



# HMII?

How Much Information?



How Much Information? 2009  
Report on American Consumers

Roger E. Bohn  
James E. Short

Global Information Industry Center  
University of California, San Diego

December 9, 2009

# How Much Information? 2009 Report on American Consumers

<b>1 INTRODUCTION.....</b>	<b>8</b>
1.1 DATA AND INFORMATION .....	10
1.2 WHAT IS INFORMATION?.....	10
1.3 HOW MANY HOURS?.....	11
1.4 HOW MANY WORDS?.....	12
1.5 HOW MANY BYTES?.....	13
1.6 STORAGE VS. CONSUMPTION .....	14
1.7 VALUING INFORMATION.....	14
<b>2 TRADITIONAL INFORMATION IN U.S. HOUSEHOLDS .....</b>	<b>15</b>
2.1 TELEVISION .....	15
2.2 RADIO.....	17
2.3 TELEPHONE .....	17
2.4 PRINT.....	18
2.5 MOVIES .....	18
2.6 RECORDED MUSIC .....	18
<b>3 COMPUTER INFORMATION IN U.S. HOUSEHOLDS .....</b>	<b>19</b>
3.1 COMMUNICATING AND BROWSING THE INTERNET .....	19
3.2 INTERNET VIDEO.....	20
3.3 COMPUTER GAMING .....	21
3.4 OFF-INTERNET HOME COMPUTER USE .....	22
3.5 SMART PHONES .....	22
<b>4 TRENDS, PERSPECTIVES AND THE FUTURE OF U.S. INFORMATION CONSUMPTION .....</b>	<b>24</b>
4.1 ANALYZING THE GROWTH OF INFORMATION.....	24
4.2 WHERE ARE THE MISSING BYTES? .....	25
4.2.1 DARK DATA .....	25
4.2.2 TWO KINDS OF QUALITY: VARIETY AND RESOLUTION .....	26
4.3 ANALYZING INFORMATION CONSUMPTION.....	26
4.3.1 HOW MUCH INFORMATION IS DELIVERED VIA THE INTERNET? .....	27
4.3.2 THE RISE OF INTERACTION .....	28
4.4 THE FUTURE OF CONSUMER INFORMATION.....	28
<b>APPENDIX A: UC BERKELEY HMI? STUDIES .....</b>	<b>30</b>
<b>APPENDIX B: DETAIL TABLE.....</b>	<b>32</b>
<b>ENDNOTES.....</b>	<b>34</b>

## Tables and Figures

Figure 1 Information Flows in a Home.....	10
Figure 2 INFO <sub>H</sub> Hourly Information Consumption.....	11
Figure 3 Formula for Size Calculations.....	12
Figure 4 INFO <sub>W</sub> Consumption in Words.....	12
Figure 5 INFO <sub>C</sub> Consumption in Compressed Bytes.....	13
Figure 6 Evolution of Reading.....	18
Figure 7 Average Daily Consumption of Bytes, INFO <sub>C</sub> .....	20
Figure 8 Example of a Graphics Processing Card.....	21
Figure 9 Screen Shot from <i>NBA Live 10</i> .....	23
Figure 10 Shares of Information in Different Formats.....	26
Figure 11 Contrasting Measurements of INFO <sub>H</sub> INFO <sub>C</sub> and INFO <sub>W</sub> .....	26
Figure 12 Internet as a Source of Information.....	27
Table 1 Three Measures of Information.....	9
Table 2 Partial Breakdown of Delivery Methods Analyzed.....	11
Table 3 Television and Radio Consumption.....	16
Table 4 Telephone Consumption.....	17
Table 5 Conventional Media.....	19
Table 6 Computer Use Non-Gaming.....	19
Table 7 Computer Game Playing.....	21
Table 8 Explaining the Gap Between Consumption and Capacity Growth.....	25
Table 9 Summary of Information for Major Groups.....	27

# Acknowledgements

This report is the product of industry and university collaboration. We are grateful for the support of our industry partners, sponsor liaisons, university research partners, and administrative staff at the University of California, San Diego. Special thanks for research and writing assistance provided by L. Lin Ong and Doug Ramsey. Early support was provided by the Alfred P. Sloan Foundation of New York.

Financial support for HMI? research and the Global Information Industry Center is gratefully acknowledged. Our sponsors are:

**AT&T**

**Cisco Systems**

**IBM**

**Intel Corporation**

**LSI**

**Oracle**

**Seagate Technology**

The authors bear sole responsibility for the contents and conclusions of the report.

Questions about this research may be addressed to the Global Information Industry Center at the School of International Relations and Pacific Studies, UC San Diego:

Roger Bohn, Director, [rbohn@ucsd.edu](mailto:rbohn@ucsd.edu)

Jim Short, Research Director, [jshort@ucsd.edu](mailto:jshort@ucsd.edu)

Pepper Lane, Program Coordinator, [pelane@ucsd.edu](mailto:pelane@ucsd.edu)

Press Inquiries:

Doug Ramsey, Communications Director, [dramsey@ucsd.edu](mailto:dramsey@ucsd.edu), (858) 822-5825

<http://hmi.ucsd.edu/howmuchinfo.php>

# How Much Information? 2009 Report on American Consumers

Roger E. Bohn  
James E. Short

## Executive Summary

In 2008, Americans consumed information for about 1.3 trillion hours, an average of almost 12 hours per day. Consumption totaled 3.6 zettabytes and 10,845 trillion words, corresponding to 100,500 words and 34 gigabytes for an average person on an average day. A zettabyte is  $10^{21}$  bytes, a million million gigabytes. These estimates are from an analysis of more than 20 different sources of information, from very old (newspapers and books) to very new (portable computer games, satellite radio, and Internet video). Information at work is not included.

We defined “information” as flows of data delivered to people and we measured the bytes, words, and hours of consumer information. Video sources (moving pictures) dominate bytes of information, with 1.3 zettabytes from television and approximately 2 zettabytes of computer games. If hours or words are used as the measurement, information sources are more widely distributed, with substantial amounts from radio, Internet browsing, and others. All of our results are estimates.

Previous studies of information have reported much lower quantities. Two previous *How Much Information?* studies, by Peter Lyman and Hal Varian in 2000 and 2003, analyzed the quantity of original content created, rather than what was consumed. A more recent study measured consumption, but estimated that only .3 zettabytes were consumed worldwide in 2007.

Hours of information consumption grew at 2.6 percent per year from 1980 to 2008, due to a combination of population growth and increasing hours per capita, from 7.4 to 11.8. More surprising is that information *consumption* in bytes increased at only 5.4 percent per year. Yet the *capacity* to process data has been driven by Moore’s Law, rising at least 30 percent per year. One reason for the slow growth in bytes is that color TV changed little over that period. High-definition TV is increasing the number of bytes in TV programs, but slowly.

The traditional media of radio and TV still dominate our consumption per day, with a total of 60 percent of the hours. In total, more than three-quarters of U.S. households’ information time is spent with non-computer sources.

Despite this, computers have had major effects on some aspects of information consumption. In the past, information consumption was overwhelmingly passive, with telephone being the only interactive medium. Thanks to computers, a full third of words and more than half of bytes are now received interactively. Reading, which was in decline due to the growth of television, tripled from 1980 to 2008, because it is the overwhelmingly preferred way to receive words on the Internet.

<http://hmi.ucsd.edu/howmuchinfo.php>

# 1 INTRODUCTION

The world is awash in information and data, the ‘raw material’ of information. The goal of the *How Much Information?* Project is to create a census of the world’s data and information in 2008. How much did people consume, of what types, and where did it go?

This first report conveys our findings about information at the U.S. consumer level. In other words, **how much information was consumed by individuals in the United States in 2008?** Our statistics include information consumed in the home as well as outside the home for non-work-related reasons, including going to the movies, listening to the radio in the car, or talking on a cell phone. It does not include information consumed by individuals in the workplace. Future reports will focus on information in companies and on a global scale.

We have reached a variety of conclusions about the uses of information in the digital age, especially in the nearly 30 years since IBM launched its first PC in 1981 (which went on to become Time magazine’s “Man of the Year”). A few highlights:

- Americans spend a huge amount of time at home receiving information, an average of 11.8 hours per day.
- Bytes of information consumed by U.S. individuals have grown at 5.4 percent annually since 1980, far less than the growth rate of computer and information technology performance.
- Roughly 3.6 zettabytes (or 3,600 exabytes) of information were consumed in American homes in 2008. Americans spend 41 percent of our information time watching television, but TV accounts for less than 35 percent of information bytes consumed.
- Computer and video games account for 55 percent of all information bytes consumed in the home, because modern game consoles and PCs create huge streams of graphics.

Our estimate of 3.6 zettabytes for U.S. household information consumption is many times greater than the findings of previous studies. One zettabyte is  $10^{21}$  bytes, or 1,000 exabytes, while one exabyte is  $10^{18}$  bytes – a billion gigabytes (see inset). A 2007 IDC report estimated that total worldwide digital data would not reach one zettabyte until 2010. (One major factor accounting for this discrepancy is that IDC probably did not include video gaming, or most TV, in its calculations.)

## Counting Very Large Numbers

Byte (B)	=	1 byte	=	1	=	One character of text
Kilobyte (KB)	=	$10^3$ bytes	=	1,000	=	One page of text
Megabyte (MB)	=	$10^6$ bytes	=	1,000,000	=	One small photo
Gigabyte (GB)	=	$10^9$ bytes	=	1,000,000,000	=	One hour of High-Definition video, recorded on a digital video camera at its highest quality setting, is approximately 7 Gigabytes
Terabyte (TB)	=	$10^{12}$ bytes	=	1,000,000,000,000	=	The largest current hard drive
Petabyte (PB)	=	$10^{15}$ bytes	=	1,000,000,000,000,000	=	AT&T currently carries about 18.7 Petabytes of data traffic on an average business day
Exabyte (EB)	=	$10^{18}$ bytes	=	1,000,000,000,000,000,000	=	Approximately all of the hard drives in home computers in Minnesota, which has a population of 5.1M
Zettabyte (ZB)	=	$10^{21}$ bytes	=	1,000,000,000,000,000,000,000	=	



The rate of growth of information bytes consumed, 5.4 percent yearly on average, is low in light of the more familiar, exponential growth linked to Moore's Law: the rate of improvements in computer processing power, memory, storage, and other digital technologies. But over 28 years this growth adds up, constituting a four-fold increase in bytes and a 140 percent increase in words "consumed" by Americans from 1980 to 2008.

Our results are based on our own definitions of information and how to measure it. Appendix A provides some comparisons with a previous generation of studies, conducted by Professors Peter Lyman and Hal Varian at UC Berkeley, published in 2000 and 2003. We also compare some of our numbers with two industry reports produced by EMC and the International Data Corporation (IDC) in 2007 and 2008. These studies asked different questions, used different definitions, and got different results. Comparisons with information flows in 1980 largely draw on the groundbreaking work of Ithiel de Sola Pool, who used words as the metric for his observations. For example, he quite literally counted how many words were uttered in radio programs. Pool did not analyze bytes, so we have reanalyzed some of Pool's data to make them more directly comparable to those we use in 2008.

We have looked at words, bytes, and the number of hours spent consuming information in the household. These three measures show very different pictures about the volume of information in any given medium. (**Table 1 Three Measures of Information**) Take radio: Americans spent nearly 19 percent of their information hours listening to the radio, which accounts for 10.6 percent of words received – but barely one-quarter of one percent (0.3%) of the total bytes received each day. This points to radio's continuing role as a highly byte-efficient delivery mechanism for information.

## What does this report cover?

This study reports on consumers in the United States in 2008. Where 2008 data was not available, we extrapolated from earlier years. In some cases, behavior is changing so fast that January 2008 and December 2008 might be quite different; our goal was to report usage for the entire year but we were not always able to make an accurate adjustment for this situation. All of our data are estimates; see endnotes for sources.

While this report looks at all three measures of information consumption – hours, words and bytes – probably the most attention will be paid to bytes, given the prevalence of digital media. Measuring bytes is bound to be controversial, because it appears to emphasize types of information that stream at very high rates (such as computer games) yet account for only a fraction of the words or hours we spend consuming information each day. As we will show, moving pictures account for the vast majority of bytes. Therefore, we report the other measures as well.

The current report focuses on the U.S. household sector, while subsequent HMI? reports will expand the focus to a) the workplace, b) other regions, and c) new types of data and information that have no historical antecedent – notably the 'dark data' that is increasingly transmitted from machine to machine. This report is divided into five sections:

- Section 1 introduces our concepts and measurement methods.
- Section 2 looks at information consumed by U.S. consumers from traditional sources of information.
- Section 3 considers digital information and the computer revolution.
- Section 4 examines results, discusses some interesting special topics, and outlines future research.
- Appendices cover additional topics.

**Table 1: Three Measures of Information**

What is measured	Variable name	US 2008 Consumption
Hours of consumption	INFO <sub>h</sub>	1,273 billion hours
Words consumed	INFO <sub>w</sub>	10,845 trillion words
Compressed Bytes consumed	INFO <sub>c</sub>	3,645 exabytes

## 1.1 Data and Information

We distinguish between data and information. Information is a subset of data – but what is data? For our purposes, we define data as artificial signals intended to convey meaning. ‘Artificial,’ because data is 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.

Past high-level studies have generally measured data of only two kinds: data that gets stored, and data that is transmitted over long distances, such as over the Internet backbone. We greatly expand on these two categories. For example, we include data that is transmitted over a local area network (LAN), such as a home Wi-Fi (802.11) wireless network, and data that is never stored in a permanent way.

Indeed, data in the 21st century is largely ephemeral, because it is so easily produced: a machine creates it, uses it for a few seconds and overwrites it as new data arrives. Some data is never examined at all, such as scientific experiments that collect so much raw data that scientists never look at most of it. Only a fraction ever gets stored on a medium such as a hard drive, tape or sheet of paper. Yet even ephemeral data often has ‘descendants’—new data based on the old. Think of data as oil and information as gasoline: a tanker of crude oil is not useful until it arrives, its cargo unloaded and refined into gasoline that is distributed to service stations. Data is not information until it becomes available to potential consumers of that information. On the other hand, data, like crude oil, contains potential value.

## 1.2 What Is Information?

There are probably hundreds of definitions of information, and even the way we use the term in daily conversation changes depending on the topic. For looking at consumers, we choose to define information as data that is delivered for use by a person. Our measures of information include all data delivered directly to people at home, whether for personal consumption (such as entertainment), for communication (e.g., email) or for any other reason. Some data delivered to machines could also be considered information, but only if it is factored into a decision or action. We will not analyze it in this report. **Figure 1 Information Flows in a Home** shows 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.

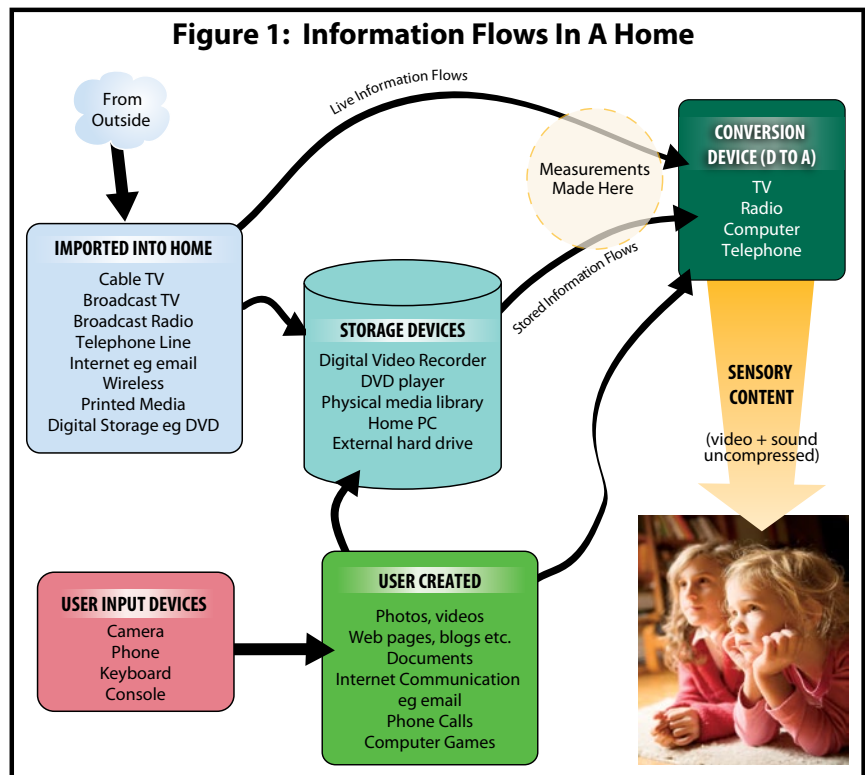
## Flows versus stocks of information

Our definition emphasizes *flows* of data – data in motion. We count every flow that is delivered to a person as information. Another approach goes to the opposite extreme: it counts data that is *stored* somewhere, such as a book, whether or not it is subsequently used.

As we will show, there are a wide variety of types of information consumed daily, such as:

- Text in readable form such as on a printed page or cell phone display;
- Moving pictures on a TV, in a movie theater or on a computer screen;
- An MP3 audio track received through earphones or speakers;
- An electronic spreadsheet

For the purposes of this study, information is ‘useful’ in itself, while data is only a means to ultimately produce information. In many situations, a lot of data is created and then filtered and manipulated to produce a relatively small amount of information. For example, the “signal strength” bar on your cellphone is the result of continuously monitoring radio signals from a cellphone tower. A 30 second TV commercial is the result of shooting, converting, and editing hours of raw footage. Data



## Simultaneous information

We do not adjust for double counting in our analysis. If someone is watching TV and using the computer at the same time, our data sources will record this as two hours of total information. This is consistent with most other researchers. Note, though, that this means there are theoretically more than 24 hours in an information day!

The use of multiple simultaneous sources of information is analyzed extensively in *Middletown Media Studies: Media Multitasking ... and how much people really use the media* by Robert A. Papper, Michael E. Holmes, and Mark N. Popovich.

can also be expanded, as when that TV commercial is sent to millions of TVs simultaneously.

### 1.3 How Many Hours?

In focusing on U.S. household consumption of information, a natural question is how much time Americans spend with different sources of information. Our time statistics for U.S. households in 2008 – including use of mobile phones and movie-going – are tabulated in **Figure 2 INFO<sub>H</sub> Hourly Information Consumption**. We estimate that an average American on an average day receives 11.8 hours of information a day. 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 receiving information, much of it electronic.<sup>1</sup> This is, indeed, the “Information Age.” We define “hours

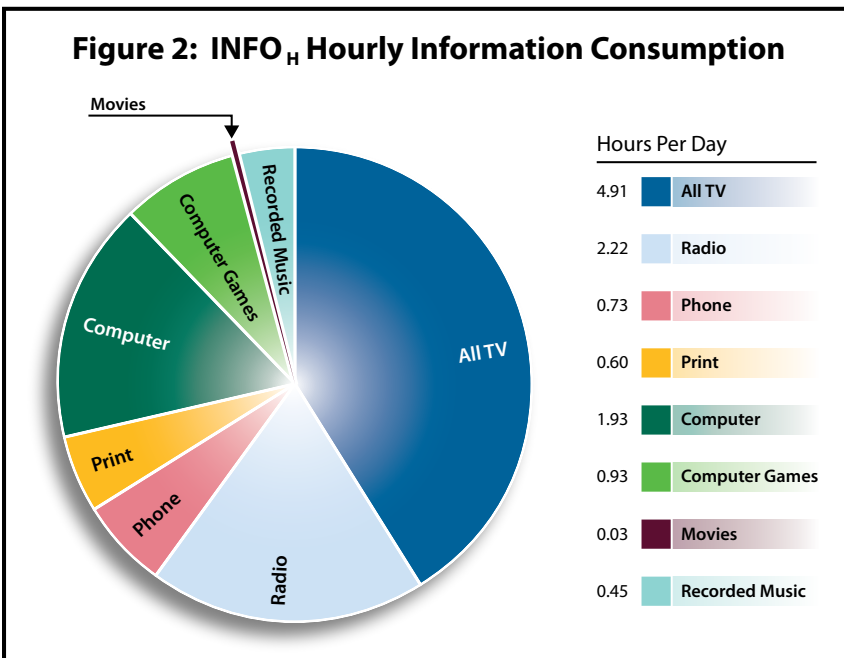
spent receiving information” as INFO<sub>H</sub>, one of our three measures of information.

Our calculations for measuring information used by consumers start by breaking information down into about 20 categories of delivery media. (**Table 2 Partial Breakdown of Delivery Methods Analyzed**) For each medium, we estimate the

**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

**Figure 2: INFO<sub>H</sub> Hourly Information Consumption**



number of people who use it, and the average number of hours per user each year. The data on numbers and hours comes from various sources, including the US Census and other government sources, Nielsen and other industry sources, and a variety of studies on special topics. These sources, in turn, used a variety of surveys and observation/methods.<sup>2</sup> Our hourly statistics confirm that a large chunk of the average American’s day is spent watching television. We estimate that on average 41 percent of information time is watching TV (including DVDs, recorded TV and real-time watching). An additional 19 percent of our information time 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 we spent consuming information: more than seven hours watching TV and listening to the radio, for more than 60 percent of total information hours. By comparison, computers accounted for 24 percent of  $INFO_H$  time (including browsing the Internet, playing computer games, texting, watching videos on the PC, and so on). So more than three-quarters of U.S. households' information time is spent with non-computer sources – despite the widespread belief that the seemingly ubiquitous computer now dominates modern life. (**Figure 2  $INFO_H$  Hourly Information Consumption**)

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, the facts show otherwise: American teenagers watch less than four hours per day while the largest amount is watched by older Americans, those 60 to 65, who watched more than seven hours per day.<sup>3</sup>

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 percent from 1960 levels – from 7.4 hours per day in 1960, to 11.8 in 2008. The forms of information media have also changed. When de Sola Pool did his analysis in 1980, he included a variety of media that either don't exist any more or are very small for consumers today. They included Direct mail, First-class mail, Telex, Telegrams, Mailgrams, and Fax.

### 1.4 How Many Words?

In 1960, digital sources of information were non-existent. Broadcast television was analog, electronic technology used vacuum tubes rather than microchips, computers barely existed and were mainly used by the government and a few very large companies, music recording used vinyl disks called “records,” and newspapers and magazines had black and white pictures, if they had any at all.<sup>4</sup> The concept that we now know as *bytes* barely existed. Early efforts to size up the information economy therefore used *words* as the best barometer for understanding consumption of information.

Using words as his only metric, Pool estimated that 4,500 trillion words were ‘consumed’ in 1980.<sup>5</sup> 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. This measure of information, 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 information consumed per unit of time. (**Figure 3 Formula for Size Calculations**) To get total consumption, we sum over the various media. All our numbers are estimates - see the on-line appendix and the endnotes for more information about data sources and methods. <[http://hmi.ucsd.edu/howmuchinfo\\_research.php](http://hmi.ucsd.edu/howmuchinfo_research.php)>

### Figure 3: Formula For Size Calculations

**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

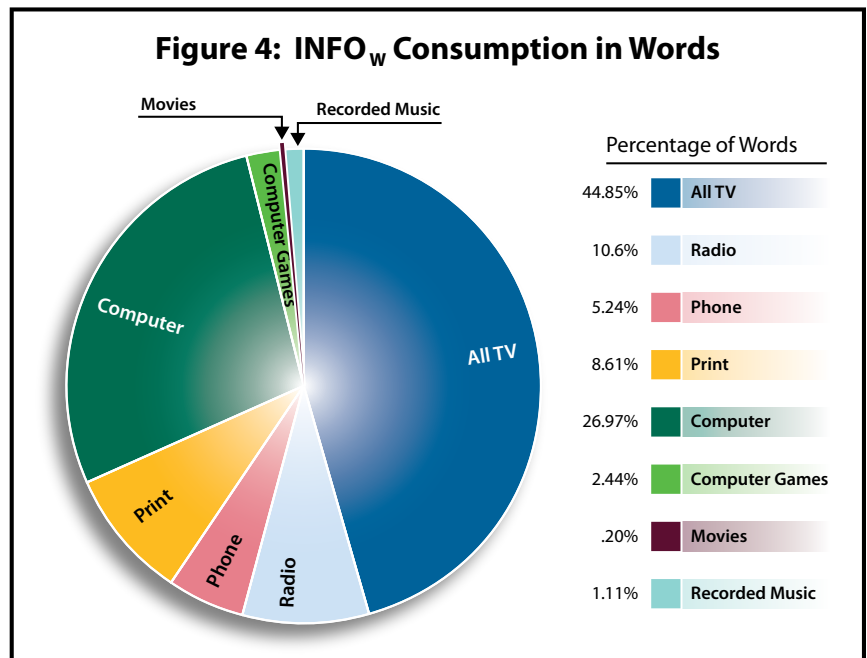
x 365 days per year

x 3600 seconds per hour

x Information per second for Z (bytes or words)

Total for technology Z = Sum over all population segments M

Comparing the 2008 statistics by type of media, TV remains the single largest source of information – over 45 percent of all words consumed. (**Figure 4  $INFO_w$  Consumption in Words**) In many categories, the percentage distribution of  $INFO_w$  and  $INFO_H$  is similar, such as with computers (24% of  $INFO_H$ , 27% of  $INFO_w$ ). A bigger difference between words and hours is for radio: in 2008 radio



accounted for about 10.6 percent of our daily information intake in words, even though we spent nearly 19 percent of our information time listening

to the radio. The reason is simple: a lot of radio programs are mostly music, with comparatively few words per minute.

### 1.5 How Many Bytes?

While the statistics based on hours and words are useful, especially when trying to draw conclusions about long-term trends, we now live in a digital age

measured in bits per second. Multiplying the bandwidth by the number of hours, and adjusting for the conversion between seconds and hours and between bits and bytes, gives the number of bytes for that category.

Determining the correct bandwidth to use, however, is quite literally “tricky.” The reason is that computer and communications engineers use a variety of tricks 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 one percent of the uncompressed rate, as we will discuss in the section on television. But not all information is actually “transmitted” in the usual meaning of the term. For example newspapers and movies are, for the most part, still delivered physically on analog media (paper and film, respectively). In these cases, we developed measures of bandwidth “as if” the information were transmitted over a digital link.

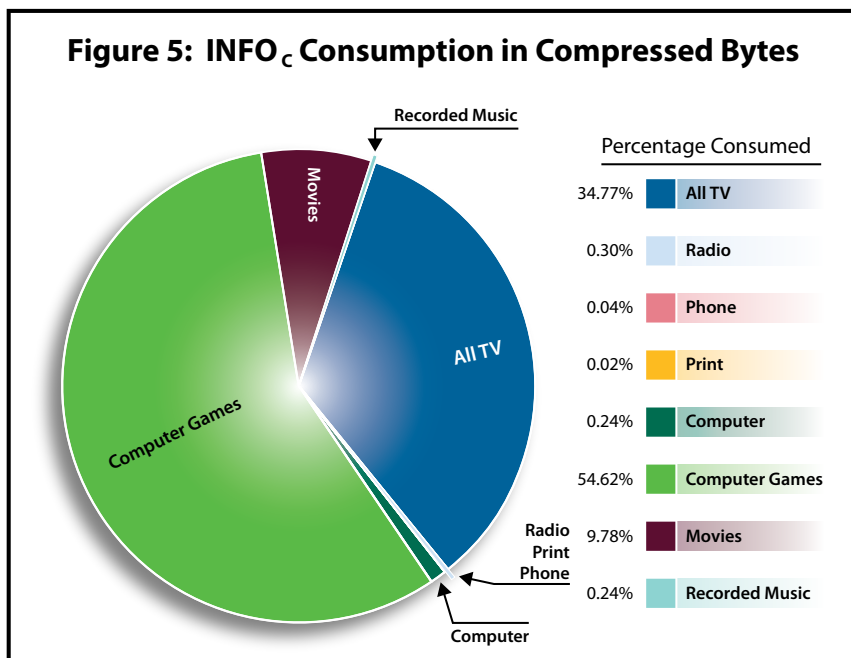
Whatever the precise definitions used for measuring  $INFO_C$ , one fact stands out: when measured by bytes, moving pictures dominate all other types of consumer information. Even photographs are tiny by comparison with most video. A high-resolution digital picture might be 10 megabytes, but this is equivalent to only 20 seconds of a standard TV picture.<sup>7</sup>

This led to a big surprise: only three activities contribute a significant amount of information based on  $INFO_C$ : television, computer games, and movies in theaters. Everything else adds up to less than one percent! (**Figure 5  $INFO_C$  Consumption in Compressed Bytes**)

In total, we estimate that an average American consumed about 34 gigabytes ( $3.4 \times 10^{10}$ ) bytes per day in 2008. 34 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 a computer. About 35 percent was from television, 10 percent from movies, and 55 percent from computer games. Computer games are a big story in themselves, and we will discuss them extensively in Section 3.

Compared to the 140 percent increase in total words consumed from 1980 to 2008, there was a 350 percent increase in the number of bytes consumed, to 3.6 zettabytes. The higher growth of bytes than words reflects faster growth in visual media (TV and computers) than in verbal and textual media (radio and print). We will discuss these growth rates fully in Section 4.1.

**Figure 5:  $INFO_C$  Consumption in Compressed Bytes**



when most of the information we consume comes in the form of 0s and 1s, of bits and bytes. Music is consumed via MP3 devices, ‘newspapers’ can be read online, and virtually all electronic devices are now based on digital integrated circuits.<sup>6</sup>

So it stands to reason that in the digital age, 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.”

Much of our research has gone into estimating  $INFO_C$ . 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 estimated the rate at which information is delivered, called the “bandwidth,” traditionally

## How much is 3.6 zettabytes?

If we printed 3.6 zettabytes of text in books, and stacked them as tightly as possible across the United States including Alaska, the pile would be 7 feet high.

## 1.6 Storage vs. Consumption

One implication of our definition is that *stored data is not necessarily information*. Storage is vital to shift data consumption forward in time, because someday it may be useful to create information. Some previous studies define stored data as “information.” But we classify it as data, until such time as the data is transmitted to the consumer for use.

For the purposes of this study, we measure data as information each time consumers use it. This measurement is feasible in the household sector, where the primary storage media include books, DVDs, CDs, MP3 players, computers and, increasingly, digital video recorders (DVRs). Indeed, our statistics for consumption of information are many times larger than total storage of data in those devices. According to some estimates, the total amount of hard disk storage worldwide at the end of 2008 was roughly 200 exabytes. In other words, the 3.6 zettabytes of information used by Americans in their homes

during 2008 was roughly 20 times more than what could be stored at one time on all the hard drives in the world.

The data ‘footprint’ of a storage device is not just how many bytes it holds, but how many bytes are created (both reads and writes) over time. Hence, for most storage devices, their nominal capacity is much *smaller* than the data that can be housed on the device over a period of time (as files are erased and replaced).

## 1.7 Valuing Information

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 consumers, development cost, and audience size. But there is no clear way of comparing value, especially when comparing information of different kinds, and

### **Recording devices can process more than their capacity**

Digital video recorders in 2008 typically store between 80 and 160 gigabytes (GB) of recorded video. But consumers erase programs after viewing them, and overwrite the data with more recent programs. So the nominal storage capacity in a DVR is almost irrelevant when measuring how much information is accessed by members of a household.

Similarly, take the example of a home video surveillance system with four cameras. A DVR stores the video streams, and new video over-writes older images. How long it stores the images before overwriting depends on the ratio between the size of the DVR’s hard disk drive and the bit rate of each surveillance camera stream. In turn, the bit rate is determined by the quality of the original video: is it in color or black-and-white? How much video compression is used in storing the feed? A typical medium-quality video stream occupies about one gigabyte per hour of video. So if the DVR has a 160 GB hard drive, the system will hold approximately the most recent 36 hours of video frames.

So how do we classify these data streams? The data stored on the DVR is not yet information, because those bytes are not necessarily used by anyone. In the case

of the surveillance system, only a fraction of the data becomes information, i.e., data delivered for use. This would include any time the homeowner is watching the live feed (rare if ever), or that a recent period is played back for evidence in the event of a burglary. The same size of DVR used in the home to store TV programs (to zip through commercials or simply time-shift viewing) is likely to produce much more information, and as a program is viewed, the same bytes can be written over with future programming. In both DVR examples, we measure only what the consumer sees.

To complicate matters further, in the home video surveillance example when a security camera creates a frame, it actually creates at least four frames of data – the original plus three descendants. Because of compression, the three descendant frames have fewer bytes than the original.<sup>1</sup> If the frame is later recalled and displayed on a monitor, two additional descendants are created. So according to our measurement, the original and most descendants are data; only the final descendant is information.

<sup>1</sup> This explanation oversimplifies issues such as the byte-equivalent of analog images.

particularly from different time periods. Take for example a landmark speech, Abraham Lincoln's Gettysburg Address versus a current TV series, a 2008 episode of "Heroes" on NBC.

The Gettysburg Address took roughly 2.5 minutes to deliver and was 244 words in length, i.e., 1,293 characters, or bytes of text. Nobody is sure exactly what Lincoln said; his handwritten texts do not match contemporary accounts. (On the other hand, a presidential speech today will be recorded electronically for posterity, as they have been since the early Fifties). 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).<sup>8</sup>

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 much more information than the Gettysburg Address offered. 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 this report on information, we measure neither original delivery time nor bytes of storage, but total bytes of all copies across all recipients. "Heroes" episodes in the 2008-09 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 per episode.

Optimistically, let's 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 one billion readings. So measured by the pure number of information consumers, Lincoln's one-billion readership trumps "Heroes" 10 million average weekly viewership by a factor of 100. On the other hand, looking at bytes, a compressed episode of "Heroes" on an average TV comes in at about 500 MB, times 10 million views, which adds up to 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.

Another approach to measuring information then and now is to calculate the amount of time people spend receiving different kinds of information. The Gettysburg Address takes barely 2.5 minutes to read, but more time is spent understanding the

background: the American Civil War, the carnage of the battle, the political importance of Lincoln rallying the North to continue the war, and so forth. So let's call it 20 minutes. Now Lincoln's Address is measured at 40 billion minutes, or 0.7 billion hours. In contrast, a "Heroes" episode is watched only about 14 million hours. So the Address is bigger by a factor of 50.

So which of the two 'information' events is larger in an absolute sense? Perhaps none of our quantitative measures captures this. The pure volume of information does not necessarily determine its value or impact. The right information, delivered at a key time and place, can move mountains. At the other extreme, raw bytes are now so inexpensive that we often pay only minor (or zero) attention to them. So this study eschews efforts to determine the value of one type of information over another, in favor of estimating the volume of information consumption.

## 2 TRADITIONAL INFORMATION IN U.S. HOUSEHOLDS

Information can be roughly classified into "information for consumption," primarily in households and mobile uses, and "information for production" in workplaces and between machines (both of which will be the subject of future *HMI?* reports).

This section discusses traditional information in U.S. households – information delivered and consumed from media that preceded the home computer era. While most of these media are increasingly digital, thanks to the power of modern computing and networking technologies, they remain "traditional" in the sense that the content and the consumption experience are conventional – think of people watching television, speaking on the telephone, reading a book or magazine, or going out on a Friday night to the movies.

### 2.1 Television

Americans are heavy users of TV, and on two of our three measures of information (hours and words), TV is by far the largest source, although it is only second as measured by bytes. However, television usage measured in hours per person is rising only slowly. After all, whether you have 150 channels on digital cable or just a handful of channels of over-the-air broadcast TV, you still have only a limited number of hours to watch TV. The total time has not changed dramatically despite today's broader channel choices and higher-definition TV reception.

The estimated 292 million U.S. viewers average nearly five hours of TV viewing per day.<sup>9</sup> Total TV viewing accounts for 41 percent of total hours of information consumption, and nearly 35 percent of total bytes.

**(Table 3 Television and Radio Consumption)** We receive TV programs over a wide variety of media, including cable, satellite, and plain old broadcast, in descending order. Digital television is compressed for transmission and then uncompressed for viewing, and we measure the compressed bit rate.<sup>10</sup> And if two people are watching the same show on the same TV set, it will show up twice in our measurements, because we use A.C. Nielsen for TV data, and that's how they do their counting.

While HDTV began to take off with consumers in 2008, more homes had HDTV sets (53 percent in January 2009, according to estimates from the Consumer Electronics Association) than actually get HDTV signals (approximately 40 percent, although estimates vary). It is quite common for TV owners to not realize that their "high definition" TV set is actually showing only standard definition TV signals. For those households that do receive HDTV, we don't have good data, but in 2008 roughly 40 percent of their viewing hours were high definition.<sup>11</sup>

Even so-called "high definition" television programs vary considerably in quality. One reason is that original content varies, but another is that cable companies often choose to offer a higher number of channels, with lower bandwidth and lower quality per channel, rather than the reverse. Over the air, cable, and satellite (digital) TV are transmitted at an average of 4 megabits per second, although this depends on what compression methods are used. We estimate that high definition TV averages about 12 megabits per second. Putting all of this together, we use an estimate of 4 megabits per second for standard TV, and 7.2 megabits per second for the weighted average bandwidth of TV into homes that receive HDTV.

Not surprising given TV's still-dominant role in consumer information, it has spawned a variety of special delivery methods beyond cable and satellite (which, in their time, were novel). These include video cassette recorders (VCRs), digital video discs

(DVDs), and home video recorders that use hard drives, called DVRs or PVRs (digital or personal video recorders). While roughly 80 million video cassette recorders (VCRs) remain in U.S. home, their usage is so low that VCRs are no longer broken out as a separate category in home-video playback statistics. Meanwhile, the high quality of DVDs led to their use in essentially all American households. And according to a recent estimate, DVR penetration was 28 percent in late 2008, and 33 percent in 2009.<sup>12</sup> Nielsen lumps most DVR use into its "live" TV hours, but it reports DVD use

**Table 3: Television and Radio Consumption**

ACTIVITY	Total # of Users (millions)	Hours per User/month	Total Info. Exabytes/year	% of Total Hours	% of Total Bytes
<b>TELEVISION AND TV DEVICES</b>					
Television (incl. Delayed View)	292	148.5	1,197	39.22%	32.83%
DVD Players	254	9.3	70	2.21%	1.91%
Mobile TV	10	3.6	0	0.03%	0.00%
<b>Subtotal</b>			<b>1,266</b>	<b>41.47%</b>	<b>34.74%</b>
<b>RADIO</b>					
AM/FM Radio	233	80.6	10	17.62%	0.27%
Satellite Radio	19	65.8	1	1.17%	0.04%
<b>Subtotal</b>			<b>11</b>	<b>18.79%</b>	<b>0.30%</b>
<b>TOTAL</b>			<b>1,277</b>	<b>60.25%</b>	<b>35.04%</b>

separately with an average per user viewership time of 9.2 hours every month. DVDs have a variable bandwidth averaging 5 megabits per second, slightly better than a standard definition TV, and contribute 2.2 percent of INFO<sub>H</sub> hours, and 1.9 percent of INFO<sub>C</sub> in bytes. In comparison, "live" television was 41 percent of hours and 35 percent of INFO<sub>C</sub> in bytes.

DVD usage is expected to increase as the price of Blu-ray high-definition players declines and more people buy HDTV sets on which to watch Blu-ray discs. Ironically, until now the biggest purchasers of Blu-ray discs have been gamers, because Sony built Blu-ray technology into its Playstation 3 game console, making it the most widely used Blu-ray player in the world (This decision probably raised the price of the Playstation 3 significantly, hurting its competition with other game consoles). But in 2008, sales and rentals of Blu-ray disks had little impact on home TV viewing, and they are not included in our DVD viewing estimates for this report.

The leading TV ratings service, A.C. Nielsen, began collecting and reporting on the use of mobile telephones to view video content in 2008, and we



have used their new measurements to calculate the information received over this mobile data service. Just over 10 million U.S. subscribers watched video content on their mobile phones, and they averaged 3.6 hours of video viewing per month. Since both the hours per user and number of users are low relative to the juggernaut of mainstream TV, this works out to only about 0.04 percent of words of information  $INFO_w$ . Because of the small screen size and the scarcity of mobile telephone bandwidth, the bit rates and resolution of these signals can be quite low, less than a quarter of even a conventional TV signal. Therefore, their total impact on bytes is even smaller – we estimate .002 percent of total bytes  $INFO_c$ . On the other hand, iPods and similar devices have the ability to watch TV programs and movies via downloads from the Internet, using a service like iTunes. These can certainly be considered “mobile TV,” but currently Nielsen provides no viewer data for these devices.

## 2.2 Radio

Video never did “kill the radio star” as the British pop group, The Buggles, warned in their 1979 chart-topping single that famously became the first video on MTV when it began broadcasting in 1981. Radio today is thriving on new technology, including HD audio, satellite transmission, online radio and other new services. But in a census of total information consumed in U.S. households, audio requires very low data rates. Even without factoring HDTV into the equation, video requires roughly 30 times more

infancy, but users of services such as Sirius listen to more than 2 hours a day on average (almost as much as listeners to standard AM-FM radio), pushing satellite radio information to more than 1.3 exabytes. Data for Internet radio – a new and small but potentially important segment of the market – are not yet reliable, and it was not included in the radio totals.<sup>13</sup>

## 2.3 Telephone

While most U.S. households had a telephone 25 years ago, today it is common to have at least one landline in the home and more than one mobile phone. There were 263 million wireless users in 2008, versus only 154 million wired lines.<sup>14</sup> On the other hand, on average wired lines are used for almost twice as many minutes per day, so information words  $INFO_w$  are slightly higher for fixed lines. For the sake of accuracy, therefore, this report divides the ‘telephone’ sector into two parts: the traditional or conventional phone usage covered in this section, and wireless phone service that is fast evolving into mobile computing, and therefore is covered in Section 3 below. (**Table 4 Telephone Consumption**)

First though, some comparisons of wired versus wireless voice telephony. It is quite likely that by 2010, the total number of hours that Americans spend on their cell phones will overtake their use of landline phones in the household. As a factor of total hours of information consumed by U.S.

households in 2008, it was already a close race: fixed-line phones accounted for 3.2 percent of total time consuming information, while mobile phones accounted for 2.9 percent. Translated into bytes though, landline calling per person per day remained 12 times greater. This is due to two factors: much more sophisticated compression and lower voice quality of wireless phones.

**Table 4: Telephone Consumption**

ACTIVITY	Total # Subscribers (millions)	Hours per Subscriber/month	Total Info. Exabytes/year	% of Total Hours	% of Total Bytes
<b>VOICE TELEPHONY</b>					
Fixed Line Voice	154	22.5	1.19	3.26%	0.03%
Cellular Voice	263	11.9	0.17	2.94%	0.00%
<b>TOTAL</b>			<b>1.36</b>	<b>6.20%</b>	<b>0.04%</b>

Note 1: Fixed Line Voice includes residential, business and most VoIP subscribers.  
Note 2: Cellular Voice includes residential and business subscribers.

data throughput than audio. Or to compare satellite services, the throughput of satellite TV (1,800 megabytes per hour) compares to only 8 megabytes per hour for satellite radio.

The country’s 233 million AM-FM radio listeners received 10 exabytes of information in 2008. This includes radio in and out of the home (mostly at home on weekends) and in the car during weekday morning and afternoon commutes. Satellite radio, with nearly 19 million listeners, is still in its

Our calculations of information consumed by telephone fixed landline users in 2008 (also known as ‘POTS,’ for ‘plain old telephone service’) are for voice traffic – not DSL nor dial-up Internet service supplied through a wired telephone connection to the home. We calculate that occupants in every U.S. household used their home phone for an average of 22.5 hours each month. Using these numbers, we calculate 1.2 exabytes as the total voice-traffic information consumed by people using landline

telephone service in 2008. Adding mobile voice traffic to the mix, total voice communications created and consumed 1.4 exabytes of information.<sup>15</sup>

## 2.4 Print

The most traditional media consumed in the home are the different flavors of print publications. Taken together, U.S. households in 2008 spent about 5 percent of their information time reading newspapers, magazines and books, which have declined in readership over the last fifty years. From the perspective of the information measured in words  $INFO_W$ , printed media account for almost 9 percent of all words consumed. However, translated into bytes, they barely register: two-hundredths of a percent (0.02%) of  $INFO_C$ . The alphabet is a very compact way to transmit words, and although magazines have color photographs, they are only still images. (Table 5 Conventional Media)

It is expected that readership of print publications will continue to decline, even if newspapers and magazines are able to find a sustainable model for publishing their content on the Web. Our print data do not take into account any Internet editions, which are instead included as computer information in the home (see Section 3). Printed books – on which Americans spent barely 2 percent of their information time  $INFO_H$ , and 4 percent of words – may someday be displaced by digital devices such as the Amazon Kindle, but the electronic book platforms had more potential than actual readers in 2008. Yet, in many ways electronic documents have already taken over for paper – see the sidebar “The Evolution of Reading.”

## 2.5 Movies

Although Americans spend much more time watching movies on television, broadcast and through DVDs, watching movies in a theater remains a popular attraction. No other medium offers anything like the data throughput of a large-screen theatrical projection – roughly 250 million bits per second, which is 20 times the bandwidth of high definition TV. How is this possible, especially since movies are shown at only 24 frames per second, versus 30 for television? Movies have the advantage on the three other determinants of bandwidth: they have larger screen resolutions with more pixels, they have finer color rendition corresponding to more bits per pixel, and they use arguably less compression since they use film and not electronic transmission.<sup>16</sup>

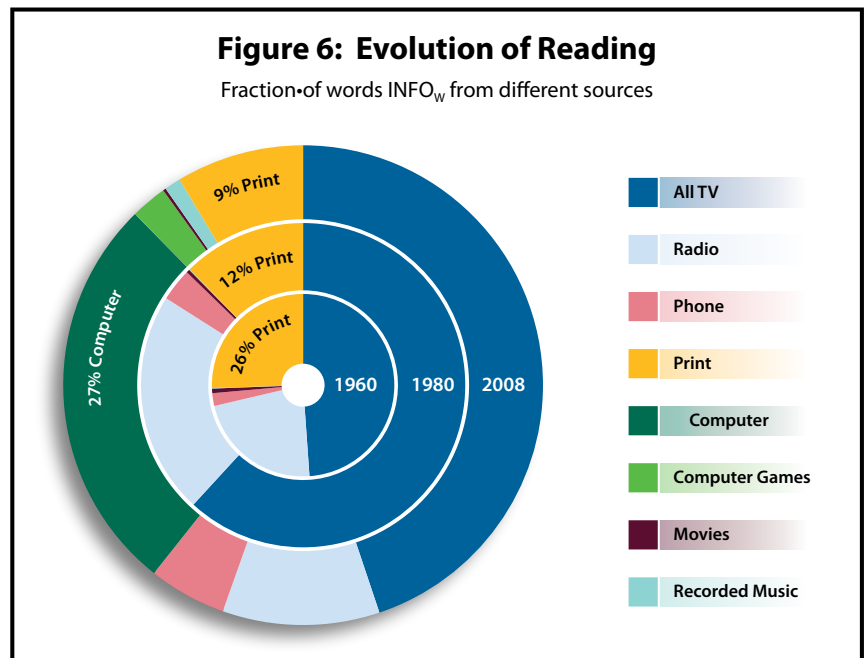
So even though the average American spends less than one hour per month at the movies, it adds up to

## The Evolution of Reading

The use of different media has changed dramatically over time. It is a cliché that reading is in decline. But on the other hand we get considerable information from the Internet, which is a heavy print medium. Do we really read less?

We show this evolution in **Figure 6 Evolution of Reading**.

Conventional print media has fallen from 26 percent of  $INFO_W$  in 1960 to 9 percent in 2008. However, this has been more than counterbalanced by the rise of the Internet and local computer programs, which now provide 27 percent of  $INFO_W$ . Conventional print provides an additional 9 percent. In other words, reading as a percentage of our information consumption has *increased* in the last 50 years, if we use words themselves as the unit of measurement.



3,300 megabytes of information  $INFO_C$  per person per day — a surprising ten percent of the total daily bytes.

Digital projection is gradually coming to movie theaters, but the penetration in 2008 was limited. At present, IMAX technology which is based on 70 mm (analog) film provides the highest quality.

## 2.6 Recorded Music

While the technology for listening to recorded music has changed dramatically, and retail sales of recorded music have declined, the popularity of this medium appears to be intact. We estimate that Americans spend an average of 14 hours per

month listening to recorded music – on CDs and MP3 players. That is nearly 4 percent of all the hours spent consuming information, contributing to a volume of 8.8 exabytes of information. Although the amount of time INFO<sub>H</sub> used for recorded music

digital devices for entertainment, information and other purposes: 3G phones, PDAs, MP3 players, television sets, DVRs, home computers, game devices, and so on.

**Table 5: Conventional Media**

ACTIVITY	Total # of Users (millions)	Hours per User/month	Total Info. Exabytes/year	% of Total Hours	% of Total Bytes
<b>CONVENTIONAL MEDIA: MOVIES, READING, MUSIC</b>					
Movies	295	0.9	356.31	0.25%	9.78%
Books, Newspaper, Magazine	295	32.8	0.67	5.09%	0.02%
Recorded Music	295	13.8	8.85	3.83%	0.24%
<b>TOTAL</b>			<b>365.83</b>	<b>9.2%</b>	<b>10.04%</b>

In this section we report on five major categories of home computer use:

is much lower than radio, the compressed bytes INFO<sub>C</sub> are almost as large, due to higher audio quality of recordings.

### 3 COMPUTER INFORMATION IN U.S. HOUSEHOLDS

New digital technologies continue to remake the American home. Ten years ago 40 percent of U.S. households had a personal computer, and only one-quarter of those had Internet access. Current estimates are that over 70 percent of Americans now own a personal computer with Internet access, and increasingly that access is high-speed via broadband connectivity.<sup>17</sup>

Adding iPhones and other ‘smart’ wireless phones, which are computers in all but name, personal computer ownership increases to more than 80 percent. Many households now boast dozens of

- Accessing the Internet such as web browsing, communications (including email) and social networking;
- Uploading, downloading and watching videos on the Internet;
- Playing computer games;
- Mobile devices and applications; and
- Offline computer activities that don’t require Internet access; such as writing a letter in Word, putting together an Excel spreadsheet, or editing home photos.

The average American spends nearly three hours per day on the computer, not including time at work. That is 24 percent of total information hours, and over 55 percent of all information bytes INFO<sub>C</sub>. We estimate that 2,000 exabytes of information, or 2 zettabytes, were consumed by Americans using home computers, gaming consoles and mobile computing devices in 2008. The vast majority of this information is attributed to computer games, whereas the majority of the time Americans spend on the computer involves the less graphics-intensive but more commonplace Web browsing, email and such.

**Table 6: Computer Use Non-Gaming**

ACTIVITY	Total # of Users (millions)	Hours per User/month	Total Info. Exabytes/year	% of Total Hours	% of Total Bytes
Communications and web browsing	226	65.7	8.01	13.99%	0.22%
Internet video	95	1.8	0.89	0.16%	0.02%
Offline Programs	226	11.1	0.68	2.37%	0.02%
<b>TOTAL</b>			<b>9.58</b>	<b>16.51%</b>	<b>0.26%</b>

#### 3.1 Communicating and Browsing the Internet

The Internet has revolutionized the way Americans communicate. In 1980, email was non-existent in U.S. households, and sending a fax was the hot new way to send messages faster and more cheaply than

via Telex or first-class mail. Today, 220 million Americans spend 14 percent of their information hours  $INFO_H$  on the Internet, almost all of it on applications such as web browsing and email (**Table 6 Computer Use, Non-Gaming**).

In 2008 email remained the most widely used application, accounting for nearly 35 percent of all hours on the Internet. Studies show that the average user can process 30 to 60 emails an hour, involving a sequence of read, respond, assign, delay or delete actions for each message.<sup>18</sup> However, because email is largely text-based, it accounted for relatively few bytes.

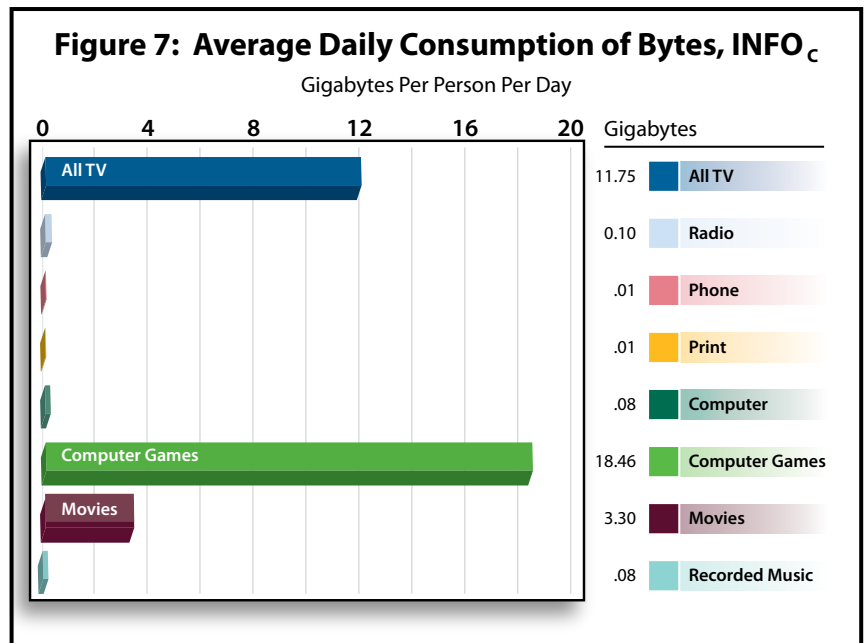
By comparison, Americans spent fewer hours on web browsing (30% of our time on the Internet). Studies show that people cycle quickly through Web sites and doing searches to find content, and they estimate that most users spend only 8-9 seconds looking at most Web pages. Users tend to continue this behavior until they find the page of interest, change their minds, get bored or shift to another task.<sup>19</sup>

Web pages generally include both photos and text, and rapid browsing behavior creates delays as each page is loaded. Internet use continued to evolve rapidly during 2008. Web use was changing due to the rapid uptake of social networks such as Facebook and MySpace. Facebook reported over 175 million users worldwide with an average Facebook user spending 27.5 minutes a day on the site.<sup>20</sup>

For our byte measure  $INFO_C$  we track the amount that actually moves across the “pipe” into the home. This is limited by the average download speed, which varies considerably by technology, by region, by what plan the consumer is signed up for, and even by time of day.<sup>21</sup> However, bandwidth levels are increasing over time as people sign up for higher levels of service, and as Internet service providers strengthen their networks. We assume an average speed of 100k bits per second, which gives an estimated total of 8 exabytes in 2008. All of the text Internet applications combined represent a drop in the bucket when estimating the total number of bytes of information consumed in 2008 by U.S. households – just two-tenths of a percent (0.2%), even though Americans spend 76 percent of their Internet time on email and other text. The reason: Internet video and especially computer gaming involve computer graphics that deliver much higher data throughput to the user’s computer screen.

### 3.2 Internet Video

We measure Internet video, such as YouTube, in its own category. Although there were 95 million viewers in 2008, their average viewing time was



less than 2 hours per month. Hulu and other sites for viewing “regular” television shows may have a big effect in the future, but were used only sparingly in 2008.<sup>22</sup> Furthermore, the resolution of Internet video was very low. Again, the speed of the pipe into the house limits how much can be received while the consumer is actively trying to watch. Although in principle delayed download methods such as peer-to-peer and Apple TV (from iTunes or similar web sites) can increase video download sizes, surveys of consumers don’t yet indicate much use. Furthermore, whatever the pipeline into the home, providing high quality video costs more for the provider, be it YouTube, Hulu, or otherwise, because they must pay for all of the bandwidth used at their end. YouTube only made so-called HD video available late in 2008, and even that has a much lower resolution than high definition television.

As a result, Internet video is still small by most measures. Time consumption  $INFO_H$  was only .2 percent of the total, and bytes  $INFO_C$  were under 1 exabyte, less than text-based Internet use.<sup>23</sup> The higher bandwidth of video compared with web browsing is more than counterbalanced by the smaller number of users (95 million versus 226 million) and much smaller number of average hours per user (1.8 versus 65.7 hours per user per month)

In the future, the small role of Internet video may change. YouTube and other video sites are growing exponentially in both the number of unique visitors to the sites each month, and in the number of videos uploaded and viewed daily.<sup>24</sup> We return to this topic in the conclusion.

**Table 7: Computer Game Playing**

ACTIVITY	Total # of Users (millions)	Hours per User/month	Total Info. Exabytes/year	% of Total Hours	% of Total Bytes
<b>VIDEO AND COMPUTER GAMES</b>					
PC (high performance)	21	86.9	1,405	1.70%	38.56%
PC (standard)	124	18.1	194	2.10%	5.33%
Consoles	89	30.3	368	2.53%	10.09%
Handheld Devices	129	12.6	24	1.53%	0.64%
<b>TOTAL</b>			<b>1,991</b>	<b>7.86%</b>	<b>54.62%</b>

### 3.3 Computer Gaming

While non-game activities account for more of the time Americans spend on computers, computer gaming has come to dominate the total number of information bytes – for a total of nearly 2,000 exabytes (2 zettabytes) in 2008. That is the lion’s share of total bytes from *all* home computing and all sources in general (**Figure 7 Average Daily Consumption of Bytes, INFO<sub>C</sub>**), even though gaming accounted for less than 8 percent of total information hours INFO<sub>H</sub>.

In 2008, an estimated 70 percent of adults in the U.S. played computer games, averaging slightly less than one hour a day. Players were split roughly evenly between men and women (although gender played a role in the differing types of games played). Another estimate in 2009 was that 87 percent of males of all ages, and 80 percent of females, play some form of computer game.<sup>25</sup> Approximately 15 million dedicated gaming

machines (consoles and portables) are sold annually in the US.

It is difficult to talk about computer gaming in aggregate, because there are many different categories of gaming and each type is associated with different percentages of players as well as hours and bytes consumed. Our headcounts and estimated hours of play are from a 2008 industry report on computer

gaming, which described seven types of gamers, ranging from “extreme gamers” (3% of the gaming population) to casual gamers (20%).<sup>26</sup> Many gamers play on more than one type of machine, which is not surprising since almost everyone owns a cellphone and most cellphones in 2008 had the capability to play at least simple games.

Hardware is the critical factor in determining the volume of information generated by videogames and computer games. We report hardware in four categories:

- High performance gaming computers, which were used by 21 million players in 2008;
- Standard computers – 124 million users;
- Console game machines, such as Microsoft’s Xbox, Sony’s Playstation and Nintendo’s Wii – 89 million users; and
- Portable game machines, including the Sony PSP, Nintendo DS, and others – 129 million users.

**Figure 8: Example of a Graphics Processing Card**

For each hardware type, we estimated the video throughput for an “average machine” in the class, playing an “average game.” High-performance gaming PCs use the most powerful processors in the world, called Graphics Processing Units (GPUs), to generate graphics. Some GPUs have over one billion transistors, and more than 200 parallel processors running at once. (**Figure 8 Example of a Graphics Processing Card**) We estimate the effective compressed bandwidth of these machines at approximately 100 megabits per second – eight times that of high definition TV. An estimated 21 million users spend an average of 87 hours every month playing games on these computers. (**Table 7 Computer Game Playing**) They account for a huge share of all information bytes consumed by U.S. households: 1,400 exabytes (1.4 zettabytes) annually – or approximately 39 percent of all INFO<sub>C</sub>. This large role of high-end computer gaming is particularly surprising, because

it accounts for less than 2 percent of the hours Americans spend consuming information. The quality of visual effects on high-end machines and the rapidity with which the player is confronted with changing scenes on the screen are why these devices and games represent such a huge portion of total information to U.S. households, as well as why the games are so immersive to play. **Figure 9 Screen Shot from *NBA Live 10*** shows a screen shot from the Playstation 3 version of a recent computer game. (This resolution is considerably below the best possible from computers in 2008.)

By contrast, six times more Americans play games on *standard* PCs than on high-end PCs, with an average of 18 hours a month. The quality of their on-screen graphics on these PCs is on average far inferior, so this translates into 194 exabytes, barely 10 percent of the total amount of the INFO<sub>C</sub> from games.<sup>27</sup> Nearly 90 million Americans play games on dedicated game machines, and the average player uses a console 30 hours a month.<sup>28</sup> By our calculations, game consoles account for 368 exabytes – 10 percent of total household information. Most of the consoles are used offline, but increasingly, users are playing over the network as well, so the line that divides online and off-the-Internet gaming is rapidly fading. Even handheld game devices, used by 129 million players, created 24 exabytes of information in 2008, or about triple the total bytes of information received in the form of recorded music (primarily CDs and MP3s).

Looking at all the game platforms, we calculate that total information from this relatively young form of entertainment (2 zettabytes) is *50 percent* larger than the volume of information from established media that are more than 50 years old, TV and radio (1.3 zettabytes). (Of these 2 zettabytes, 70 percent is from 21 million high-end gaming computers.) In the short run, TV's share of bytes may increase as the percentage of U.S. households with HD television reception grows at a rapid pace. But manufacturers of high-end gaming computers and consoles are already working on even more powerful new machines and photorealistic games – so in the long run, gaming is likely to continue accounting for the bulk of information consumed by U.S. households, as measured by INFO<sub>C</sub>. On the other hand, measured by words and hours, computer games are a modest 2.5 percent and 8 percent of total consumption, respectively.

### 3.4 Off-Internet Home Computer Use

In many households, considerable computer time is spent locally, without going online except perhaps to send an email. After all, fewer than 60 percent

of adult Americans had broadband connections in 2008.<sup>29</sup> Off-line use includes activities like updating a resume, editing photos, or running a household finance program. Time-use statistics for such off-Internet, non-gaming computer use are no longer reported directly by U.S. government or industry sources. We relied on partial data provided by the American Time Use Study (ATUS) conducted by the Bureau of Labor Statistics (BLS), and time-of-use studies published by the Center for Research in Information Technology and Organization (CRITO) at the University of California, Irvine.<sup>30</sup>

We estimate that non-Internet, non-gaming home computer use was very widespread, but averaged only 17 minutes per day per average American. Because these applications are primarily text based, they add up to only 0.7 exabytes per year.

### 3.5 Smart Phones

The growth of new media, viewing video, sending text messages, or playing games on a feature phone or smartphone are growing quickly, driven by consumer sales of new phones and the provision of new content services, both free and subscriber based.<sup>31</sup> The contributions of these new devices and their use, however, do not figure prominently in our information totals – their numbers are still too low to be a significant fraction of the total information consumed by Americans when compared to the information volume delivered by traditional media.<sup>32</sup>

Approximately 263 million Americans carry cell phones, and in 2008 approximately 50 million of these phones were smartphones such as the Apple iPhone.<sup>33</sup> With first-generation analog cell phones and 2G digital handsets, consumers were largely limited to using their phones for voice calling (cellular phone voice traffic was discussed in Section 2, 2.3 on telephones). So while voice communication accounted for over two-thirds of cell phone hours in 2008, the spoken word is such an efficient medium for conveying information that voice traffic, measured in bytes, accounted for only 0.2 exabytes, a negligible fraction of total INFO<sub>C</sub>.

While Americans spent approximately 7 billion hours text messaging in 2008, because SMS text messages are so small, their byte total is insignificant.<sup>34</sup> For now, new media volumes are small compared with traditional media volumes, but this is changing.

Figure 9 Screen Shot from NBA Live 10



## 4 Trends, Perspectives and the Future of U.S. Information Consumption

Our analysis, while incomplete, has uncovered a variety of trends and patterns, and also some paradoxes.

### 4.1 Analyzing the Growth of Information

While at one level, the estimated five-fold increase from 1980 to 2008 in  $INFO_C$  bytes consumed is impressive, this is an annual growth rate of only 5.4 percent. This is far less than the rate of increase in most measures of digital technology, which tend to be driven by Moore's Law: the number of transistors on an integrated circuit doubles approximately every two years. For example, the cost of hard disk storage in an 1982 personal computer was about \$50 per megabyte for a 10 megabyte drive; today it was less than \$1 per gigabyte for a drive of 100 gigabytes, a 50,000-fold improvement.<sup>35</sup> William Nordhaus studied long-term trends in the cost of computation, and found that it fell faster than 60 percent *per year* from the mid 1980s to 2006. The total reduction was five orders of magnitude, a cost/performance improvement by a factor of 200,000.<sup>36</sup>

If the total revenue of an industry is constant, then the quantity of its output, measured in terms of total performance, must grow in inverse proportion to its price/performance ratio. And 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 percent (the lower end of Moore's Law estimates) and 60 percent per year. How is it possible that *consumption*  $INFO_C$  grew at only 5.4 percent per year, less than twice as fast as growth in GDP 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 percent per year, from 226 million to 295 million (ages 2 and up)
- Average hours of information consumption per person grew at 1.7 percent per year, from 7.4 hours to 11.8 hours of  $INFO_H$
- Average bandwidth (across all media) grew at 2.8 percent 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 percent, from 9.8 to 33.8 Gigabytes of  $INFO_C$ . Not coincidentally, 4.4 percent 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 US population is well known, and the 1.6 percent per year growth in hours of consumption per person is understandable given the constant 24 hour length of a day. But the 2.8 percent compound annual growth rate in bytes consumed per hour remains a drop in the bucket compared to the doubling every two years in the 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 reason for this result: very slow growth of  $INFO_C$  from television. The dominant source of bytes in 1980 – color television – 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. Intel and the rest of the semiconductor industry sell more devices, and more transistors per device, every year, and America's



share of worldwide consumption has been roughly constant. So if these transistors 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 depending on what measure is used so does the Internet. Will their growth raise the historical trajectory for the future?

## 4.2 Where are the Missing Bytes?

We have tentatively identified four places where the missing bytes have gone, although further research

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.

Examples of dark data occur in the home, although most of it is elsewhere. Data can be created in an automated fashion without the consumer intervening. For example, a consumer can set a DVR just by specifying the name of the program, not when it is broadcast. Information is exchanged over the Internet between the cable company's computer and the DVR, and the DVR decides when to record, and what channel. We recognize the

results of the dark data when we turn on the DVR and it is converted to information on our TV screen.

The family auto (or automobiles) is a more typical example of dark data. Luxury and high-performance cars today carry more than 100 microcontrollers and several hundred sensors, with update rates ranging from one to more than 1,000 readings per second.

One estimate is that from 35 to 40 percent of a car's sticker price goes to pay for software and electronics.<sup>38</sup> As microprocessors and sensors 'talk' to each other, their ability to process information becomes critical for auto safety. For example, airbags use accelerometers, which measure the physical motion of a tiny silicon beam. From that motion, the car's acceleration is calculated,<sup>39</sup> and approximately 100 times each second, this data is sent to a microprocessor, which uses the last few seconds of measurements to decide whether and at what intensity to inflate the airbag in the event of a collision. Over the life of an auto, each accelerometer will produce more than one billion measurements. Yet in a crash, only the last few data points are critical.<sup>40</sup> Each sensor creates several gigabytes of data without a single byte that counted as "information" in our analysis of consumer information.

The phenomenon of dark data permeates modern digital technology, and goes far beyond the range of this report. We hope to analyze it carefully in the future.

**Table 8: 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

will be needed to confirm and measure them. (**Table 8 Explaining the Gap between Consumption and Capacity Growth**) First, we have measured information *consumed* by consumers, but the amount of information *available* to them has grown much faster.<sup>37</sup>

Second, this report looks only at consumer information. We are working on a study of information in enterprises, which follows different growth patterns. 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.

### 4.2.1 Dark Data

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 one 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

### 4.2.2 Two Kinds of Quality: Variety and 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 about 130, of which the average household actually watches 18.<sup>41</sup> 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.

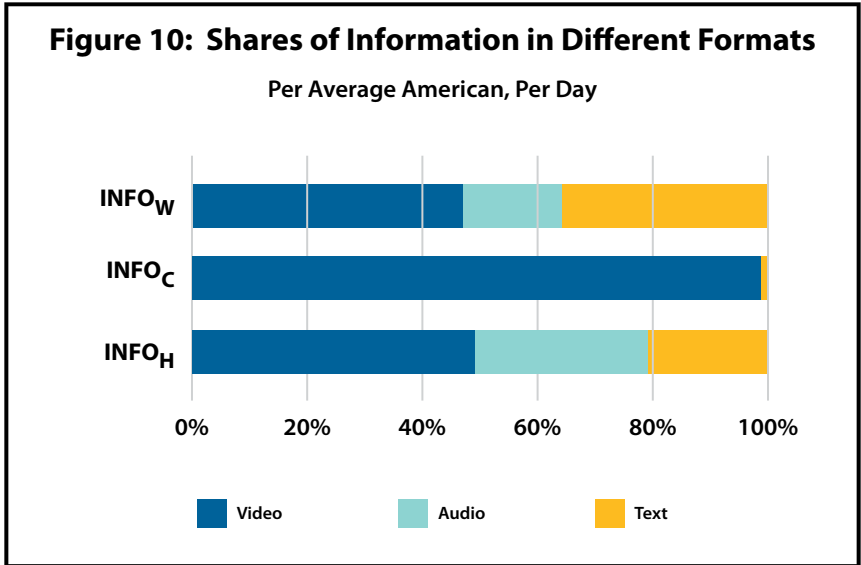
The additional channels of TV, however, have come at a cost: higher compression and therefore lower video resolution for the channels we receive. The issue is straightforward: bandwidth costs money (all those transistors). For a fixed budget, a cable TV company, and especially a satellite TV company,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

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 1080i that many TV sets are now capable of receiving.<sup>42</sup> In fact, channels advertised as “HDTV” are sometimes so compressed that the pictures are far below the theoretical capability of the TV set.<sup>43</sup> The same issue comes up for broadcast stations, which are each given the use of 16 Mbps of bandwidth, and typically divide it into two or three different 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.

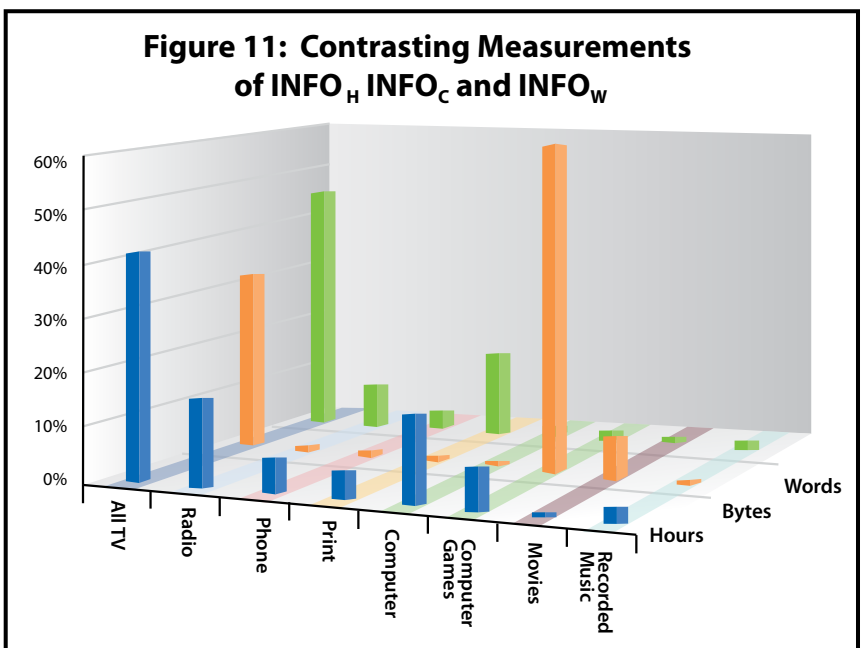
### 4.3 Analyzing Information Consumption

We have discussed each medium of information in turn, using three different measures (hours, compressed bytes, and words), and a range of reference points including percentages, yearly totals, and daily consumption. Appendix B provides



much of the underlying detail, from which the summary numbers were drawn. (However, it does not include details of calculations for the more complex topics, such as computer games.)

As **Figure 10 Shares of Information in Different Formats** illustrates,  $INFO_C$  bytes are completely dominated by video sources: movies, TV, and computer games. Consumption time,  $INFO_H$  on the



**Table 9: Summary of Information for Major Groups**

ACTIVITY	Total per year (entire population)			% of Total			Per average American per day		
	INFO <sub>h</sub> in hours	INFO <sub>c</sub> in bytes	INFO <sub>w</sub> in words	% Hrs	% Bytes	% words	Hours	Giga- bytes	Words
All TV	5.30E+11	1.27E+21	4.86E+15	41.62%	34.77%	44.85%	4.91	11.75	45,100
Radio	2.39E+11	1.10E+19	1.15E+15	18.79%	0.30%	10.59%	2.22	0.10	10,645
Phone	7.89E+10	1.36E+18	5.68E+14	6.20%	0.04%	5.24%	0.73	0.01	5,269
Print	6.49E+10	6.72E+17	9.34E+14	5.09%	0.02%	8.61%	0.60	0.01	8,659
Computer	2.08E+11	8.69E+18	2.93E+15	16.35%	0.24%	26.97%	1.93	0.08	27,122
Computer Games	1.00E+11	1.99E+21	2.65E+14	7.86%	54.62%	2.44%	0.93	18.46	2,459
Movies	3.24E+09	3.56E+20	2.14E+13	0.25%	9.78%	0.20%	0.03	3.30	198
Recorded music	4.88E+10	8.85E+18	1.20E+14	3.83%	0.24%	1.11%	0.45	0.08	1,112
<b>TOTALS</b>	<b>1.27E+12</b>	<b>3.64E+21</b>	<b>1.08E+16</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>	<b>11.80</b>	<b>33.80</b>	<b>100,564</b>

5.3E+11 means  $5.3 \times 10^{11} = 530,000,000,000$

other hand, 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 11 Contrasting Measurements** shows in more detail how different media dominate each measure of information. Only television is a large contributor to all three measures.

**Table 9 Summary of Information for Major Groups** aggregates the information in Appendix B by major categories, such as television and print.

### 4.3.1 How Much Information is Delivered via the Internet?

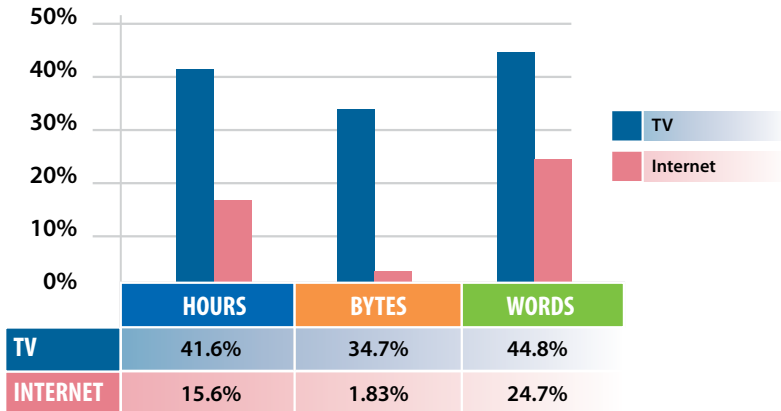
Another question we investigated is the quantitative importance of the Internet: how much does it contribute to our information consumption?

Our basic finding is that the Internet provides a substantial portion of some kinds of information, but very little of others. Measuring with hours or words, the Internet provided a significant fraction of our information, although less than television. (**Figure 12 Internet as a Source of Information**) We spent 16 percent of our information hours using the Internet (versus 41 percent for TV), and receive 25 percent of our words INFO<sub>w</sub> from it (versus 45 percent from TV). The Internet was the source of only 2 percent of our INFO<sub>c</sub> bytes (versus 35 percent for TV).

Yet surveys show that many of us view the Internet as very important, to the extent that we will cut spending on cable TV before we cut Internet access. How can this importance be reconciled with its smaller quantitative measurements? Our analysis explains why the unique properties of the Internet make it considerably more useful per byte or word of information for certain purposes.

We classify our information consumption into three mutually exclusive purposes: two-way communication, entertainment, and research/current events. Two-way communication is self explanatory. Before the Internet, the only ways to have a two-way exchange without being in the same room were telephone and first-class letters. The

**Figure 12: Internet as a Source of Information**



Internet adds multiple additional methods, including email, social networking, and instant messaging. We estimate that Americans averaged 1.6 hours per day conducting two-way communication, of which 57 percent was via the Internet, with the rest of the time on cellular or landline telephones. Correspondingly, the Internet provides 79 percent of the bytes and 73 percent 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.

The other uses we classified information into were entertainment and research/current events, by which we mean gathering factual information of any kind – basically any non-fiction information, to distinguish it from entertainment. We calculate that Americans average 6.5 hours per day on entertainment and 3.7 hours on research/current events.

The Internet's contribution to pure entertainment information is very small: less than 2 percent, whether measured by hours, bytes, or words. The reasons stem 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 can deliver only a fraction of that or less at peak times. Broadband providers in many areas do offer premium-priced service levels, but the speed is not sufficient for live TV, for several reasons. Even when the "last mile" to a house is capable of adequate speeds, this is based on statistical multiplexing, meaning that it assumes that only a fraction of users will be operating at this speed at the same time. If everyone turned on their "Internet TV" at 7pm, many parts of the network would be unable to handle the load. On the other hand, video on the Internet is growing rapidly. The popularity of video download sites indicates that demand exists, even with lower visual quality than standard television.

In our third and final use category, research and current events, the Internet provides 23 percent of our hours and 31 percent 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 or 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, the Internet provides only 1.3 percent of our research/current event bytes.

### 4.3.2 The Rise of Interaction

Most sources of information in the past were consumed passively. Listening to music on the radio, for example, does not require any interaction beyond selecting a channel, nor any attention thereafter. Telephone calls were the only interactive form of information, and they are only 5 percent of words and a negligible fraction of bytes. However, the arrival of home computers has dramatically changed this as computer games are highly interactive. Most home computer programs (such as writing or working with user generated content) are as well. Arguably, web use is also highly interactive, with multiple decisions each minute about what to click on next.

As a result, we estimate that a full third of our  $INFO_w$  in words is now received interactively, and 55 percent of our  $INFO_c$  bytes. This is an overwhelming transformation, and it is not surprising if it causes some 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 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 be 135 gigabytes of  $INFO_c$ , about four times the average daily consumption today.

### 4.4 The Future of Consumer Information

There are some patterns of information consumption in the first half-decade of the twenty-first 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. On the other hand, the number of TV channels has grown steadily for 50 years, and actual video quality has not grown nearly as fast as a simplistic theory of technological progress (Moore's Law) seemingly predicted.

Two nascent developments might 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 as a complement rather than a substitute for conventional TV program delivery mechanisms. 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 US, 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, and 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 may 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 exabytes. 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.

## Appendix A: UC Berkeley HMI? Studies

*How Much Information? 2009* follows two University of California, Berkeley research reports, HMI? 2000 and HMI? 2003, conducted by Professors Peter Lyman and Hal Varian. HMI? 2009 builds on the two Berkeley studies, but there are important differences. First, Lyman and Varian report on World and U.S. information totals for calendar years 1999 and 2002. HMI? 2009 reports only on the U.S. for calendar year 2008. Second, the two studies measure information differently and use different methods to count it. Lyman and Varian measured “original” information - that is, the first instance of new information being created, such as a voice telephone call, or someone composing an email. They analyzed the quantity of original content created – how many hours of radio broadcasting were produced worldwide, how many books were published, and so on. HMI? 2009 defined information as flows of data delivered to people. We measured the amount of information delivered to people for consumption. Our contrasting definitions led to differences in calculating information totals, and later we work through two examples to illustrate the importance of these differences. Third, Lyman and Varian divided total information into two measures and reported them separately - the first, the annual size of the “stock” of new information contained in storage media; the second, the volume of information seen or heard each year in information flows. We measured information consumption as the number of hours information was received by people (INFO<sub>H</sub>), the number of bytes delivered (INFO<sub>C</sub>), and the number of words consumed (INFO<sub>W</sub>). We reported annual totals for each measure.

We also consulted industry sources, including two reports on digital information growth completed by the International Data Corporation (IDC) published in 2007 and 2008.

### HMI? 2000

The first UC Berkeley report estimated that in 1999 the world produced between 2 and 3 exabytes of new information, or roughly 500 megabytes for every man, woman, and child (we define an exabyte of information elsewhere in this report).<sup>44</sup>

Lyman and Varian identified three key conclusions in summarizing their 2000 report:

- First, they referred to the “paucity of print.” Printed materials of all kinds made up less than .003 percent of the total amount of annual information produced in the world. They cautioned that this number did not mean print was insignificant. On the contrary, they noted it simply meant the written word was a very efficient way to convey information.
- Second, they referred to a growing “democratization of data” – the fact that a vast amount of new information is created and stored by individuals. For example, original documents created by office workers were more than 80% of all original paper documents (the other 20% included original copies of newspapers, books, magazines, and other print material). And photographs taken by consumers and X-rays together were 99% of all original film documents.
- Third, they noted the increasing “dominance of digital” content. Not only was digital information production the largest in total, it was also the most rapidly growing. They concluded that while unique content produced on print and film was hardly growing at all, magnetic storage was by far the largest medium for storing information and was the most rapidly growing medium, with shipped hard drive capacity doubling every year.

### HMI? 2003

In 2003, Lyman and Varian extended their earlier study. They added a new section on the Internet, sampling the World Wide Web to estimate the size of the surface web and to determine the source and content of Web pages. And they added an analysis of desktop disk drives, to determine how people consumed information received on the Internet. They concluded:

- Print, film, magnetic, and optical storage media produced about 5 exabytes of new information worldwide in 2002. Ninety-two percent of the new information was stored on magnetic media, mostly on hard disks.
- Information flowing through electronic channels – telephone, radio, TV and the Internet – contained almost 18 exabytes of new information worldwide in 2002, three and a half times more than was recorded on storage media. Ninety eight percent of this total was the information sent and received in telephone calls – including both voice and data on both fixed line and wireless phones.
- They estimated that the total amount of new information stored annually on paper, film, magnetic, and optical media worldwide had doubled in the last three years.

Lyman and Varian drew a number of implications from their 2003 study. Perhaps most important, they noted that our ability to store and communicate information was far outpacing our ability to search, retrieve and present it.

## Comparing HMI? 2009 with HMI? 2003 and 2000

As noted, our contrasting definitions and measures produce different annual information totals, and these totals are not directly comparable. For example, in Television and Radio we calculated total annual information in the U.S. (INFOc) was 1,277 exabytes per year. Lyman and Varian's total for the U.S. was seven one-hundreds of an exabyte. Why? They counted a television (or radio) program once, the first time it was aired. We counted every time a television viewer watches a program, which could be 20 million people.

Here is how each respective total was calculated: Berkeley estimated that in 2002 there were approximately 3.6 million hours of original information broadcast by U.S. television stations, and 19.8 million hours of original information broadcast by U.S. radio stations (reported in their Table 1.11). Using a conversion factor of 1.3 GB to 2.25 GB per hour for television, and 0.05 GB per hour for radio, they calculated total U.S. Television and Radio information was between a lower bound estimate of 5,718 terabytes and an upper bound estimate of 9,175 terabytes, of new information in 2002. In HMI? 2009, we counted the number of television viewers (292 million people), the amount of time they view television (on average 148.5 hours a month), and calculated 1,197 exabytes of data was delivered to their television screens that year. Adding in DVD players and Radio brought the total to 1,277 exabytes (Table 3 Television and Radio Consumption). We also contrast Telephone information, where a more thorough explanation is necessary, in an extended endnote.<sup>45</sup>

## International Data Corporation (IDC) 2007 and 2008

International Data Corporation (IDC) published two reports on the growth of digital data in 2007 and 2008. IDC's definition of digital information and their methods for counting it were not explained in sufficient detail to reliably compare their totals with the HMI? 2000 and 2003 reports, or HMI? 2009. IDC's numbers for the entire world were 12 times less than our 2009 numbers for U.S. households alone. But it is not clear whether the large discrepancy was due to our including more types of information sources (such as non-Internet computer use and game consoles), our inclusion of analog as well as digital sources, our different approach to measuring bytes, or for other reasons.

The main conclusions of IDC's 2008 report include:

- The amount of digital data created in 2007 was 281 billion gigabytes (281 exabytes), equivalent to 45 gigabytes per capita, roughly the size of a Blu-Ray disc. (The maximum capacity of the new Blu-ray HD format is 50 GB on a dual-layer disc.)
- Digital data was projected to grow at a compound annual growth rate of almost 60%, reaching 1.8 zettabytes (1,800 exabytes) by 2011.
- More than 80 percent of bytes are images: pictures, surveillance videos, TV streams, and so forth.
- Individuals' "Digital Shadows" – information generated as a result of activities such as web surfing and shopping, but not by them directly – surpasses the amount of digital information individuals create themselves.

## SOURCES

Peter Lyman and Hal R. Varian, How Much Information, 2000. <http://www2.sims.berkeley.edu/research/projects/how-much-info/>

Peter Lyman and Hal R. Varian, How Much Information, 2003. <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>

John F. Gantz, et. al., The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011, IDC (March 2008).

## Appendix B: Detail Table

ACTIVITY	Users	Throughput		Total per year (entire population)			Per User / Per Day			Per average American / per day			% of Total		
	# of Users (millions)	bits per sec. (bps) - comp.	Words per minute	Hours (billion) INFO <sub>H</sub>	Exabytes INFO <sub>C</sub>	Words (trillion) INFO <sub>W</sub>	Hours	Megabytes	Words	Hours	Gigabytes	Words	% Hrs	% Bytes	% words
Cable TV - SD	95.7	4,000,000	153	163.0	293.0	1,493	4.66	8,380.0	42,740	1.51	2.71	13,843	12.8%	8.0%	13.8%
Cable TV - HD*	69.3	7,200,000	153	118.0	382.0	1,081	4.66	15,085.0	42,740	1.09	3.54	10,024	9.3%	10.5%	10.0%
Over air TV - SD	27.8	4,000,000	153	47.0	85.0	434	4.66	8,380.0	42,740	0.44	0.79	4,027	3.7%	2.3%	4.0%
Over air TV - HD *	20.2	7,200,000	153	34.0	111.0	314	4.66	15,085.0	42,740	0.32	1.03	2,916	2.7%	3.0%	2.9%
Satellite - SD	45.5	4,000,000	153	77.0	139.0	710	4.66	8,380.0	42,740	0.72	1.29	6,586	6.1%	3.8%	6.5%
Satellite - HD*	33.0	7,200,000	153	56.0	182.0	514	4.66	15,085.0	42,740	0.52	1.68	4,769	4.4%	5.0%	4.7%
DVD	253.8	5,500,000	153	28.0	70.0	258	0.30	751.0	2,787	0.26	0.65	2,394	2.2%	1.9%	2.4%
Other TV (delayed view)	50.0	3,000,000	153	3.9	5.3	36	0.21	289.0	1,966	0.036	0.05	333	0.31%	0.14%	0.33%
Mobile video	10.3	300,000	153	0.4	0.1	4.1	0.12	16.0	1,089	0.004	0.00	38	0.03%	0.002%	0.04%
Internet video	94.7	1,000,000	153	2.0	0.9	18	0.06	26.0	527	0.018	0.01	169	0.16%	0.024%	0.17%
Newspapers	51.2	18,235	240	9.0	0.4	124	0.46	3.8	6,628	0.080	0.00	1,149	0.68%	0.011%	1.14%
Magazines	250.0	18,000	240	29.0	0.2	421	0.32	2.6	4,616	0.27	0.00	3,906	2.3%	0.007%	3.9%
Books	250.0	1,330	240	27.0	0.0	389	0.30	0.2	4,261	0.25	0.00	3,605	2.1%	0.000%	3.6%
Satellite Radio	18.9	192,000	80	15.0	1.3	71	2.16	186.0	10,354	0.14	0.01	662	1.2%	0.035%	0.66%
AM & FM Radio	232.5	96,000	80	224.0	10.0	1,077	2.64	114.0	12,686	2.08	0.09	9,982	17.6%	0.27%	9.9%
Conventional Telephone (POTS)	154.0	64,000	120	41.0	1.2	299	0.74	21.0	5,311	0.38	0.01	2,768	3.3%	0.033%	2.8%
Cellular Voice	263.0	10,000	120	37.0	0.2	270	0.39	1.8	2,809	0.35	0.00	2,501	2.9%	0.005%	2.5%
High-end Computer gaming**	20.8	Varies	50	22.0	1,405.0	65	2.85	185,100.0	8,548	0.20	13.03	602	1.7%	38.6%	0.60%
Computer gaming**	123.7	Varies	50	27.0	194.0	80	0.59	4,299.0	1,777	0.25	1.80	744	2.1%	5.3%	0.74%
Console gaming**	88.8	Varies	50	32.0	368.0	97	0.99	11,349.0	2,980	0.30	3.41	896	2.5%	10.1%	0.89%
Handheld gaming**	128.9	Varies	20	20.0	24.0	23	0.41	500.0	497	0.18	0.22	217	1.5%	0.64%	0.22%
Internet text (email, web, etc.)	226.3	100,000	240	178.0	8.0	2,564	2.16	97.0	31,032	1.65	0.07	23,771	14.0%	0.22%	23.60%
Offline programs	226.3	50,000	200	30.0	0.7	361	0.36	8.0	4,375	0.28	0.01	3,352	2.4%	0.019%	3.3%
Movies	295.5	244,737,638	110	3.2	356.0	21	0.03	3,304.0	198	0.03	3.30	198	0.25%	9.8%	0.20%
Recorded Music inc. MP3	295.5	403,200	41	49.0	9.0	120	0.45	82.0	1,112	0.45	0.08	1,112	3.8%	0.24%	1.11%
<b>MASTER SUM</b>				<b>1,273</b>	<b>3,645</b>	<b>10,845</b>				<b>11.80</b>	<b>33.80</b>	<b>100,564</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>

\* HD numbers are a blend of High Definition and Standard Definition use in HD households.

\*\*Computer gaming users and bandwidths are averages from more detailed calculations.

All our numbers are estimates - see the on-line appendix and the endnotes for more information about data sources and methods.

<[http://hmi.ucsd.edu/howmuchinfo\\_research.php](http://hmi.ucsd.edu/howmuchinfo_research.php)>



## Endnotes

<sup>1</sup> A 40-hour per week job is 22 percent of a year. Slightly less than half of the US population is employed. Therefore an “average person” is at work 2.7 hours per day. Source: Bureau of Labor Statistics 2008. < <http://www.bls.gov/news.release/empsit.nr0.htm>>.

<sup>2</sup> HMI? 2009 draws on an unusually large number of data sources from university research, government and industry. Reconciling the many differences in definitions, sample populations and measurement approaches has been a major preoccupation of the research team, especially where sample populations may vary in age or other demographic characteristics, or where double-counting could take place in cases where multiple measurements have been taken of the same population. We have done the best we can in isolating such cases and accounting for them. We have also consulted other large-scale media studies facing the same methodological challenges, for example, the Video Consumer Mapping (VCM) Study conducted by the Council for Research Excellence and Ball State University’s Center for Media Design (CMD). < [http://www.researchexcellence.com/news/032609\\_vcm.php](http://www.researchexcellence.com/news/032609_vcm.php)>.

<sup>3</sup> Teenage viewing is analyzed in Nielsen, How Teens Use Media: A Nielsen report on the myths and realities of teen media trends, June 2009. Statistics for various age groups are from The Council for Research Excellence, Video Consumer Mapping Study: Appendix - Additional Findings & Presentation Materials, June 2009.

<sup>4</sup> In 1960, transistors were used only in a few applications, including some computers and a new kind of consumer electronics, “portable radios.” Integrated circuits were not even invented until later in the decade.

<sup>5</sup> We have adjusted Pool’s numbers for some differences in assumptions.

<sup>6</sup> Analog integrated circuits are also very important, but even devices with analog circuitry such as radios generally are controlled by digital processors.

<sup>7</sup> Standard Definition TV (SDTV).

<sup>8</sup> Lincoln’s salary at the time was \$25,000 per year, or about \$8 an hour. The salary today is \$400,000, or about \$200 an hour. < <http://www.lib.umich.edu/govdocs/fedprssal.html>>.

<sup>9</sup> Nielsen, A2/M2 Three Screen Report, January 2009. Based on data collected in 4Q 2008, Nielsen reported U.S. viewers watched an average of 151 hours per month. This number probably has some seasonality in it.

<sup>10</sup> Over the air analog (NTSC) television is not compressed. NTSC is an analog color TV standard developed in the U.S. in 1953 by the National Television System Committee. Television signals that are compressed and then uncompressed for viewing are MPEG-2 or higher. MPEG-2 is a standard for the coding of moving pictures and associated audio information. It describes a combination of lossy video compression and lossy audio data compression methods.

<sup>11</sup> Further adding to the confusion, in 2009 all US

broadcasters shifted from analog to digital broadcasting. Some cable companies and most satellite broadcasters made the shift years before, but there are still some cable signals that are analog. In any case, digital TV signals can have a number of different resolutions, so whether a show is high definition does not depend on whether it is broadcast in digital or analog.

<sup>12</sup> Bill Carter, “DVR, Once TV’s Mortal Foe, Helps Ratings,” New York Times 1 November 2009.

<sup>13</sup> Our source for radio data is Arbitron Inc. We reviewed Arbitron’s Radio Today: How America Listens to Radio, 2007, 2008 and 2009 Editions, and Arbitron Radio Listening Report, The Infinite Dial 2008: Radio’s Digital Platforms Online, Satellite, HD Radio and Podcasting. Arbitron reports AQH (average quarter hour) listenership by location. For the most popular radio formats, for example News/Talk/Information, at work listener share was 12.8 percent in 2008, and for other popular formats, averages 20 percent or less. We did not deflate average listening hours by format by location in our estimates.

<sup>14</sup> Our cellular and fixed line telephone numbers include both residential and business lines. Additionally, our fixed line number also includes most Voice over IP (supplied by the cable TV companies). Voice over IP (also referred to as VoIP, IP telephony, and Internet telephony) refers to technology that enables routing of voice conversations over the Internet or a computer network. Sources: CTIA-The Wireless Association; Federal Communications Commission, Local Telephone Competition: Status as of December 31, 2007, Industry Analysis & Technology Division, Wireline Competition Bureau, September 2008; Federal Communications Commission, Local Telephone Competition: Status as of June 30, 2008, Industry Analysis & Technology Division, Wireline Competition Bureau, July 2009.

<sup>15</sup> The 154 million figure includes residential and business lines and most VoIP (which is supplied by the cable TV companies). What it does not include is “over the top” VoIP like Vonage and Skype. Our telephone numbers, therefore, may overstate consumer information by approximately 30 percent. Estimates differ as to the number of VoIP subscriber lines in 2008. By the end of 2008, the top 10 ISPs (Internet Service Providers) had approximately 19.6 million residential customers in the US. If we use fixed line usage as an approximation for VoIP usage, VoIP subscribers would have spent approximately 5.3 billion hours making VoIP telephone calls in 2008. Our estimate was calculated by adding up the total number of VoIP customers listed in annual reports and in SEC disclosures by the top ten ISPs providing VoIP services in 2008. We have not included these calculations in our voice telephony information totals. We also do not include international calls. Sources: Customer data obtained from SEC 10-K and 10-Q disclosures and Annual Reports for Comcast, Time Warner, Vonage, Cox, CableVision, Charter, Insight Communications, Mediacom, SureWest and CBeyond.

Industry sources included VOIP-News.com, ISP-Planet, BusinessWire and information services companies including Pike and Fischer, Nielsen, TeleGeography, iLocus and In-Stat.

<sup>16</sup> There is no easy way to rate the “bits per second” of film. For example, film resolution is measured in a different way than video – lines per inch, rather than pixels. And the quality of 35 mm films actually shown in theaters degrades over time as the negatives get scratched. Even when first shown, theater films are usually third generation copies of the original negative, and since the reproduction process is analog, resolution is lost from the original. See Vittorio Baroncini, Henry Mahler and Matthieu Sintas, *The Image Resolution of 35mm Cinema Film in Theatrical Presentation*, for details of a human-observer study of film resolution. Even different brands of film differ.

<sup>17</sup> Since 1998, American households went from less than 10 percent of homes owning a personal computer, to over 70 percent of homes having personal computers wired with Internet access. In High Definition television, HD ownership has doubled in the last two years - a quarter of all US households owned HD in 2007, to just under 50 percent of American homes in 2008. In the ubiquitous cell phone market, sales of smartphones such as Apple’s iPhone were over 20 percent of all new handset sales in the US in 2008, up from 12 percent in 2007. Sources: U.S. Census, *Computer and Internet Use in the United States: 2003, October 2005*; Nielsen Wire, *Household TV Trends Holding Steady: Nielsen’s Economic Study 2008*, 24 February 2009; ComScore, *Key Trends in Mobile Content Usage & Mobile Advertising*, 12 February 2009.

<sup>18</sup> Microsoft Email productivity consultants state that effective email users can view and handle 30 percent of their incoming email box in 2 minutes, based on Microsoft Productivity Study (MPS) Statistics. MPS statistics show that on average, people can process up to 60 e-mail messages an hour, where “process” means to complete the full action necessary (not just scan/read – the full sequence is read, respond, assign, delay, or delete). Sources: <<http://office.microsoft.com/en-us/help/HA011464801033.aspx>>; <<http://www.microsoft.com/atwork/manageinfo/email.msp>>; <<http://www.mcgheep productivity.com/library/index.html>>.

<sup>19</sup> Studies of web behavior and navigation find high variability of document display and view time. For example, Weinreich et. al. report: “Our data confirms the rapid interaction behavior with heavy tailed distributions already reported in previous studies... participants stayed only for a short period on most pages. 25 percent of all documents were displayed for less than 4 seconds, and 52 percent of all visits were shorter than 10 seconds (median: 9.4s). However, nearly 10 percent of the page visits were longer than two minutes. Figure 4 shows the distribution of stay times grouped in intervals of one second. The peak value of the average stay times is located between 2 and 3 seconds; these stay times contribute 8.6 percent of all visits.” See Weinreich et al., “Not Quite the Average: An

Empirical Study of Web Use,” *ACM Transactions on the Web* 2, no. 1 (2008): p. 5:18 <ISSN:1559-1131>.

<sup>20</sup> Worldwide user total reported by Facebook at <<http://www.facebook.com/press/info.php?statistics>>.

Average daily time use reported by Silverbean, “Mobile users visit Facebook ‘3 times per day’ - 18th February 2009,” *Online Marketing News* <<http://www.silverbean.co.uk/stories/mobile-users-visit-facebook-3-times-per-day>>.

<sup>21</sup> There are many technology factors affecting the average download speed to a home, including backbone network speed, access connection speed, web server speed, the home network itself, and physical factors such as inside wiring.

<sup>22</sup> Depending on who is counting, Hulu had either 9 million or 42 million viewers in May 2009. See Brian Stelter, “Hulu Questions Count of Its Audience,” *New York Times* 14 May 2009. <<http://www.nytimes.com/2009/05/15/business/media/15nielsen.html>>.

<sup>23</sup> This is with an assumed average video download speed of 1 Mbps.

<sup>24</sup> YouTube’s number of unique visitors grew nine-fold between March of 2006 and March of 2007, and video page-views grew at a rate of 25 times over the same period. In July of 2008, YouTube reported 72 million unique visitors to its US site, 4.7 billion page-views per month, and hundreds of millions of videos viewed daily. Source: YouTube. *You Must Know – July 2008*.

<sup>25</sup> United States National Gamers Survey 2009 available at <[http://corporate.newzoo.com/press/TodaysGamers\\_SummaryReport\\_US.pdf](http://corporate.newzoo.com/press/TodaysGamers_SummaryReport_US.pdf)>.

<sup>26</sup> Anita Frazier, *The Games People Play*, NPD Group, July 2008.

<sup>27</sup> We analyzed computer games in more detail than is reported here. We used a total of 12 categories, which we have summarized down to 4 categories in our tables. For example, our fastest computer runs a screen resolution of 2080 by 1024 at 60 frames per second. This is based on data from Steam.com and other computer game sources. For a low-end laptop, we estimated 800 by 600 at 15 frames per second.

<sup>28</sup> There are no estimates of bandwidths for high-end computer games, and our estimates are therefore plus or minus 25 percent.

<sup>29</sup> See for example John Horrigan, *Home Broadband Adoption 2009*, Pew Internet & American Life Project. Available at <<http://pewinternet.org/Reports/2009/10-Home-Broadband-Adoption-2009.aspx>>.

<sup>30</sup> Chuan-Fong Shih and Alladi Venkatesh, “A Comparative Study of Home Computer Use in Three Countries: U.S., Sweden, and India,” Center for Research on Information Technology and Organizations, University of California, Irvine, Paper 378, 2003; Alladi Venkatesh, “Smart Home Concepts: Current Trends,” Center for Research on Information Technology and Organizations, University of California, Irvine, Paper 377, 2003.

<sup>31</sup> We reviewed multiple sources for data on mobile Internet, text messaging, and mobile gaming use. For

mobile Internet, the Council for Research Excellence (CRE) reported mobile web use of 1 minute per day for the average media consumer in 2008. ComScore M:Metrics reported the average U.S. smartphone user spent 4.6 hours per month browsing the mobile web. When we calculated total annual hours for the U.S. population, we obtained on the order of 2 billion hours for 2008. We therefore did not include this category. Sources: Council for Research Excellence (CRE), *A Day in the Media Life: Some Findings from the Video Consumer Mapping Study*, April 3, 2009. ComScore M:Metrics, "Americans Spend More Than 4.5 Hours Per Month Browsing on Smartphones, Nearly Double the Rate of the British," ComScore Press Release, 21 May 2008.

<sup>32</sup> For mobile gaming, our primary data sources did not break out gaming on smartphones from gaming on dedicated handheld devices. We also reviewed secondary sources on mobile gaming for 2008. All indications are that it is quite small.

<sup>33</sup> Luke Simpson, "Smartphones vs Feature Phones: What's the Difference?" *WirelessWeek*, February 28, 2009 <<http://www.wirelessweek.com/Articles/2009/03/Smartphones-vs-Feature-Phones--What%E2%80%99s-the-Difference/>>.

<sup>34</sup> We estimate that Americans spent 7 billion hours text messaging in 2008. We calculated this amount as follows: Nielsen reported that in mid 2008, the average US mobile customer sent or received 357 text messages a month. Assuming that each text message is sent or received in 30 seconds, and that approximately 200 million American cell phone users subscribed to or paid for text messaging service, multiplying the number of users (200M) by the number of messages (357 per month) by the average time per message (30 seconds) works out to an estimated 7.14 billion hours for the year. Because SMS text messages are so small, the byte totals are insignificant. Sources: Nielsen Wire, "In U.S., SMS Text Messaging Tops Mobile Phone Calling," *Insights*, 22 September 2008; Nielsen Telecom Practice Group, "Flying Fingers: Text-messaging overtakes monthly phone calls," *Insights*, November 2008.

<sup>35</sup> For an analysis of storage costs over time see E. Grochowski and R. D. Halem, "Technological impact of magnetic hard disk drives on storage systems," *IBM Systems Journal* 42, No 2, 2003. <<http://www.research.ibm.com/journal/sj/422/grochowski.pdf>>.

<sup>36</sup> William D. Nordhaus, "Two Centuries of Productivity Growth in Computing," *The Journal of Economic History* 67, No.1, March 2007. (Tables 5 and 6).

<sup>37</sup> This observation was first made by Ithiel de Sola Pool, and was studied in 2005 by Russell Neuman and colleagues. See W. Russell Neuman, Yong Jin Park and Elliot Panek, "Tracking the Flow of Information into the Home: An Empirical Assessment of the Digital Revolution in the U.S. from 1960 – 2005," *International Communications Association Annual Conference*, Chicago, IL. 2009.

<sup>38</sup> Robert N. Charette, "This Car Runs on Code," *IEEE*

*Spectrum*, February 2009. Available at <<http://www.spectrum.ieee.org/feb09/7649>>.

<sup>39</sup> For a description of airbags and how they are activated, see "Inside the Toyota Prius: Part 1 - The airbag control module," *Automotive DesignLine*, 16 April 2007. Available at <<http://www.automotivedesignline.com/howto/199001244>>.

<sup>40</sup>  $100 \text{ Hz} \times 5000 \text{ hours of life} \times 3600 \text{ sec/hour} = 1.8\text{E}+9 = 1.8 \text{ gigabytes}$ .

<sup>41</sup> 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 percent of the total station channels available. Source: Nielsen, 2008 Television Audience Report. Available at <[http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/07/tva\\_2008\\_071709.pdf](http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/07/tva_2008_071709.pdf)>.

<sup>42</sup> "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.

<sup>43</sup> Hard data on this topic is, probably not surprisingly, hard to come by.

<sup>44</sup> Lyman and Varian's original estimate was between 1 and 2 exabytes of information, published in their 2000 report. In their 2003 report, they updated their earlier estimate to 2 to 3 exabytes of information.

<sup>45</sup> Our total for annual fixed line voice information in the U.S. is similar to the totals reported by Lyman and Varian in *HMI? 2003*. Our approaches were different, however. Lyman and Varian asked the question, how much storage would be needed to store all of the fixed line voice calls taking place in the U.S. in 2002? They reported two answers: 9.25 exabytes of uncompressed storage; and 1.2 to 1.5 exabytes of compressed storage (assuming compression would reduce storage requirements by a factor of 6 to 8). To calculate these totals, they consulted Federal Communications Commission (FCC) sources reporting the number of fixed wirelines in the U.S. (approximately 190 million in 2002), and the total number of minutes of use of these lines (4,819 billion DEMS, or Dial Equipment Minutes). Dial Equipment Minutes (DEMS) are measured by telephone switching equipment as "calls enter and leave telephone switches so two dial equipment minutes are recorded for every conversation minute." As Lyman and Varian counted the production of original information, they were interested only in measuring the time of the phone call itself, not how many callers were on the call. Therefore in using DEMS to estimate time usage, they divided the dial equipment minutes in half. They then multiplied usage time by a conversion factor they had previously

defined for storing audio information on storage media – 64,000 bytes per second (uncompressed). Using these numbers and adjusting for the different units of measure, they calculated total annual storage for voice traffic in the U.S. of 9.25 exabytes. They then ran a second calculation for compressed bytes, noting that “compression could reduce storage requirements by a factor of 6 to 8, resulting in a total of 1.2 to 1.5 exabytes.” Our methodology has been to use wherever possible actual device usage time by people in all of our calculations. Importantly, this led us to interpret DEM data differently in estimating conversational minutes of use for phone subscribers. In our approach, two people speaking on the telephone for one hour is counted as two hours – one hour each for each phone caller. Therefore, our information total was calculated as follows: we relied on similar FCC documents as Lyman and Varian to estimate the number of wireline subscribers in the U.S. in 2008 (154 million, including residential, business and some VoIP). To estimate the time usage of these lines, we reviewed studies conducted by the FCC on household wireline penetration and conversational minutes of use. In one study, “Recent developments in US wireline telecommunications,” Paul Zimmerman, an FCC economist, reported that conversational use of wireline phones averaged 900 minutes a month in 2003 (Zimmerman reported this data in Figure 2 Average Monthly Wireline and Wireless Usage by Year (1993-2003), and in Footnote 24, p. 430). We used Zimmerman’s data to estimate an average use time of 22.5 hours per month per subscriber (further details available upon request). We then calculated our annual information total by multiplying the total number of subscribers (154 million), by the average rate of use per subscriber (22.5 hours per month), by the compressed throughput of wireline telephone calls (64,000 bits per second, or 64 kbps). Adjusting for the different units of measurement we calculated a total of 1.2 exabytes a year of information (compressed) for fixed line voice (see Table 4 Telephone Consumption). Interestingly, our total for compressed bytes of annual information is similar to the totals calculated by Lyman and Varian, but we arrived at them by different means. Sources: Federal Communications Commission, Trends in Telephone Service, August 2003; Federal Communications Commission, Trends in Telephone Service, August 2008; Federal Communications Commission, Trends in Telephone Service, July 2009; Paul Zimmerman, “Recent developments in US wireline telecommunications,” Telecommunications Policy 31 (2007), pp. 419-437; Paul Zimmerman, “Strategic incentives under vertical integration: the case of wireline-affiliated wireless carriers and intermodal competition in the US,” Journal of Regulatory Economics 31 (2008), pp. 282–298.

