for casual games, and 50 for others. More detailed results are in Appendix B of our (Bohn & Short) 2009 report.

The results surprised us: Computer games were 2.4% of words, 8% of hours, but the largest category of bytes—55% of the total. The reason for the high byte content is that games can have 50 times higher resolution than can TV signals after both are compressed. Other forms of computer use were more balanced, as discussed in the following Internet section. These results compare with the finding of Hilbert and López (2011) that graphics processing units constituted 97% of all computational capacity in the world in 2007. GPUs are intensively used when playing games, either in dedicated game consoles or in desktop personal computers. Then again, such machines and their GPUs have relatively low load factors (ratio of actual use to maximum possible use), which explains why actual bytes are a lower percentage than is MIPS capacity.[5]

Media Quality: Variety Versus Visual Resolution

Some of the benefit of cheaper information technology has been in the form of more choices of what to consume. The number of TV channels per average household has now reached over 130, of which the average household actually watches 18.[6] Both numbers are considerably higher than they were in 1980. This is an example of a more general phenomenon: The ratio of information available to information consumed grows over time. (Neuman et al., 2012, this Special Section)

The additional channels of TV, however, have come at a cost: higher compression and therefore lower video resolution for the channels we receive. The issue is straightforward: Bandwidth costs money (all those transistors). For a fixed budget, a cable TV company, and especially a satellite TV company have only a fixed total capacity in megabits per second. Suppose it allocates 600 megabits per second (Mbps)

---

[5] We are not suggesting that the ratio of MIPS to pixels or video bytes is fixed, but that more processing power (MIPS) allows more complex graphics to be rendered.

[6] Nielsen's (2008) *Television Audience Report* states that the average U.S. television household received a total of 130.1 station channels as tuning options that year (the total includes digital cable and satellite channels, and 17.7 channels of over-the-air broadcast). Growing digital cable and satellite penetration has increased the tuning options for the average household. In 2006, the average total available was 104 channels. In 2008, the average household actually watched 18 channels or approximately 14% of the total station channels available.

to broadcast TV. If it divides this capacity into 130 channels, their bandwidth must average 4.6 megabits per second. Total bandwidth can be split between high definition channels (at roughly 12 Mbps each) and standard definition channels (4 Mbps each), but most of the 130 will have to be standard definition. Or they could provide half as many channels and double the average bandwidth, or any other combination, as long as the total is 600 Mbps. For example, CSPAN or weather could be given 2 Mbps, while a sports channel could receive 16 Mbps.

It appears that most TV carriers have chosen to go for lower bandwidth per channel and more channels. Almost no broadcasts are close to the full resolution of 1080i that many TV sets are now capable of receiving. In fact, channels advertised as "HDTV" are sometimes so compressed that the pictures are far below the theoretical capability of the TV set.[7] The same issue comes up for broadcast stations that are each given the use of 16 Mbps of bandwidth and typically divide it into multiple channels.

Assuming that this accurately reflects what TV viewers want, this tells us that American consumers generally prefer variety (more choices) over sheer visual quality. But one result has been the very slow growth in average bytes per hour of $INFO_C$ bandwidth. Presumably, over the next ten years, the mass migration to HDTV-capable sets will gradually lead to an increase in average bandwidth and information consumption. It's not clear how quickly carriers, networks, and display manufacturers will give consumers the full HD experience that many consumers assume they are already getting.

How Much Information is Delivered via the Internet?

Another question is the quantitative importance of the Internet: How much does it contribute to our information consumption? Our basic finding is that the Internet provides a substantial portion of some kinds of information, but very little of others. Figure 5 shows that, the Internet provides a significant fraction of our information, although less than does television, when measuring with hours or words.

[7] "Resolution" is more than the number of pixels. It includes frames per second and the degree of compression. A program can theoretically be 1080i, but still be so heavily compressed that it is no more attractive visually than a standard definition (480i) program would be.

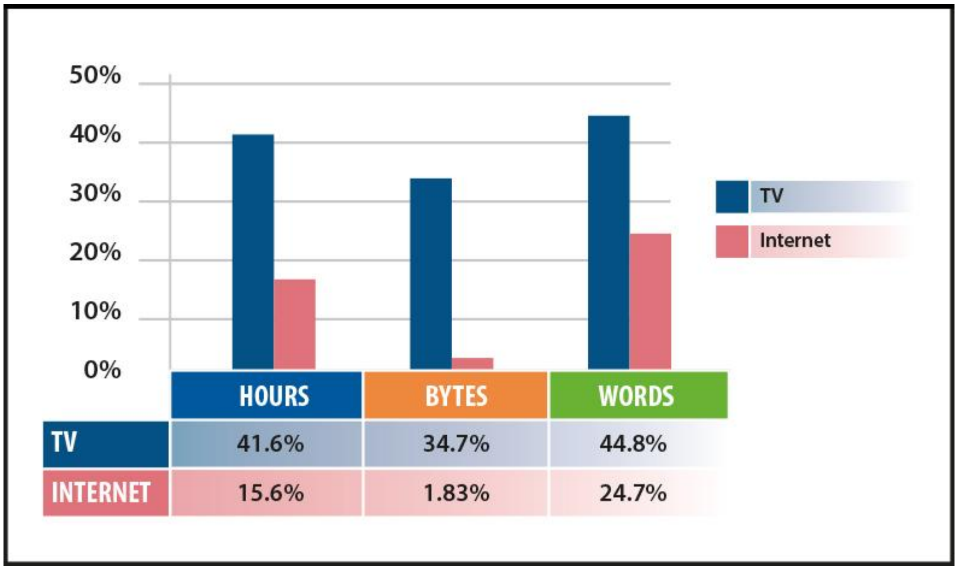

Figure 5. Internet as a source of information.

The Internet was the source of only 2% of our $INFO_C$ bytes (versus 35% for TV). Our low estimate of Internet bytes has been criticized. Perhaps, this goes back to the instinct that bytes "ought to be" the correct measure of value, suggesting that our 2% number therefore implies that the Internet is "not very valuable." Surveys show that many Americans view the Internet as very important, to the extent that they will cut spending on cable TV before they cut Internet access. How can this be reconciled with the Internet's very low numerical share of bytes? Mathematically, the low byte count occurs because low bandwidth to most homes, plus low resolution of Internet video in 2008 (compressed 240p on YouTube, for example), was much less capable of pushing bytes into homes than were cable or broadcast TV. We estimated average Internet bandwidth at 1 Mbps for Internet streaming video and only 100 kbps for text consumption, such as web browsing. We will discuss this further.

$INFO_W$ and $INFO_H$ are arguably more meaningful metrics for the human importance of Internet services. Americans spent 16% of their information hours using the Internet (versus 41% for TV), and received 25% of words $INFO_W$ from it (versus 45% from TV). We expect all three Internet/TV information ratios (bytes, words, hours) will rise rapidly as bandwidth to the home and medium quality video content become more available.

We can classify information consumption into three purposes: two-way communication, pure entertainment, and gathering factual information of any kind, which we labeled "research/current events." There is not much information about how TV programs fit into these categories, but based on our own rough analysis, we estimate that Americans average 6.5 hours per day on entertainment and 3.7 hours on research/current events. The distinction we make between these categories is that entertainment is fictional, so that game shows and other "reality" programs are classified as research/current events. The

Internet's contribution to pure entertainment information is very small—less than 2%, whether measured by hours, bytes, or words. This stems from entertainment's dominance by video activities: TV shows, movies, and computer games. Video requires very high bandwidth, and Internet speed to most Americans is still far below what is needed to watch conventional live television. A standard TV program requires approximately 4 megabits per second of bandwidth, while most Internet connections in 2008 could deliver only a fraction of that, especially at peak times. Even when the "last mile" to a house is capable of adequate speeds, the Internet provider is using statistical multiplexing, meaning that it assumes that only a fraction of users will be operating at full speed at the same time. If everyone turned on their "Internet TV" at 7 p.m., many parts of the network would be unable to handle the load.

Two-way communication is self-explanatory. Before the Internet, the only ways to have a two-way communication at a distance were telephones and first-class letters. The Internet adds multiple methods. We estimate that Americans averaged 1.6 nonwork hours per day conducting two-way communication, of which 57% was via the Internet, with the rest of the time spent on cellular or landline telephones. Correspondingly, the Internet provides 79% of the bytes and 73% of the words in two-way communication. The Internet is so important for two-way communications because of its unique technical characteristics, including a nearly universal network, very low variable costs, and the ability to handle both real-time and delayed activity.

In our third-use category, research and current events, the Internet provides 23% of our hours and 31% of our $INFO_W$. It connects to vast amounts of factual information, making it very good for current events that can be delivered in the form of text. We classify about one-third of television programming as research /current events (including not only news but also reality shows, talk shows, and the like), so television dominates the total bytes in this category. Given the much higher bandwidth of TV, again the Internet provides only 1.3% of our research/current event bytes.

Interactive Media

Most sources of media information in the past were consumed passively. Listening to music on the radio, for example, did not require any interaction beyond selecting a channel and setting the volume control. Telephone calls were the only interactive form of information, and they are only 5% of words and a negligible fraction of bytes. The arrival of home computers, however, dramatically changed this, as computer games are highly interactive. Most home computer programs, such as writing or working with user generated content, are interactive as well. Arguably, web use is also highly interactive, with multiple decisions each minute about what to do or click on next.

As a result, we estimate that one-third of our $INFO_W$ in words is now received interactively and 55% of our $INFO_C$ in bytes. This is an overwhelming transformation, and it is not surprising if it causes cognitive changes. These changes may not all be good, but they will be widespread. On the other hand, we are only measuring artificial forms of information. For most of human evolution, we spent most of our days interacting with our environment and with each other, without artificial assistance. In fact, if we include "personal conversation" as a source of information, it is possible that we receive fewer bytes $INFO_C$ than our ancestors did 100 years ago. The reason is that conversation is very "high bandwidth." A full fidelity video link between two locations, including stereo vision and sound, is not possible with present technology; the observer will realize they are not physically in the same location. If we could do it, however, it would require (conservatively) 100 million bits per second. Three hours of personal conversation a day at this bandwidth would generate 135 gigabytes of $INFO_C$, about four times the average daily consumption today.

Media Information Value

Hours, words, and bytes measure the volume of information, not its value. There are many potential criteria for measuring the value of a stream of information, including subjective judgment, selling price, willingness to pay by the users of the information, development cost, and audience size. But there is no clear metric for value, especially when comparing information of different kinds.

Take for example a landmark speech, Abraham Lincoln's Gettysburg Address versus a current TV series, a 2008 episode of "Heroes" on NBC. The speech took roughly 2.5 minutes to deliver and was 244 words in length, i.e., 1,293 characters or bytes of text. No one is sure exactly what Lincoln said; his handwritten texts do not match contemporary accounts. The direct cost of writing and delivering the Address was probably less than $5,000 (valuing Lincoln's time at $200 an hour in today's dollars). In contrast, a 2008 episode of "Heroes" on NBC ran 44 minutes in length (without commercials), the master version occupies 10 GB of digital storage, and each special effects-laden "Heroes" episode cost an estimated $4 million to produce. So, by any quantitative measure, the popular TV program would be considered to offer much more information than did the Gettysburg Address. Yet Lincoln's words were far more important, and most of the world would agree that they were, in most senses of the word, far more "valuable" and worthy of saving for posterity.

In our analysis of information, we measure neither original delivery time nor bytes of storage, but total bytes of all copies across all recipients. "Heroes" episodes in the 2008–2009 season had just over 10 million viewers, including those views on DVRs within a week of the broadcast. Reruns could push that figure to 18 million viewers per episode. Optimistically, let us guess that the Gettysburg Address has been read twice by every American who reached 6th grade since Lincoln uttered the words in 1863. This is approximately 500 million people, multiplied by two readings, which equals 1 billion readings. So, measured by the pure number of information consumers, Lincoln's one-billion readership trumps the 10 million average weekly viewers for "Heroes" by a factor of 100. In contrast, looking at bytes, a compressed episode of "Heroes" on an average TV comes in at about 500 MB, times 10 million views, which equals 5 petabytes. In contrast, the 1 billion readings of the Gettysburg Address are only 2.4 terabytes. Looked at this way, NBC's "Heroes" wins by a factor of 2,000.

Which of the two "information" events, then, is larger in value? The pure volume of information does not determine its value or impact. The right information, delivered at a key time and place, can move a nation. At the other extreme, raw bytes are now so inexpensive that we often pay no attention to them. Therefore, our study eschews efforts to determine the value of one type of information over another in favor of estimating the volume of information consumption.

Conclusions: Analyzing the Growth of Information Consumption

While the estimated five-fold increase from 1980 to 2008 in $INFO_C$ bytes consumed is impressive at one level, this is an annual growth rate of only 5.4% and is far less than the rate of increase in most measures of digital technology, which tend to be driven by Moore's Law. For example, the cost of hard disk storage in a 1982 personal computer was about $50 per megabyte; today, it is less than $1 per gigabyte, a 50,000-fold improvement.[8] Nordhaus (2007) found that the cost of computation fell faster than 60% per year from the mid-1980s to 2006. The total reduction was five orders of magnitude, a cost/performance improvement factor of 200,000.

If the total revenue of an industry is constant, then the quantity of its output must grow in inverse proportion to its price/performance ratio. In fact, revenue for both the semiconductor industry and the electronics industry grew between 1980 and 2008. So the capacity to process bytes must have grown at a rate somewhere between 30% (the lower end of Moore's Law estimates) and 60% per year. How is it possible, then, that consumption $INFO_C$ grew at only 5.4% per year, only a few percentage points faster than GDP did over the period?

We analyze this by decomposing growth in $INFO_C$ into three components. Total information consumed is the product of three factors: the American population, average hours per person spent consuming information, and average information per hour. We decomposed total growth into these components for the period 1980 to 2008:

- Population grew at 0.95% per year, from 226 million to 295 million (ages 2 and up).
- Average hours of information consumption per person grew at 1.7% per year, from 7.4 hours to 11.8 hours of $INFO_H$.
- Weighted average bandwidth across all media grew at 2.8% per year from 2.9 Mbps (megabits per second) to 6.4 Mbps. This is a measure of "information intensity" of our consumption.
- Gigabytes per person per day grew at an annual rate of 4.4%, from 9.8 to 33.8 Gigabytes of $INFO_C$. Not coincidentally, 4.4% is the sum of the growth rates in hours per person and in average bandwidth.

If there is one major surprise in this study, it is that $INFO_C$ consumption and information intensity per hour grew at these low rates from the dawn of personal computing in 1980 to today, despite Moore's Law and the revolutionary shift from analog to digital technology in most information media. Slow growth in the U.S. population is well-known, and the 1.6% per year growth in hours of consumption per person is understandable, given the constant 24-hour length of a day. But the 2.8% compound annual growth rate in bytes consumed per hour remains a drop in the bucket compared to the doubling every two years in the

[8] For an analysis of storage costs over time, see Grochowski and Halem, 2003.

number of transistors on an integrated circuit. Given how cheap information processing is today compared with 1980, why aren't we consuming hundreds of times more bytes per hour than we did in 1980?

There is one basic mathematical explanation of this result—very slow growth of $INFO_C$ from television. Color television, the dominant source of bytes in 1980, remained largely unchanged until the very recent switch to high definition TV in the U.S. market. And because high definition TV in 2008 was in less than half of households and accounted for less than half of the TV viewing hours in those households, it had little impact on average bytes per hour from TV. Finally, TV viewing time as a share of our information day was approximately unchanged. Putting together the slow growth in hours of TV and the minimal change in the quality of TV signals, bytes from TV grew slowly.

But this arithmetic does not get at the essence of the issue. First, why did TV picture quality stay stagnant for so long? Second, the capacity of information technology has been increasing at Moore's Law speeds, as documented in this Special Section. Intel and the rest of the semiconductor industry sell more devices, with more transistors per device, every year, and America's share of worldwide consumption has been roughly constant. If these transistors, then, were not being used to consume more bytes, where did they go? Third, personal computers now occupy a major share of our information consumption, and when measured in hours or words, so does the Internet. Will their growth change the historical trajectory in the future?

Where are the Missing Bytes?

We have tentatively identified four places where the missing bytes have gone, although further research will be needed to confirm and measure them (Table 3).

Table 3. Explaining the Gap between Consumption and Capacity Growth.

Cause	Explanation	Example
Growth of information available over information consumed	We have far more choices of what to consume	Average household now receives 120 TV channels, but still watches only about 10 hours per day
Dark data	Much of the world's data now flows between machines, without human intervention or awareness	Automobiles now contain more than 50 processors each
Enterprise information	This report only considers consumer information	
Low load factor	We can afford multiple redundant devices	TVs in the kids' bedrooms

First, we have measured information consumed by consumers, but the amount of information available to them has grown much faster. Second, our analysis looks only at consumer information. We have completed a study of information in enterprises, which shows different and much faster growth patterns (Short, Bohn, & Baru, 2010). Third, is the reduction of load factors. Our houses today are full of electronic devices that we use for only hours or minutes a month. Even devices that we use every day, such as cell phones, contain transistors that have capabilities that we may never use, such as built-in GPS and Bluetooth.

A final factor is the rise of "dark data." When electronics were expensive, devices were naturally reserved for high-value activities. People and information worked closely together. But now 1 million transistors costs less than one cent, yet people's time is still valuable. We can no longer afford, nor do we need, to have people closely scrutinizing data as it is created and used. Instead, we hypothesize that most data is created, used, and thrown away without any person ever being aware of its existence. Just as cosmic dark matter is detected indirectly only through its effect on things that we can see, dark data is not directly visible to people.

The Future of Information Consumption

We have discussed each medium of information in turn, using three different measures (hours, compressed bytes, and words), as well as a range of reference points, including percentages, yearly totals, and daily consumption. As Figure 6 illustrates, $INFO_C$ bytes are completely dominated by video sources: movies, TV, and computer games. On the other hand, consumption time ($INFO_H$) is primarily used for video and audio (radio, telephone, and recorded music). Words, finally, come heavily from text sources (newspapers, magazine, books, and Internet use).

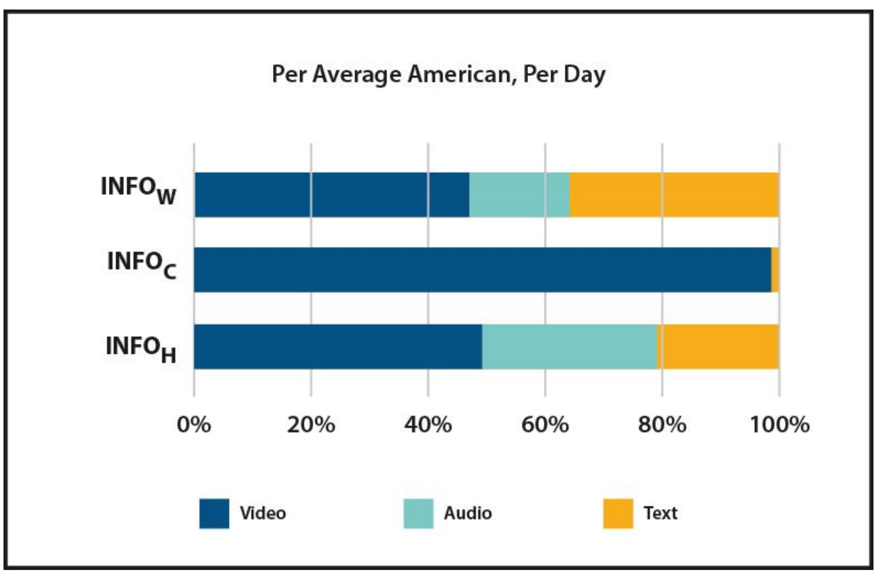

Figure 6. Shares of information in different formats.

There are some patterns of information consumption in the first half-decade of the 21st century that may be considerably changed by 2015. The significance of these changes, however, is not clear and may not become clear for some time. Perhaps the most visible is shifts in television. We have already discussed rapid changes in the delivery of television from 2005 to today, including the shift to digital broadcasting, the mass acceptance of high definition TV sets (although not high definition programming), and digital video recorders becoming a mass-market product. Then again, the number of TV channels has grown steadily for 50 years, yet actual video quality has not grown nearly as fast, as a simplistic theory of technological progress (Moore's Law) seemingly predicted.

Two nascent developments may also cause significant dislocations: mobile television and video over the Internet. So far, mobile TV has low utilization and is very much a niche product. On the other hand, video by Internet is quite widespread, but to date as a complement rather than as a substitute for conventional TV program delivery. YouTube and its cousins have made a huge variety of novel and specialized video material available to anyone with a mediocre broadband connection. But at least in the U.S., the quality of video over the Internet is far below what is available by more "conventional" means such as cable TV. The reason again is basically bandwidth constraints. A minimal standard definition TV signal requires 4 megabits per second, while a "medium" version of so-called high-definition TV requires double or triple that. The result is that Internet videos are generally small, or grainy, or downloaded gradually rather than streamed. If and when a substantial number of Americans are able to receive streaming video at sustained speeds of roughly 10 megabits per second and low latency, it will dramatically alter the way they receive video. Internet-based television, rather than being reserved for material where low quality is compensated for by a very wide selection (the "long tail effect"), might become common for mainstream programming as well.

Beyond television, computer games will be an area for growth of consumption $INFO_C$. The performance of GPUs follows Moore's Law and will continue to do so. In consequence, game-playing enthusiasts will consume rapidly increasing numbers of zettabytes. Casual gamers have shown little interest in high-resolution graphics so far. But at least for a few years, rapid growth in consoles and high-end computers will drive faster growth in $INFO_C$ bytes.

Consumption of words and hours, $INFO_W$ and $INFO_H$, are destined to continue their slow growth. They are constrained by human physical limits, including the length of a day and reading speed. Their growth will never exceed a few percent per year.

References

Bohn, R. E., & Short, J. E. (2009). *How much information? 2009 report on American consumers.* Global Information Industry Center, UC San Diego, CA. Available at http://hmi.ucsd.edu/howmuchinfo.php

Bureau of Labor Statistics. (Annual). *American time use survey.* Available at http://www.bls.gov/tus

Grochowski, E., & Halem, R. D. (2003). Technological impact of magnetic hard disk drives on storage systems. *IBM Systems Journal*, *42*(2), 338–346.

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, *332*, 60–65.

Hilbert, M., & López, P. (2012). How to measure the world's technological capacity to communicate, store and compute information? Part I: Results and scope. *International Journal of Communication, This Special Section on How to Measure "How Much Information?"*

Lyman, P., Varian, H. R., Dunn, J., Strygin, A., & Swearingen, K. (2000). *How much information 2000?* University of California at Berkeley. Retrieved from http://www2.sims.berkeley.edu/research/projects/how-much-info

Lyman, P., Varian, H., Swearingen, K., Charles, P., Good, N., Jordan, L., & Pal, J. (2003). *How much information 2003?* University of California at Berkeley. Retrieved from http://www2.sims.berkeley.edu/research/projects/how-much-info-2003

Neuman, R., Park, Y. J., & Panek, E. (2012). Tracking the flow of information into the U.S. home. *International Journal of Communication*, *6*, *This Special Section on How to Measure "How Much Information?"*

Nielsen (2008). *Television audience report.* Retrieved from http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/07/tva_2008_071709.pdf

Nordhaus, W. D. (2007). Two centuries of productivity growth in computing. *The Journal of Economic History*, *67*(1), 128–159.

Papper, R. A., Holmes, M. E., & Popovich, M. N. (2005). Middletown media studies: Media multitasking. *The International Digital Media & Arts Association Journal*, *1*(1), 5–43.

Pool, I. (1983). Tracking the flow of information. *Science*, *211*, 609–613.

Short, J. E., Bohn, R. E., & Baru, C. (2010). *How much information? 2010: Report on enterprise server information.* Global Information Industry Center, UC San Diego. Available at http://hmi.ucsd.edu/howmuchinfo.php