Digital Magic and the Data Barons

S. Burbeck, (v5.4) 12/27/2018

Abstract

Prior to the invention of computers we cared most about things made of atoms. We needed food, property, land, domesticated animals. We also wanted art, sculpture and music. Once writing became common, we also needed or wanted documents, books, etc. Our cultures, customs, and laws were based on such assumptions. Now that we are in the digital revolution we find that bits play as meaningful a role in the human world as do atoms. Yet human culture has barely begun to adapt to the new realities. In a world of information those who see and exploit the power of bits are having outsized effects on the cultural and economic changes we are experiencing. If we do not shape the creation and provide bounds for acceptable exploitation of bits, the “data barons” – those who monopolize the zetabytes of data about everything we think, do, or buy -- will continue to exploit these new realities and damn the torpedoes.

This paper explores the evolution of the explosive new digital world.

Table of Contents

Digital Robber Barons....3
Magical Digital Worlds....5
    The Evolution of Big Data Computing....7
    Cyberspace and Harry Potter....8
    Digital Spells Flock Together....9
The Data Barons of Personal Data....12
    Wireless Communication....12
    The Effectiveness of Targeted Advertising....13
    Freemail....13
    Anonymity, Private Data, and communications....14
    Magic Cookies....15
    Web crawling and indexing....15
    “The Cloud”....15
    Face recognition and other Photo Surveillance....16
    Speech Processing....17
    GPS and Geotagging....18
    Big Data and Data-centers....19
    Lessons Learned from the Data-Centric Internet....21
Data Barons and the Internet of Things....23
    The RFID IoT....23
        Privacy Concerns....24
    The IoT of “Smart” Things....25
        IoT Botnets....27
Discussion....29
    Data Barons and the Economics of the Digital World....29
    The Evolution of Magic in the Digital World....31
    Out of Control Wizards....31
    Digital Magic in Society....32
    Morality, Privacy, and Insecurity....34
Conclusions....36
Appendix: Civil Law and “Ownership” in Cyberspace....38
    History of Civil Law in Cyberspace....38
    Legal Chicanery: EULAs, “Privacy Policies” and Wiretapping....40

Digital Robber Barons

Eleven of the world's richest 50 billionaires gained their wealth by exploiting bits in various ways. Bill Gates, with $95 billion in 2018, was ranked as the world's richest man for 15 of the past 20 years.[1] Gates' wealth is based on Microsoft's two decade near monopoly of personal computer software. The Microsoft monopoly stemmed from his indomitable will to control PC software rather than some special insight about the future of digital technology. Once his market power was established, Gates demanded and received a software license fee for nearly every PC sold in the world, whether or not it included any Microsoft software[2]. Microsoft demonstrated in a spectacular manner that monopolies can be based on bits as well as atoms, and many other founders of digital companies have followed his lead.

Monopolists, dubbed Robber Barons in 1870[3], have imposed their will upon societies to gain wealth and power throughout the ages. The various products or services that conferred great wealth and power had little in common. Written text is one of the oldest examples. Literacy in the form of ability to write on and read clay tablets or scrolls of papyrus provided a form of power as early as the third millennium BC. For example, in Mesopotamia in 2350 BC,

> "Scribes effectively maintained a monopoly of knowledge where literacy was restricted to a relative few who were trained from birth to belong to the administrative class. Scribal culture was also key in the diffusion of written systems for record keeping and codifying religion that spread throughout the region, particularly to Egypt and other surrounding kingdoms."[4]

Transport of food grain from where it was grown to where it was needed also became a source of monopoly power in the 4th and 5th centuries BC. Half of the grain that fed Athens was shipped from farmers on the shores of the Black Sea through the narrow Bosporus strait that connects the Black Sea to the Mediterranean. Thus much of Greece's food supply was under the control of Leukon, king of Bosporus[5]. Other monopolies were built upon Tea shipped to America (which led to the Boston Tea Party and the American revolution), American transcontinental railroads, oil, salt, diamonds, trade with Asia and India, mail delivery, steel, caviar, telephone service, and the fur trade[6]. Of the above, only the diamond monopoly retains a shadow of its monopoly position today, and it is under increasing attack[7].

Monopolies typically require a confluence of circumstances that reinforce each other. Transport of the grain that fed ancient Athens also required ship builders capable of building the very large grain ships, financiers to bankroll ship construction and risk their money on each cargo, sailors and ship captains to sail the ships, and Black Sea farms to grow the grain. However, monopolies also require at least one person who perceives the confluence of these players and has the foresight and strong will to craft ways to exploit the situation.

[1] As of July, 2017, Jeff Bezos, founder of Amazon, is the richest with $112B
[2] http://en.wikipedia.org/wiki/Bundling_of_Microsoft_Windows
[3] http://en.wikipedia.org/wiki/Robber_baron_%28industrialist%29
[4] http://earlyworldhistory.blogspot.co.uk/2012/02/scribes.html
[5] http://www.pontos.dk/publications/books/bss-6-files/bss6_04_moreno
[6] https://ministryoffear.wordpress.com/2009/01/28/10-greatest-monopolies/
[7] http://www.economist.com/node/2921462

Bill Gates was one such. He perceived that software, especially a PC operating system, could be the focal point of a monopoly because the IBM PC was built from off-the-shelf commodity parts. Prior to the PC, hardware was assumed to be the central controlling factor in computing. That belief lived in IBM's genes hence its executives did not recognize that basing the PC on off-the-shelf hardware allowed the operating system to become the control point. Foolishly, IBM allowed Microsoft rights to MSDOS, and thereby brought about the "PC compatible" computing market. The monopoly control of PCs shifted to what became known as the Wintel duopoly[8]. Since bits are free, Microsoft got the lion's share of the profits. Intel did well too, but computer chips are physical devices manufactured in multi-billion dollar chip fabs, so Intel's time horizon, investment decisions, and design cycles were far longer and more costly than Microsoft's.

Bill Gates was the first Digital Baron to be widely known. However money became digital well before the IBM PC appeared in 1981. Corporate accounting began to move to computers in the 1955 when a UNIVAC computer began running payroll for one of General Electric's factories, and IBM mainframes began to dominate corporate business computing in the 1960s. Credit cards grew popular in the 1960s and 70s as Diners Club, Visa, and the MasterCard consortium grew up. Then bank ATMs made "digital money" publicly visible in the 1970s. However, these advances were all gradual precursors to the magic of digital money. In 1981, when "personal computers" were still primarily toys, magical digital money in the form of computerized financial trading of artificial and abstract bundles of mortgages was growing rapidly inside Wall Street Investment Banks. Fannie Mae issued its first mortgage pass-through, called a *mortgage-backed security*[9] *in 1981*. In 1983, Freddie Mac issued the first collateralized mortgage obligations. By 1982, annual sales of these securities were in the tens of billions of dollars (see history of such new financial instruments[10]). Fortunes were being made in the back rooms of Wall Street investment banks. Digital wizards called "rocket scientists" or "quants"[11] were revolutionizing digital finance.

Many quants are legitimate mathematicians[12]. But many are ethically challenged. As such instruments became popular, the banks and their wizards scooped billions into their pockets without much regard for legality because only the wizards could understand what was happening. Goldman Sachs sold fraudulently structured mortgage instruments[13]. Bear Stearns sold instruments that their own experts called "a sack of sh*t"[14]. And Morgan Stanley fraudulently sold instruments that were were known to be far riskier than they were portrayed[15]. Those sorts of instruments led directly to the 2008 financial crisis.[16] When the party collapsed in what is now called the *Great Recession*, the entire world economy suffered. Yet most of the Financial Barons responsible escaped prosecution, are generally anonymous, and are very wealthy.

Digital technology changes rapidly. The circumstances that come together to facilitate a particular

[8] http://en.wikipedia.org/wiki/Wintel
[9] https://en.wikipedia.org/wiki/Mortgage-backed_security
[10] https://www.huduser.gov/Publications/pdf/HUD-11648.pdf
[11] http://articles.latimes.com/1986-09-27/news/mn-10370_1_term-rocket-scientist
[12] https://www.forbes.com/sites/nathanvardi/2015/09/29/rich-formula-math-and-computer-wizards-now-billionaires-thanks-to-quant-trading-secrets/#5561b1486712
[13] http://www.theguardian.com/business/2010/apr/16/goldman-sachs-fraud-charges
[14] http://www.businessinsider.com/jpmorgan-bear-ambac-lawsuit-2011-1?IR=T
[15] http://www.sec.gov/News/PressRelease/Detail/PressRelease/1370542355594
[16] http://america.aljazeera.com/articles/2014/7/14/citigroup-financialcrisis.html

digital monopoly can be fleeting. Such circumstances also tend to dissipate because the monopoly itself changes the landscape. As Bill Gates learned in 1998, Digital Barons are not forever above the law any more than the Oil Barons were a century ago. Even as the anti-trust suit limited Microsoft's monopoly over personal computing, other computing entrepreneurs found new niches. Apple created a new niche by turning to mobile computers such as the iPod, iTunes, and then the iPhone. Google created another when it monitized search and then bought Android to compete with Apple in mobile computing. And in the late ‘90’s the value of software, per se, was undercut by the Open Source Software movement[17]. Open source alternatives to Windows, Internet Explorer, and MS Office, came from Linux, Netscape’s Firefox, and Open Office. Microsoft, in its overconfidence, complacency, or ineptitude, responded slowly and poorly with Zune (an iPod clone that failed), Vista (a poor attempt to follow on to Windows XP), and Bing (a poor attempt to copy Google search).

Magical Digital Worlds

Thousands of years of human understanding about ownership, ethics and law have evolved to deal with the properties of atoms. The notion of “ownership” of a physical object is a sensible concept, if fraught with all sorts of subtlety. In contrast, the notion of “ownership” of bits is inherently nonsensical -- which copy of the bits, in which location, at what time, and for what purpose. Bits have no inherent location or permanence[18], no mass, color, or any other inherent properties, including any inherent meaning. Whatever meaning bits may appear to have depends solely upon the design constraints, efficiency issues, whims (and mistakes) of the programmers who construct the code that creates and interprets the bits. For example, who “owns” the encryption key which is just a very large bunch of bits that can be displayed as an integer number. Some lawyers once argued that “publication” of an encryption key such as 13,256,278,887,989,457,651,018,865,901,401,704,640 should be illegal[19] However, that set of 16 bytes can be a digital representation of a vast number of different phenomena in the world. Thus the bit pattern is arbitrary and meaningless without the associated information that it is an encryption key of a given type for a particular use. Its use as an encryption key cannot prevent its use in other applications. The notion of digital data is simply too new for human customs and laws to have come to terms with its magic. That uncertainty enables many, if not most, of the ways that data barons have enriched themselves at everyone else's expense.

Unlike the magical incantations[20], spells, and potions used by traditional wizards, sorcerers, or witch doctors, the efficacy of digital magic does not depend upon our belief in magic. Nonetheless, Gandalf’s admonition about magic in the Lord of the Rings still applies:

> ”Perilous to us all are the devices of an art deeper than we possess ourselves.”

Today's magicians are computer geeks of many sorts: software and hardware developers at the big computing powers such as Google and Apple. Or corporate IT experts who create, manage, and protect their company's computing infrastructure. Other skilled magicians become Web designers, SEOs[21],

[17] http://en.wikipedia.org/wiki/History_of_free_and_open-source_software

[18] Bits in DRAM, for example, must be refreshed many times per second. See http://en.wikipedia.org/wiki/Dynamic_random-access_memory. But even bits stored in magnetic media such as tape or disk, degrade slowly

[19] http://en.wikipedia.org/wiki/AACS_encryption_key_controversy

[20] http://www.livescience.com/48833-ancient-egyptian-handbook-spells-deciphered.html

[21] “SEO” is shorthand for Search Engine Optimization intended to raise the ranking position of a page in the results of a

mobile app developers, free-lance white hat and black hat hackers[22], organized criminal hackers, and various government-backed "cyber warriors", even "Cyber-jihadists." There are also a sprinkling of independent computing wizards in nooks and crannies all over the world. Less skilled magicians include so called "script kiddies" who buy, borrow, or steal their hacking tools and local computer shop technicians who fix hacked computers by using publicly available tools to detect and remove rogue software, or just reinstall the entire OS while muttering incantations to their customers to boost their credibility. Digital "help line" staff are given a few magic solutions for frequent problems – the most common incantation is ***reboot***[23] and if that doesn't work, power everything off, wait a minute or so, and turn it all back on (in exactly the order specified in their spells, mind you).

Digital technology is seductive because digital commands often appear to be magical incantations. That fact has fascinated those who came into contact with computers since they were invented in World War-II[24]. The few with access to the rare early computers tinkered with them to contrive new ways to use them for purposes not envisioned by their inventors. Novel applications of these "computing machines" then seduced even more converts to the wonders of digital magic. Early examples of new applications included business accounting, scientific simulations, computer aided design and manufacturing, graphics, and computer games such as Pong. Almost from the beginning, the notion of artificial intelligence lurked in the background as well. As computers became more reliable, cheaper, smaller, and faster, new types of digital data (text for example), and new types of new input-output devices brought the digital age into music, art and film special effects. Protocols for computers to communicate with each other[25] led to the Internet, the World Wide Web and hence Cyberspace. The Internet ignited even more new sorts of digital magic such as web search, on-line retail, social media, "cloud" computing, and monetization of many sorts of data about individuals' views, preferences, and social networks that are of interest to advertisers, politicians...and police or spooks.

The innovativeness we see in the digital revolution is similar to the early stages of prior major transitions. The personal and cultural consequences of new technologies are seldom obvious in advance. The advent and spread of agriculture 12,000 years ago (called the Neolithic revolution[26]) slowly changed most human societies from nomadic hunter-gatherers to agricultural settlements that grew into cities and led to such uses for agricultural grains as beer [27] and pizza[28]. Permanent settlements led to the advent of stone masonry[29] in Egypt and Babylon and its full flowering in the Gothic cathedrals of the 12th century. The Bronze Age (~3000 BCE[30]) and then the Iron Age (~2000 BCE[31]) led to major advances in tools and weapons, and thereby more complex societies. The advent and spread of writing in (3200 BCE[32]) allowed accurate and long lasting communications. The rapid

Google search *en.wikipedia.org/wiki/Search_engine_optimization*

[22] See http://en.wikipedia.org/wiki/Hacker_%28computer_security%29

[23] Few even know the origins or meaning of the terms 'boot' or 'reboot'. See: http://en.wikipedia.org/wiki/Booting

[24] http://www.computerhistory.org/timeline/?category=cmptr

[25] Called TCP/IP protocols. See: http://en.wikipedia.org/wiki/Computer_network#Ethernet

[26] http://en.wikipedia.org/wiki/Neolithic_Revolution

[27] http://en.wikipedia.org/wiki/History_of_beer

[28] http://en.wikipedia.org/wiki/History_of_pizza

[29] http://en.wikipedia.org/wiki/Stonemasonry

[30] http://en.wikipedia.org/wiki/Bronze_Age

[31] http://en.wikipedia.org/wiki/Iron_Age

[32] http://en.wikipedia.org/wiki/History_of_writing

expansion of printing following Gutenberg's metal movable type (1455[33]) dramatically expanded the range and variety of written knowledge. The invention of accurate timepieces (1741[34]), allowed for navigating across oceans which led to easier world-wide commerce. “The Industrial Revolution was the transition to new manufacturing processes in the period from about 1760 to sometime between 1820 and 1840.[35]” The rapid spread of many industrial techniques revolutionized Western society, economics, and culture. It was followed by the discovery and exploitation of petroleum in the 1850s. The advent of aviation in the early 20th century, and the flowering of electronic communication such as telegraphy, radio, television and other analog devices, increased the spread of knowledge and culture.

Technical revolutions begin in obscurity and, as they spread, attract many sorts of related or derivative innovations; think of the many innovations in agriculture developed since its first adoption[36]. Some invent wholly new ways of using or augmenting the basic innovation; primitive astronomy, called astrology, was invented to guide planting and harvesting of crops[37]. Others think of ways to generalize the original innovation to serve new purposes. Some think of ways to spread the technology to a wider audience or, conversely, to keep the innovation secret to be the province of a priesthood: for example, the Freemasons maintained a monopoly on the techniques of stone masonry used to build the Gothic cathedrals[38]. And with each new technology, there are a few whose innovative business practices aim to own and control most of the benefits of the technology. For example, Rockefeller's innovations gave him a virtual monopoly on the distribution of oil. These innovators become the Robber Barons of each new technology. Rockefeller was dubbed an Oil Baron. And Bill Gates became the first publicly known Data Baron.

The Evolution of Big Data Computing

Early computers were large, rare, and exceedingly expensive. Thanks to the operation of Moore's Law[39] digital hardware became smaller and cheaper which both made the large business computers more and more powerful, and enabled personal computing devices optimized for use by a single user. The early PCs sat on desktops and provided text processing, spreadsheets, and games. Now they can live in our pockets, or on our wrists, or in our athletic shoes. Smartphones at present are the pinnacle of that branch. They increasingly are filled with data about everything of interest to the owner.

Huge data-centers, each containing tens of thousands of powerful connected servers, are the descendants of the 1950s mainframe computers that used to serve business and government needs for data storage and analysis. However massive data-centers, e.g., those serving Google, Amazon, Facebook and many others are increasingly occupied with capturing, storing and mining data about individuals obtained from their poorly secured smartphones.

Personal computing as seen by the users is increasingly ever present, addictive, and even whimsical. It gives rise to novel apps at an astounding rate. And each of these apps generates (and captures) new

[33] http://en.wikipedia.org/wiki/History_of_printing

[34] http://en.wikipedia.org/wiki/History_of_navigation

[35] http://en.wikipedia.org/wiki/Industrial_Revolution

[36] http://archaeology.about.com/od/neolithic/tp/ancient_farming.01.htm

[37] http://en.wikipedia.org/wiki/Agricultural_astrology and http://ancienthistory.about.com/od/basics101/a/090510-Ancient-Calendar.htm

[38] http://www.bbc.co.uk/history/british/middle_ages/architecture_medmason_01.shtml

[39] http://en.wikipedia.org/wiki/Moore%27s_law

types of data about users. Such data is for sale or rent to advertisers or governmental surveillance organizations. We will explore the culture of personal computing in some depth to see what lessons can be learned.

Business computing, requiring as it always has, big expensive and above all stable computing systems, began as a service to the company. The Web became a valuable capability as businesses created their own web sites and did business over the Web. Then various disasters befell companies that did not architect their IT systems for security. Companies such as Target, Home Depot, Sony, Anthem and others have discovered that being hacked can be an existential threat to the business. Recently, IBM among others is demonstrating that savvy data mining can convert their computing infrastructure and data into a source of revenue and important business insights. And, as Google has demonstrated, gathering and mining data from a billion personal computers can provide very large revenues.

We create new bits in a digital world when we type, touch a screen, move a cursor, speak or make gestures that the user interface of our digital device detects. Only a tiny minority of people are aware of the degree to which the consequences of such simple actions in a digital virtual world depend upon the whims of various programmers, bugs in the software, or deliberately hidden violations of our expectations. Each and every tiny mouse movement, touch on a touch sensitive screen, click or tap, or keystroke in the digital world is a separate act to be monitored by software. The resulting data can be ignored, counterfeited, recorded or reported to some interested observer. Such tricks are exploited by malware as well as by corporations that gather data about users for financial gain. The user does not necessarily even control whether the device's microphone, camera, GPS sensors or accelerometers are reporting their data to some unknown observer (human or digital).

The founders of Google, Facebook, Twitter, and others have become billionaires by collecting and claiming monopoly ownership of various sorts of personal data. They may be best thought of as *Data Barons* akin to the 19th century Robber Barons such as John D. Rockefeller (Oil), Andrew Carnegie (Steel), Leland Stanford (Railroads) and many others. We will further explore these issues to see what lessons can be learned.

Cyberspace and Harry Potter

The generation that grew up with the Web also grew up with Harry Potter[40]. J. K. Rowling's world shows us a rich sociology of wizards of various sorts together with “muggles”, the majority in society who do not have magical talents. Harry Potter fans take magic for granted in the fictional world of Hogwarts. And the Harry Potter generation also takes for granted the ever present world of digital magic. Digital Muggles are looked upon by wizards and wannabe wizards with the same disdain in the “real” world as they are in the Harry Potter books and movies.

Digital wizards come in many flavors according to their particular talents and inclinations. Some focus on what we will call “White Magic” that furthers mankind's knowledge helps us to collaborate together and share such knowledge. Some tend to become arrogant, greedy, or hungry for power and use their magic talents to exploit or control others. That's “Black Magic”. And others, perhaps the majority, are

[40] I would argue that growing up with the web predisposed this generation toward acceptance of magic and perhaps growing up with Harry Potter predisposes them toward immersion in the web. The web certainly appears magic to all but the few who understand, or think they understand, how it works.

interested in the utilitarian uses of magic and let others choose how to use the magic at their disposal. Call their contributions "Gray Magic".

Digital muggles are considered as little more than sources of free data and pliable consumers easily manipulable by appropriately targeted ads. In short, muggles are suckers who provide, for free, the very data that makes them manipulable. And digital muggles are led to believe that the companies that exploit data about them "own" that data and can use it arbitrarily. Our current legal system does little if anything to disabuse them of that myth. For example, Google data-mines any email sent to gmail addresses with no shame or acknowledgment that the sender has in no way relinquished rights to the content of the email. Google merely asserts that you have no "reasonable expectation" of privacy whether your get your mail at gmail.com or you send mail to a gmail address. If even a minor wizard were to spend a moment thinking about that fact, it would be clear that you should never send anything that for any reason should remain private to a gmail address. But muggles are oblivious to that fact. Muggles, after all, are fair game for the black magic that Google employs. And because legislators themselves are typically muggles, and Data Barons wield powerful lobbying tools, legislatures are doing little to protect the privacy of muggles. Given that such muggles are also citizens, many are voters, and all supposedly have rights to privacy, legislatures may one day be forced to rethink...but clearly not yet.

Stanford, MIT, Apple, IBM Research, Google, Facebook, and Microsoft, among many others are centers of research into digital magic. They play a role similar to Hogwarts. Each can boast of one or more Albus Dumbledore equivalents, several more specialized heavyweight digital wizards, and a host of journeymen and apprentices of varying skills and promise. There are parallels between students of magic and students of programming. As we saw in scenes from Hogwarts' classrooms, when the students get the spells wrong, the magic fails or surprising results ensue. Similarly, when students of digital magic get their spells (code) a little bit wrong the magic fails or strange bugs occur.

In our world today it may seem at times as if modern digital wizards are everywhere. In truth they are a small minority except perhaps in high tech enclaves such as Silicon Valley. Wizards run Google, Facebook, Apple, Microsoft, Twitter, and many less well known tech companies. But there are also Voldemorts in digital wizardry. As we discussed above, wizards also brought us the *Great Recession* and 7x24 worldwide surveillance. Those powerufl Voldemorts are now in an all-out battle to gather and claim "ownership" of every bit of data that they can get. In the digital world of today, people from cyber-criminals to computer vendors to governmental hackers in China or the NSA are trying to grab data because data has both military and commercial value. A new class of monopolistic, which I call Data Barons, has emerged and is attempting to taken charge.

Digital Spells Flock Together

Code can indeed be likened to magic, but its granularity is completely unlike the portrayal of magic in the Harry Potter series or in the Lord of the Rings. Small bundles of lines of code may (and ideally should) serve a single purpose much as individual spells are imagined to do. But laptops, tablets or smart phones contain thousands of such spells[41]. Each one can be likened to the entire repertoire of

[41] The Linux, OS-X, and Windows operating systems each expose thousands of Application Programming Interface calls (APIs) each of which invokes a somewhat singular spell. See http://kernelbook.sourceforge.net/kernel-api.pdf for an example set of APIs for the Linux kernel. Mac and Windows APIs are similar. Note that the kernel is but one part of the entire set of APIs for a given computer.

Hogwarts' magicians with all of their spells! Some code serves the many agendas of the operating system's masters[42], some serves the application developers' agendas or the anti-virus security developers' agendas or the SMS or email or Skype or Facebook agendas, or the software update system's agendas, and so forth.

When you turn on your computer (and your smartphone is definitely a full-blown computer) you are metaphorically giving “life” to a clamoring horde of digital magicians to begin competing for CPU time and bandwidth to execute the spells that further their many agendas. The smartphone, in its “spare time” also serves the human user's agendas. The illusion that it is mostly doing your bidding can be sustained simply because today's smart phones, tablets and laptops have the compute power of a 1990's supercomputer and can do many things at once – “multitasking” as it is called amongst the wizards. The Linux laptop on which I write currently has about 75 processes running in the background and/or waiting for some event to wake them into action.   I see very little of that activity: just a word processor seemingly passively waiting for me to press a key or move the mouse. So, how many spells are in your laptop, tablet, or Pocket? And who controls them? Answer: no one necessarily knows, not even the wizards! But if you ask who controls the data gathered by those spells, the answer is clear: The Data Barons of the 21st century have managed by hook or by crook to effectively own most of your data.

Many websites have nothing but benign purposes (unless they have been hacked so that they install malware on the visitor's computer). Wikipedia, for example is a world-wide resource created primarily by volunteers. But websites or software apps may embody many purposes, some good, some evil, and some in a gray zone. For example, Google search both helps the user find what they want (good) and gathers data from the user for sale to the highest bidder (gray or evil according to your taste). A free game may be fun to play (good) but also may infect your smart phone with a virus (evil). Facebook allows you to show your photos to friends (good), but also gives Facebook valuable commercial advantage by associating names and locations to the faces in those photos (whether or not the faces belong to Facebook users who could be construed to have given permission) and can collect all the face information across all their customers to learn who associates with who. That is at best a gray area for Facebook users but not at all welcome to those not enamored of Facebook who find that their photo, with name and perhaps location is available online.   And Cambridge Analytica used such information to bias the 2016 US Presidential election in favor of the Republicans. We will discuss these issues in more detail later.

It is nearly impossible for the layperson to discern the purpose of most software in the many digital devices they own. Software apps almost always contains many “spells” and digital hardware almost always contains multiple software libraries and apps. Perhaps the best clue to the purpose of these spells is the *proprietary interest* various wizards have in the code or data on your smartphone, tablet or laptop. Unsurprisingly, the wizards' agendas are influenced strongly by their employers. Companies that prosper by the sale of hardware – Microsoft by contractual license fees and Apple by sale of the hardware itself – want to maximize the sale of those computers. Their software developers take pride in making the computers more desirable to their users. However these companies also wish to learn how their customers use the devices and often where they use it and what other software they buy. In

[42] For example, sending your location (GPS or from nearby cell towers) to the vendor of the machine. Both Microsoft and Apple do that unless you turn off location services (and may continue doing it even so). http://news.yahoo.com/blogs/technology-news-blog/windows-8-automatically-report-every-program-download-microsoft-164331963.html

that case, the company wizards take pride in how stealthy they can be in gathering such data. For example, Google's Android operating system now gathers geographical location data in near real-time:

> “In the case of Google, according to new research by security analyst Samy Kamkar, an HTC Android phone collected its location every few seconds and transmitted the data to Google at least several times an hour. It also transmitted the name, location and signal strength of any nearby Wi-Fi networks, as well as a unique phone identifier.”[43]

[43] http://online.wsj.com/news/articles/SB10001424052748703983704576277101723453610

The Data Barons of Personal Data

The Internet emerged into full commercial use in the mid 1990s[44]. In the subsequent decade, Google search, blogs, Amazon, Twitter, Dropbox, dating sites, Facebook, YouTube, Snapchat, Instagram, Pinterest, stumbledupon, digg, tumblr, LinkedIn, Flickr and many others appeared and became popular. However, they are only the visible tip of the iceberg of Internet magic, the tip that is appealing to end users, the tip that seems like white magic. The visible portion is buoyed up by the nine tenths of the magic that is hidden under the surface. Let's examine the technologies used by many visible Internet services that we take for granted or don't notice.

Wireless Communication

Today's cyberspace would be impossible if every computing device had to be connected to the Internet via physical wires. That was essentially the case prior to 1999 when the 802.11b Wi-Fi standard was agreed upon to enable wireless local-area networks[45]. Now few personal digital devices are tethered to wired communication.

We've come a long way since then. Wireless communication is nearly ubiquitous via cell-towers. Public Wi-Fi is common in coffee shops. And GPS signals cover the planet. Texas Instruments sells a single chip transceiver designed for very low power and very low voltage wireless applications. The chip can easily be programmed for operation at any frequency in the 300-1000 MHz range. And Bluetooth Personal Area Networks support communication at higher frequencies at shorter distances. There are other high frequency protocols as well, e.g., Zigbee. Today it is possible to establish radio connection between any two digital devices.

Little noted is the fact that all this RF information is undetectable by humans; we perceive it only through our electronic devices. Most of us interact only via cell-phone and Wi-Fi enabled personal computers so we assume that all this wireless information is helpful to us and private to us. Those with differently specialized RF devices perceive a different world, in which they can eavesdrop on or collect other's not-so-private information. Where there is white magic, black magic follows. White magic conversations are at our behest, helping us to find information on the Web or communicate with our friends and associates. Black magic conversations, or eavesdropping, are “about us” at the behest of a growing number of nearby digital devices: cameras in other people's smart phones, microphones, RFID devices in our belongings or credit cards, or Big Data gatherers such as Google conversing with their “home” data centers about our use of and location of our smart phones. According to the Wall Street Journal[46], “an Android phone collected its location every few seconds and transmitted the data to Google at least several times an hour. It also transmitted the name, location and signal strength of any nearby Wi-Fi networks, as well as a unique phone identifier. …Apple, meanwhile, says it 'intermittently' collects location data, including GPS coordinates, of many iPhone users and nearby Wi-Fi networks and transmits that data to itself every 12 hours … some of the most popular smartphone apps use location data and other personal information even more aggressively than this—in some cases sharing it with third-party companies without the user's consent or knowledge.”

[44] http://en.wikipedia.org/wiki/History_of_the_Internet

[45] http://en.wikipedia.org/wiki/Wi-Fi

[46] http://online.wsj.com/news/articles/SB10001424052748703983704576277101723453610

The Effectiveness of Targeted Advertising

Monetizing the vast troves of data gathered by Google, Facebook, Twitter and many other would-be data barons primarily depends upon “Targeted advertising”, i.e., showing ads to those most likely to act upon them. The reason advertisers are willing to pay to have their ads preferentially put before the person viewing the page is that behaviorally-targeted advertising generated an average of 2.68 times as much revenue per ad as non-targeted advertising[47]. As Wikipedia explains it,

> “Behavioral targeting is the most common targeting method used online. Behavioral targeting works by *anonymously monitoring and tracking* the content read and sites visited by a user or IP [address] when that user surfs on the Internet. This is done by serving tracking codes *[pixel tags]. Sites visited, content viewed, and length of visit* are databased to predict an online behavioral pattern. A further refinement to behavioral targeting is Predictive Behavioral Targeting, where machine learning algorithms overlay behavioral patterns with sampled data, to create data-rich predicted profiles for every user.

Because targeted advertising is so effective, the big players who use it choose to ignore public preferences that clearly show public displeasure at being tracked. The purpose of most of the technologies described below is to support more and more effective methods of gathering data and targeting ads. The Data Barons dependent upon targeted advertising have created a congressional lobby focused on undercutting any efforts to regulate data gathering[48].

Freemail

In the early days, email was not only free but easily readable by the many sites that forwarded email on to other sites. No one thought about or even cared that others could read their email in part because no one imagined it could be “monitized”. That seems quaint now. Google's gmail has over 500 million active users and every email from or to a gmail account is mined for Google's behavioral modeling of individuals.

Free email exists to entice users to provide more data about themselves for the email company to exploit. Remember, if it's free, ***you are the product*** for sale. Google's bots mine your social network from email address lists and mine the contents of the email for clues useful for targeted advertising. As Eric Schmidt put it, Google's Policy is to “Get Right Up To The Creepy Line” of violating their customer's privacy[49]. In fact, Google has leaped, not tiptoed, over the line many times (e.g., recording wireless traffic from their Street View cars.) Nonetheless, they claim to be holier-than-thou even though Snowden's revelations indicated that they have cooperated at least passively with NSA's surveillance for years.

White magic, or black? Either way, it is clear that you and your attitudes and preferences and social contacts are the product and their wizards aren't in the least concerned about your privacy.

And then there's spam or “junk” email– a whole topic in itself. Spam accounted for some 70% of all email in 2013[50]. The white hat wizards at email providers, who wish to filter out spam, constantly vie

[47] http://www.bizreport.com/2010/03/nai_behaviorally-targeted_online_ads_twice_as_effective.html

[48] http://techfreedom.org/, It is staffed by ex Congressional staffers and lawyers. See http://techfreedom.org/staff

[49] http://daringfireball.net/linked/2010/10/04/schmidt-creepy

[50] http://www.zdnet.com/article/worldwide-spam-rate-falls-2-5-percent-but-new-tactics-emerge/

against the black hat wizards all over the world, especially in China, Russia, and the Ukraine, who send spam for profit. There is no difficulty here in discriminating black magic from white since very few people desire spam. The simplest way to discourage spam would be to charge a small fee for each email sent. A penny a message would discourage most of it because spam works on a statistical model that requires huge numbers of spam messages sent to expect much financial income[51]. But email users resist that idea.

Anonymity, Private Data, and communications

In our real-world lives, anonymity is expected in most situations because we humans don't attend to everything we perceive and in any case we don't remember most of what we only fleetingly notice. Witnesses to a traffic accident are notoriously incapable of describing accurately what they saw.

Anonymity is a myth in the digital world – a myth that primarily serves to lull its visitors into complacency. In the digital world, everything is recorded, nothing is forgotten, and data can later be stitched together from multiple sources to identify nearly everyone.

Within a single particular Big Data set ***perhaps*** people can be made anonymous although, as big data gets bigger, even that is problematic. Rich context often gives the game away if one collects together many sets of data. That is especially true if any of the data sets include geocoded photos or tweets, or photos of faces, or phone or email "metadata". A growing number of big data exploiters do collect together such data, so it becomes relatively trivial to de-anonymize the data. Axciom and RocketFuel (discussed below) and Google, Amazon, and others have shown that it is already relatively easy to identify by name, age, location, cultural heritage, economic status, political views, and the state of health of nearly everyone. Companies that provide "free" email, e.g. Google, Yahoo, and Microsoft, have the added benefit of the content of all its mail subscribers. And governmental agencies avail themselves of all sorts of corporations' sources of data that they buy, "borrow," or steal because buying data from corporate sources is not considered illegal. As long as these practices and business models remain unfettered, whatever anonymity remains will continue to be eroded.

For many people in authoritarian countries, the ability to have confidential email communications hidden from their governments is important, and in some countries it is a matter of life and death. One tool for preserving anonymity is Tor (**T**he **O**nion **R**outer[52]). TOR is the main portal into the "Darknet", or "Dark Web", the large portion of the Internet not accessible to people that do not have special software, e.g., TOR. The US "security" industry avidly wishes to end Tor's affront to their power. Tor itself has warned its users of an ongoing attack[53]. However it seems likely that wizards within NSA and the British GCHQ are surreptitiously warning Tor engineers about flaws that NSA or GCHQ may exploit so that Tor can fix them quickly[54]. It is therefore not clear who are the white hats and who are the black hats of TOR anonymity.

[51] http://www.newscientist.com/article/dn17577-payperemail-plan-to-beat-spam-and-help-charity.html

[52] http://eandt.theiet.org/magazine/2014/08/tor-hack.cfm

[53] http://www.rawstory.com/rs/2014/07/30/internet-privacy-service-tor-warns-users-it-was-attacked/

[54] http://www.bbc.com/news/technology-28886462

Magic Cookies

A cookie is a small chunk of data sent from a website that is stored in a user's web browser[55]. Thereafter, when the user visits that website, the browser sends the cookie back to the server to notify the website of the user's previous visits. The data is typically obscure or encrypted. Some cookies support black magic such as stealthy ways of tracking your browsing from web site to website.

> “Advertising companies use third-party cookies to track a user across multiple sites. In particular, an advertising company can track a user across all pages where it has placed advertising images or web bugs *[or pixel tags[56]]*. Knowledge of the pages visited by a user allows the advertising company to target advertisements to the user's presumed preferences.”

Most people who surf the web know little or nothing about cookies and those who do simply have to live with inescapable cookie magic.

Web crawling and indexing

Tim Berners-Lee invented the World Wide Web in 1989. It grew rapidly. By the end of 1994, there were more than 10,000 web pages, necessitating ways to search for pages of interest. By 2000, Google began to dominate Web search on the basis of its page-rank algorithms that, in addition to the presence of a word or phrase, took into account the number of links to a page from other highly referenced pages, the more links and the higher-quality of the links, the higher the page is ranked in the results of a search. Google remains dominant today although their page rank algorithms are vastly more sophisticated now. There is now a discipline, called Search Engine Optimization (SEO), that seeks to understand and manipulate Google's ranking to benefit their client's website traffic. Google and the SEOs are in an arms race. Google's wizards attempt to spot and counter SEO manipulations and SEOs wizards attempt to spot and counter Googles' efforts. It is likely that most of these wizards on both sides think of themselves as the “good” white wizards. But that situation is exceedingly complex.

Web crawling is also used for purposes other than supporting search. Robots accounted for over 52% of Web traffic in 2018. White magic bots, or “good” bots focus on supporting search engines, archiving the web, web performance tools, and search engine optimization (SEOs). But the black magic bots, or “bad” bots tend to support malicious behavior such as cyberattacks and spam. Little if any of the robot search web traffic is visible to the average Web user.

“The Cloud”

“The Cloud” is a magical term for using computers accessible via the Internet for data storage and compute-cycles rather than your own computers. Internet connection bandwidth has grown to the point where the interconnection times and speed of transmission (G bytes per second) can make their location irrelevant *from a technical and cost perspective[57]*. Many cloud server firms have data-centers full of servers for rent[58]. Although their location is irrelevant in theory, the old adage – *possession is*

[55] http://en.wikipedia.org/wiki/HTTP_cookie

[56] http://en.wikipedia.org/wiki/Web_bug

[57] Performance of some applications “in the cloud” can be spoiled by interconnection time and insufficient bandwidth. See www.computerworld.com/s/article/9223117/Bandwidth_bottlenecks_loom_large_in_the_cloud

[58] http://www.cbronline.com/news/cloud/cloud-saas/10-top-cloud-computing-providers-for-2014-4401618

nine tenths of the law – may still apply! The physical location of computers and their data stores can affect security and legal status. Data in the cloud is not physically under the control of its “owner” so data security is enforced by software magic. The “owners” of the data may not know who accesses it or for what reason. Renters of virtual PC, Mac, or Linux virtual machines in the cloud are expected to set up their own security in the operating system. Few of these renters are security experts. Ensuring that virtual disks are not accessed by unauthorized users is difficult unless they are encrypted. But encryption is a black art as well, and NSA works diligently to put backdoors into encryption algorithms[59]. Moreover, the law is not at all clear about such things as subpoena in international cases[60]. For example:

> “The jurisdictional reach of a US search warrant is currently the subject of litigation in the US, brought by Microsoft and challenging a request for data held in their Irish data centre (*In the Matter of a Warrant to Search a Certain E-Mail Account Controlled and Maintained by Microsoft Corporation*, June 2014[61]). Microsoft, with supporting submissions from Verizon, AT&T, Apple and Cisco, is arguing that US search warrants are limited to servers and infrastructure located in the US.”

Who will win the battle? If you need to ask, perhaps the cloud isn't for you.

Face recognition and other Photo Surveillance

One of the early computer AI goals was interpreting the contents of photographs. The web contains many photos, especially of people. Facial recognition systems have made great strides since the earliest research in the mid 1960s[62]. Now, “...DeepFace (Facebook's research system[63]) can analyze two photos, and irrespective of lighting or angle, say with 97.25% accuracy whether the photos contain the same face. Humans can perform the same task with 97.53% accuracy.” Since Facebook has over a two billion users, most of whom submit photos of themselves, family and (real physical) friends, it should be assumed that any face appearing on any Facebook page, or any public photo on the web that is available to Facebook, has a good probability of being identified by Facebook itself. Google is working to monetize its trove of images from Gmail. Google's terms of service warns the handful of people who read and understood it that they automatically analyze emails which, by the twisted “logic” of EULAs, makes it all OK because they aren't analyzing specific images for specific purposes. Their purpose is more grandiose: to gather all possible data on all possible photos.

The hashing and tagging technology is completely agnostic as to the human content of the images it analyzes. It works as well searching for brand images, product photos (of interest to advertisers), and photos of political figures (of interest to political campaigns). One of the most active disputes about Internet photographs is centered around Google's Street View[64] which could be used anonymously to identify terrorist targets and their surroundings, to pick out burglary targets and, more trivially, spot

[59] http://www.reuters.com/article/2013/12/20/us-usa-security-rsa-idUSBRE9BJ1C220131220

[60] http://www.informationintersection.com/2013/04/the-stored-communications-act-and-document-subpoenas-to-cloud-computing-providers-2/

[61] http://www.nysd.uscourts.gov/cases/show.php?db=special&id=398

[62] http://en.wikipedia.org/wiki/Facial_recognition_system

[63] http://www.extremetech.com/extreme/178777-facebooks-facial-recognition-software-is-now-as-accurate-as-the-human-brain-but-what-now

[64] http://en.wikipedia.org/wiki/Google_Street_View_privacy_concerns

violations of building codes. Google points out that they deliberately blur parts of their photos. But blurring does not affect many illicit uses and Google quite likely retains unblurred copies.

More effective image analysis interacts with the rapid growth in the use of CCTV. In US cities there are now many CCTV cameras on poles and overpasses monitoring and identifying cars (by license plate) and drivers (by video and face recognition). In UK cities there are many CCTV cameras that can, after the fact, place people and their actions to the second and the meter. In the US, there are many "security cameras" in cities. For example, the security cameras along the Boston Marathon course identified the Boston Marathon bombers within a day of their attack. All to the good, one might say. Clearly this counts as white magic, at least as long as they are used only after the fact of some serious crime. But it is after the fact primarily because it takes lots of human labor to go through all the video. Face recognition software changes that tradeoff. In the UK, police and intelligence agencies have been especially eager to use face recognition software to identify criminals in CCTV surveillance videos.

But just who is a criminal? Our legal systems were not created for the harsh literalism of computer surveillance. Black magic lurks therein. Human enforcement of the law relies upon human judgment and flexibility. Cyber enforcement based upon watching people *en mass* can catalog everyone's whereabouts, their associates, and every little breach of our creaky legal systems. That can make everyone a criminal in the "eyes" of data mining software. Thus those who can afford the necessary data mining – wealthy individuals, corporations, and governmental agencies – can deter legal attacks on themselves by exposing every illegal act on the part of their attackers, much the way presidential campaigns (who are heavy users of data mining) attack each other. It may not be long before someone suing a party with access to big data mining will risk a defensive barrage of attacks on the attacker's "illegal" behavior that their target might have discovered in big data troves.

Speech Processing

Many of our devices can now speak to us and be controlled by speech in various languages. There is "in car" speech communication with one's smartphone. At least 400 million voice controlled "Smart speakers" such as Amazon's Alexa, Google Dot are in use.[65] You can speak to your Apple watch, and control many home IoT devices using speech, e.g., lights, thermostats, crock pots, and "Smart" TVs.

It has been pointed out that entering text by speaking is clearly superior to typing – we easily speak at 150 words per minute but most people cannot type much more than 50 words per minute. However, given that the data barons try to collect data on everything you do, wouldn't they also want to eavesdrop on everything you say in your home or car? The answer is likely to be 'Yes'.

> As with all "smart home" devices, the convenience of the technology is not without risk. For example, if you buy a security camera, or even a baby monitor, the devices ship with default passwords that are well known by hackers. If people don't change the password, then it's possible for hackers to access them. While many, even most people may read the instructions and change the password, countless others ignore this advice and simply leave the default settings, potentially letting malicious people have eyes on their home. This is dangerous in two ways: the first is that your activities can be recorded, and the second is that these cameras can show when you are not home, allowing criminals to know when it's safe to burglarize your

[65] https://www.globalme.net/blog/the-present-future-of-speech-recognition

> house. [66]

How would we know if our every word is being gathered? Answer: we wouldn’t. It is a matter of trust. And the Data Barons have not engendered much trust.

GPS and Geotagging

While bits have no geography themselves, the US government's satellite GPS systems can provide location information to any physical digital device with appropriate radio receivers and compute power. GPS was fully operational in 1995 and is now taken for granted for navigation by foot, car, boat, or air. Most mobile digital devices now support GPS, beginning with 3G iPhones and since adopted by other smartphones and iPads and their imitators. GPS magic has made its way into many digital worlds via geotagging[67]: the inclusion of the GPS location of the device in SMS messages[68], digital photos or tweets[69]. Geotagging information is also commonly available in Flikr, YouTube, and Craigslist, plus many websites that can be found by easily constructed search scripts.

There have already been unintended consequences. The US Army is concerned because “In 2007, geotagged photos of a new fleet of helicopters allowed enemy forces to mortar the base and destroy several of them; it could just as easily have been a field hospital or barracks.”[70]

A Forbes article[71] includes the following from a talk given at the 2010 Hackers On Planet Earth (HOPE) conference:

> Jackson and fellow researcher Larry Pesce plan to release two tools, affectionately named Reaper and Stalker, that use Perl scripts to harvest data from the stream of location-tagged photos that are continuously posted via Twitter on services like Twitpic, YFrog, the more risque SexyPeek. Jackson says about 3% of the photos on those services include GPS tags showing the exact latitude, longitude and direction of the pictures–data that most users likely never intended to reveal when they snapped the photos on their GPS-enabled phone. … The privacy vulnerabilities that Vet and Jackson are aiming to highlight may go well beyond a few ex-boyfriends hiding in bushes. “I’m just one student that threw this together,” says Vet. “Organizations like marketers and governments must be doing the same thing on a much larger scale.”

Thus the combination of face recognition and geotagging can easily be used to place the people in a photo at a specific location and time! See Why stalkers or thieves like geocoding[72] and the paper “Cybercasing the Joint: On the Privacy Implications of Geo-Tagging.[73]”

[66] https://www.intego.com/mac-security-blog/is-your-smart-speaker-spying-on-you/

[67] http://webtrends.about.com/od/glossary/a/what-geotagging.htm

[68] http://geosms.wordpress.com/2010/10/18/a-history-of-sms-geotagging/

[69] http://www.pcworld.com/article/182729/twitter_geotagging_what_you_need_to_know.html

[70] http://techcrunch.com/2012/03/09/army-warns-of-danger-of-geotagging/

[71] http://www.forbes.com/sites/firewall/2010/07/19/researchers-show-how-twitter-twitpic-make-stalking-simple/

[72] http://netsecurity.about.com/od/securityadvisorie1/a/Why-Stalkers-Love-Your-Geotags.htm

[73] http://www.icsi.berkeley.edu/pubs/networking/cybercasinghotsec10.pdf

Big Data and Data-centers

There are about 5 billion mobile devices world-wide in 2018. The volume of data generated by personal mobile devices is inherently peta-scale (a petabyte is 1000 terabytes). *Wired* reports[74] that:

> "To put things into perspective, 1 Exabyte (10^18) of data is created on the internet daily, amounting to roughly the equivalent of data in 250 million DVDs. Humankind produces in two days the same amount of data it took from the dawn of civilization until 2003 to generate, and as the Internet of Things become a reality and more physical objects become connected to the internet, we will enter the Brontobyte (10^27) Era."

The increasingly popular term, "Big Data analytics" implies the ability to gather, curate, store, search, *and analyze* multi-terabyte, or peta byte[75] datasets without data access performance issues unduly slowing the analytics algorithms. At that scale, even quite simple analyses using standard relational database techniques tend to break down.[76]

It is well known that companies with easy access to user contributed data – the likes of Google, Facebook, Twitter, Microsoft, and Apple – analyze that data for economic gain. Other lesser known companies even more zealously gather data. Two such are:

Acxiom is a little known company with a big reach. It wants to know everything that everyone does in the commercial world:

> "...analysts say it has amassed the world's largest commercial database on consumers — and that it wants to know much, much more. Its servers process more than 50 trillion data 'transactions' a year. Company executives have said its database contains information about 500 million active consumers worldwide, with about 1,500 data points per person. That includes a majority of adults in the United States.
>
> ...It peers deeper into American life than the F.B.I. or the I.R.S., or those prying digital eyes at Facebook and Google. If you are an American adult, the odds are that it knows things like your age, race, sex, weight, height, marital status, education level, politics, buying habits, household health worries, vacation dreams — and on and on.
>
> ...Today, Acxiom maintains its own database on about 190 million individuals and 126 million households in the United States. Separately, it manages customer databases for or works with 47 of the Fortune 100 companies. It also worked with the government after the September 2001 terrorist attacks, providing information about 11 of the 19 hijackers.[77]

Since their goal is to know everything about everyone, it is very likely that they also scrape geotagged photos from personal websites, Facebook pages and tweets to deduce important data from locations at which people think it important enough to take photos. Such information indicate what activities they favor and who their friends are (by face recognition of all the people in the photos). One of their customers is NSA. This is definitely not white magic.

[74] http://www.wired.com/2013/04/with-big-data-context-is-a-big-issue/

[75] The threshold to be considered "Big" isn't fixed See: http://en.wikipedia.org/wiki/Big_data

[76] "The Pathologies of Big Data", Adam Jacobs, ACM *Queue vol. 7, no. 6,* 2009, http://queue.acm.org/detail.cfm?id=1563874

[77] http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html?_r=1&

RocketFuel wants to know everything you do on the Web – Using stealth cookies and links that are hidden in a substantial proportion of commercial webpages, they can follow you as you browse the web. Because so many websites notify them when you surf the pages, they receive many times more hits than Google! They then use "...big data and machine learning to place online ads [targeting] the most relevant demographics for marketers." Their CEO, George John, was one of the wizards who taught artificial intelligence at Stanford.

All this data gathering exists in a confusing and gray legal environment. The US Government Accounting Office published a report in 2013 about the legality of such data gathering and selling[78]. Its introduction summarizes the status of the law today:

> 'No overarching federal privacy law governs the collection and sale of personal information among private-sector companies, including information resellers. Instead, a variety of laws tailored to specific purposes, situations, or entities governs the use, sharing, and protection of personal information. For example, the Fair Credit Reporting Act limits the use and distribution of personal information collected or used to help determine eligibility for such things as credit or employment, but does not apply to information used for marketing. Other laws apply specifically to health care providers, financial institutions, videotape service providers, or to the online collection of information about children.
>
> The current statutory framework for consumer privacy does not fully address new technologies —such as the tracking of online behavior or mobile devices—and the vastly increased marketplace for personal information, including the proliferation of information sharing among third parties. With regard to data used for marketing, no federal statute provides consumers the right to learn what information is held about them and who holds it. In many circumstances, consumers also do not have the legal right to control the collection or sharing with third parties of sensitive personal information (such as their shopping habits and health interests) for marketing purposes. As a result, although some industry participants have stated that current privacy laws are adequate—particularly in light of self-regulatory measures under way—GAO found that gaps exist in the current statutory framework for privacy. And that the framework does not fully reflect the Fair Information Practice Principles, widely accepted principles for protecting the privacy and security of personal information that have served as a basis for many of the privacy recommendations federal agencies have made.

Perhaps because the laws are in such a jumble, enforcement is feeble at best. It should be noted, however, that the European Commission takes privacy considerably more seriously than does the US government. Since digital information does not respect national borders, EU laws are beginning to impact US practices. Google, for example, ran afoul of EU privacy protection rulings on the "right to be forgotten" in 2010[79]. The results are still in contention today[80] even though Google has modified some of its practices (in the most grudging of ways).

There are thousands of data centers around the world[81], the largest of which are typically hidden away

[78] http://www.gao.gov/assets/660/658151.pdf

[79] http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

[80] http://www.reuters.com/article/2014/07/24/us-google-eu-privacy-idUSKBN0FT1AZ20140724

[81] Currently there are 3258 colocation data centers in 102 countries for use by corporate customers. There are an additional few hundred cloud server centers throughout the world. See http://www.datacentermap.com/. Some large companies with massive data usage (Google, Facebook, and Microsoft for example) each host more than a million

in nondescript warehouse buildings inside barbed-wire fences with heavy security and fed by big fiber-optic trunk lines for IP traffic, and even bigger power cables for the megawatts of power they need. A moderately large data-center may contain tens of thousands of servers and data storage units with 500 petabytes of data that can be mined in the search for important hidden patterns. Inside, they are big, noisy, dark and cold[82]. Within the most mission critical centers, called Tier-4 centers, all components are fully fault-tolerant including uplinks, storage, chillers, HVAC systems, servers etc. Everything is dual-powered. This architecture guarantees 99.995% availability[83].

Think of each of these physically isolated data centers as portions of the association cortex of a human brain, creating context associations from sensory input. They have no sensors of their own, instead their sensors are distributed all over the world in the form of mobile devices with camera's microphones, GPS, accelerometers, and compasses. Chemical sensors to provide aspects of smell and taste are already on the horizon.

Today most Big Data mining is still little more than data-base queries on steroids. Context-centric data mining is still in its infancy. But when it matures, meaning that data mining traverses context graphs instead of database tables, it will be much more capable. Will they be fast enough to keep up with sensory input? That's still a research question.

Lessons Learned from the Data-Centric Internet

- **It's a big mature ecosystem** – The top Data Barons support and are supported by an industry of specialized data gathering, storage, and analytic firms that provide functions largely invisible to the public.
- **It's about scale** – The loss of privacy happens one bit at a time. The legality of capturing each bit is murky but tends to be considered legal. The law doesn't factor in the effect of gathering data at massive scale. Gathering all possible bits from all available sources about all users of digital resources – digital credit and loans, online purchases, surfing the web, use of social media, TV watching patterns, phone conversation metadata (and perhaps content?), email contents and metadata, etc. – is qualitatively different, yet the law does not recognize effects of scale in part because politics does take account of scale and politicians write the laws. Politics itself is about scale – any one person's opinion is uninteresting but social trends matter and an important trend in reaction to Edward Snowden's revelations is forcing governments to address privacy issues.
- **Data is money** – There are many ways to monetize data. Most of them require concentrating lots of data into massive, very expensive, power-hungry data centers where exploitable patterns can be found by applying lots of CPU power to lots of data. These data centers cost $0.5 to $2 billion to build. They nonetheless can generate far more income than they cost.
- **Context distinguishes the most useful data** – You don't need to ***own*** the basic data, i.e., the identities of all the people on the planet, or all the “things” in the IoT or basic facts about the people or things. They can be from creative commons sources. If the first level facts are embedded in highly salient context graph, the graph, not the facts, provides the smarts. And if

servers. Google is reputed to have more than 2.5 million.

[82] https://www.google.com/about/careers/teams/ops-support/data-center/

[83] http://www.datacenterjournal.com/

you construct that graph, you can own it. In any case, you need to have the whole graph concentrated in a data center close to the CPU power in order to find the important connections quickly.

- **EULA-land –** people's apparent willingness to give away their data may be an illusion. They are almost completely naïve about what EULAs say or mean, and ignorant about how Google and Facebook (and many others) become fantastically wealthy by exploiting the data that users give them for free. Nonetheless, after Snowden's revelations, people seem increasingly displeased about how the data they contribute is used. Ongoing lawsuits and nascent legislation involving wiretap laws may alter that situation.
- **Profits go to the first mover –** Known as the network effect where the largest network is the most valuable to the users. Classic examples include phones, and fax machines where more people with phones or faxes make it more desirable for others to get phones or fax machines. The more modern and relevant example is Microsoft[84] where the more DOS or Windows users there were, the more software was written for those machines and thus the more value the user received by buying one. Google, Facebook and Twitter are the most recent beneficiaries of the network effect.
- **Smartphones are synergistic data providers and users –** Smartphones and other mobile devices both provide lots of data about their users and consume lots data from the big players in big data.
- **Speed is of the essence –** We want answers and we want them now! Yesterday's breathless items are of interest only to the slow, weak, and uninteresting consumers. This may be controversial since ad serving may be about slower issues, e.g., buying a new car. But the latest trending issues generate quick impulse buys that are most susceptible to a Web ad, hence serving the ads is worth more. Twitter hashtags are thus worth their weight in gold. Of course, they are bits, hence weightless). But bits are gold nonetheless.
- **Security is largely a myth –** hacking, phishing, spear-phishing and complacency... And once personal data is out of your control, it proliferates over time and **never** goes away!

[84] http://www.nytimes.com/2008/07/07/technology/07iht-07google.14282611.html?pagewanted=all&_r=0

Data Barons and the Internet of Things

The Internet of Things refers to devices that communicate wirelessly with other devices. A Primer on the Internet of Things (IoT) and RFID[85] defined the subject thus:

> "Companies and organizations explain the Internet of Things in various ways, but the Internet of Things, or IoT, is most commonly described as an ecosystem of technologies monitoring the status of physical objects, capturing meaningful data, and communicating that information through IP networks to software applications. The recurring themes in all definitions of the Internet of Things include smart objects, machine to machine communication, RF technologies, and a central hub of information."

RFID chips are the first denizens of the IoT. They are passive devices that return a fixed digital ID when queried by a wireless radio signal from a reader. That ID has to be associated in some database with whatever the RFID chip is attached to for it to have any informative value. RFID began to take off in 2003 when Walmart initiated an inventory tracking program that used RFID chips installed in most of their items for sale. Thus they could make taking inventory simple, and by tracking sales, make reordering easy. But also, since people commonly use credit cards that identify them, Walmart could get information about who bought what. More recently, the "things" participating in the IoT are "things" like doorbells, thermostats, refrigerators, lights, crockpots, stoves and ovens, and home Internet hubs that hook to cable and provide wireless Internet in the home. These two categories have very different characteristics. So we will examine them separately.

In either case – RFID and "smart things" – the IoT generates lots of data, data about every product in the consumer world, who bought it, its current location, and its status. To the corporate world, that huge trove of data represents lots of money. However, only the corporations (and governments) have the computing power to exploit all this information. Individuals are helpless *in the IoT world, because they are effectively blind, naked, and living in glass houses.*

The RFID IoT

Companies and organizations also see value in the ability to track and control activities of their employees. Therefore many corporate employee ID badges have contained RFID chips since the mid '90s. some employee uniform sellers embed RFID chips in their uniforms to allow their customers to track their employees. FedEx uses RFID wristbands that identify the driver to the truck to provide hands-free access to the truck and the ignition. However, the first large scale use of RFID was in passports. All new US Passports have contained RFID chips since August 2007.

The application that took digital dehumanization right up to Eric Schmidt's "creepy limit" was iHygene. They sold soap dispensers for company bathrooms that notify the manager if an employee fails to wash with soap before leaving. For whatever reason, the product wasn't successful.   The next step in an ID for every purpose was Italian clothier Benneton, who proposed to include RFID chips in their clothing including lingerie and underwear, with size, style and color information that could easily be read from within a meter or so, and from farther away with more specialized equipment. Benneton backed down when the plan became public.

[85] http://blog.atlasrfidstore.com/internet-of-things-and-rfid

The RFID industry is nothing if not relentless in their attempts to attach their chips to everything and everyone. The FDA has approved implanting RFID devices in people. Our livestock are already tagged, perhaps this is the next step in dehumanization: from citizens to consumers, then from consumers to livestock on the big-data farm. Pharmaceutical companies want to place RFID chips the size of a grain of sand *inside* pills! Perhaps RFID sensors inside corporate "smart toilets" could then track who is taking what pharmaceuticals and if they are taking them as prescribed. Pharmaceutical companies claim that RFID would help identify counterfeit[86] drugs that may be ineffective or dangerous by the lack of the RFID tag. But one suspects that the underlying motivation is to preserve the large price differences that the pharmaceutical companies maintain between a drug sold in the US and the same drug sold more cheaply in Canada or other countries. With RFID, reimported drugs could be identified and confiscated at the US border. The RFID could identify those who sell Canadian-priced drugs to Americans (mostly to the elderly or the poor who cannot afford the drugs at American prices) and the pharmas could punish the offending importers.

In an echo of sub-prime mortgages, there are already examples of sub-prime auto loans secured by RFID devices – lenders of sub-prime auto loans often use Internet connected ignition controllers that allow the lender to disable the car's ignition if a car payment is missed. "Beyond the ability to disable a vehicle, the devices have [GPS] tracking capabilities that allow lenders and others to know the movements of borrowers. The devices, which have been installed in about two million vehicles, are helping feed the subprime [auto loan] boom by enabling more high-risk borrowers to get loans." These auto loans are bundled and securitized and sold to insurance companies and pension funds with the justification that the technology makes the risk of default much lower.

Privacy Concerns

As early as 2003, a group of privacy advocates including the Electronic Frontier Foundation, the American Civil Liberties Union (ACLU), and the Electronic Privacy Information Center (EPIC) became concerned about the danger to privacy enabled by RFID. They published a policy guide to the dangers of RFID. Below are excerpts:

> "While there are beneficial uses of RFID, some attributes of the technology could be deployed in ways that threaten privacy and civil liberties:
>
> - **Hidden placement of tags.** RFID tags can be embedded into/onto objects and documents without the knowledge of the individual who obtains those items.
> - **Unique identifiers for all objects worldwide.** The Electronic Product Code potentially enables every object on earth to have its own unique ID. The use of unique ID numbers could lead to the creation of a global item registration system in which every physical object is identified and linked to its purchaser or owner at the point of sale or transfer.
> - **Massive data aggregation.** RFID deployment requires the creation of massive databases containing unique tag data. These records could be linked with personal identifying data, especially as computer memory and processing capacities expand.
> - **Hidden readers.** Tags can be read from a distance, not restricted to line of sight, by readers that can be incorporated invisibly into nearly any environment where human beings or items congregate. RFID readers have already been experimentally embedded

[86] Protection against "counterfeiting" seems to be a favorite stalking-horse excuse used by the RFID industry.

into floor tiles, woven into carpeting and floor mats, hidden in doorways, and seamlessly incorporated into retail shelving and counters, making it virtually impossible for a consumer to know when or if he or she was being "scanned."

- **Individual tracking and profiling.** If personal identity were linked with unique RFID tag numbers, individuals could be profiled and tracked without their knowledge or consent. For example, a tag embedded in a shoe could serve as a de facto identifier for the person wearing it. Even if item-level information remains generic, identifying items people wear or carry could associate them with, for example, particular events like political rallies.

In summary, the use and misuse scenarios of RFID tags is a quite complicated issue, especially since the legal landscape is very vague about privacy, ownership, and access to the resulting data. Countervailing technology is already being developed to block RFID. A conductive shield such as metal foil lined Tyvek will block RFID signals. An RFID chip may be neutered by strong RF pulses (e.g., 5 seconds in a microwave) that “fry” the RFID chips. And jamming signals could confuse nearby RFID readers. And one can buy or make shielded wallets for holding your RFID chipped passport and credit cards.

The IoT of “Smart” Things

Manufacturers seem to find IoT digital magic so seductive that many industries are tripping over themselves to replace analog controls in their products with digital controls and an Internet interface. Digital systems tend to be cheaper than analog systems and much sexier (within the industry). However, every device that “goes digital” brings with it the arbitrariness and susceptibility to magic inherent in digital worlds. Some of the odd places where computers lurk, and to which they can contribute their arbitrariness, include: “smart” insulin pumps worn by diabetics, “smart” heart pacemakers that can be controlled by external devices (hence disabled remotely! *… murder by IoT?*), an Internet-connected car that can have its control systems altered remotely, or an IV drip that can be remotely shut off with a click of a mouse. There is even a “smart” contact lens that senses and transmits blood sugar levels and can change its corrective power by external command. Some of these and thousands of other kinds of embedded computers will undoubtedly turn out to be susceptible to bugs or hackers.

One serious issue is: where is the demand? Techie digital addicts who want to be cooler-than-thou have bought into the IoT vision but the general public is oblivious. The players and wannabe players in Big Data are salivating over the floods of data expected to be generated by the all these Internet-connected “Smart Things”. Yet no one truly believes that the consumer will be competent, or even permitted, to control their own Things or have the right to have data about their Things expunged from all “Thing” big data farms (in a manner not unlike the EU's nascent “right to be forgotten”). By definition, that places the IoT in the category of black magic serving corporate profit maximizers. That works in the web because the web itself is seductive, even for the layman. But the IoT is not.

And imagine visiting a flea market with your hand-held RFID reader a few years hence; you will see a mind-boggling cacophony of information from used smart appliances, perhaps including the IDs of previous owners and everything the appliance has deduced about those owners. People are already very absent minded or ignorant about completely wiping data from the used smartphones they turn in when they buy new ones. I doubt they will go to the trouble of erasing the memory of their old refrigerators or other “smart” devices.

All in good fun, right? We are already used to having to reboot our computers, our phones, and our TVs, and other individual digital appliances. Recently my wife had to reboot the stove (by cutting off it's power at the breaker...the only way to do it) when it began beeping loudly and rapidly and continued to beep for many minutes with no sign that it may eventually give up. Consider the fun of rebooting your whole home and every “smart” device in it because something crashed. Will you have to “back up” the whole mess to restore it to sanity after it is all rebooted or when you move to a new home? Perhaps it isn't all in such good fun after all? White magic today may turn into black magic tomorrow. The opposite transformation is far less likely.

And there is the unexamined issue of identity theft in a world full of “Smart Things”. Today it is all too common for people to be forced to hassle for months with the banks and credit agencies when their credit record is compromised by identity theft (or just by someone's mistake). Tomorrow the same may be true for your whole life. The EU's recent ruling that Google must (at least sometimes) honor your “right to be forgotten” is just the smallest step toward dealing with life-hacking. Legal systems have yet to satisfactorily address the corporate “rights” to keep and sell information about you that they have surreptitiously gathered under cover of impenetrable EULAS or, in the case of RFID simply by fiat.

Might we find disquieting parallels to the failures of assumptions and outright bugs in the software that bankrupted more than a few investors in the housing crash. Similar problems in the IoT may kill people. Will a future Google automated car with no steering wheel notice when the driver suffers a heart attack or stroke and change course to deliver him to the nearest emergency room? Or will it “cheerfully” deliver his body to his programmed destination?

The awkward thing about code is that it mindlessly does what it was programmed to do, where “programmed” includes all misconstrued requirements, accidental bugs and faulty assumptions about the behavior of people or other collaborating code. And the companies who nominally control the programmers that program the devices aren't content to simply sell you the “smart” device. They also want to sell the data about you that they can program the device to collect and forward to them.

Last but not least, how are software updates and bug-fixes supposed to make their way into the millions of embedded computers of the IoT? See the Wired article “ The Internet of Things Is Wildly Insecure — And Often Unpatchable”. That article points out, among many sobering points, that

> “...no one entity has any incentive, expertise, or even ability to patch the [embedded] software once it’s shipped. The chip manufacturer is busy shipping the next version of the chip, and the ODM [original device manufacturer] is busy upgrading its product to work with this next chip. Maintaining the older chips and products just isn’t a priority.
>
> And the software is old, even when the device is new. For example, one survey of common home routers found that the software components were four to five years older than the device. The minimum age of the Linux operating system was four years. The minimum age of the Samba file system software: six years. They may have had all the security patches applied, but most likely not. No one has that job. Some of the components are so old that they’re no longer being patched. This patching is especially important because security vulnerabilities are found more easily as systems age.”

In September, 2014 warnings such as these came true with the discovery of the 'Shellshock' vulnerability in the Bash shell. “...the Bash shell is the default command interpreter on most Unix and Linux systems and all Macs. A more serious problem is faced by devices that use embedded Linux – such as routers, switches, and appliances. If you're running an older, no longer supported model, it may be close to impossible to patch it and will likely be vulnerable to attacks. If that's the case, you should replace as soon as possible.” Within a day of the vulnerability being made public, it was being exploited “in the wild”.

IoT Botnets

Botnets are wrecking havoc in the Internet. And the primary enabler of botnets is the Internet of Things (IoT). As the old joke goes: in the Internet, no one knows you're a dog (or a bot). A bot is a computer on the Internet that has been taken over, at least in part, by some malicious hacker who has installed some of his own software on that computer. Once infected, a bot can operate under remote direction unnoticed by its nominal owner. A botnet is a whole network of bots under the control of one malicious actor. The agent that controls the botnet is known as a "bot herder".

The IoT has dramatically boosted the number and size of botnets because billions of IoT devices are sold and installed each year. They are typically invisible to their nominal owners because they are hidden inside "smart" doorlocks, home routers, cameras, and hundreds of other devices. Even if a consumer knows that some sort of tiny computer is present in something they purchase, few of them are aware of how easily IoT devices can be compromised or how to prevent them from being compromised. And most manufacturers who install the computers in "smart" devices take minimal responsibility for making them difficult to hack or for forcing consumers to change IoT devices' passwords to sufficiently secure new passwords. By this abdication of responsibility, the hardware industry has created a monster.

Bot herders' most visible exploitation of IoT devices is using them for distributed denial-of-service (DDoS) attacks, either directly from the bot herders themselves, or "rented out" to wannabe attackers. In 2016, massive DDoS attacks were launched by a botnet named "Mirai" that seeks out and enrolls poorly-secured IoT devices such as security cameras, digital video recorders and Internet routers. A derivative of the Mirai botnet then blackmailed at least three large financial institutions. Another troublesome botnet, named "wireX" was created for Android devices. WireX first appeared August 2, 2017. Hacked Android devices conducted some relatively small online attacks. "Less than two weeks later, however, the number of infected Android devices enslaved by WireX had ballooned to the tens of thousands." Several large industry players including Google, Akamai, Cloudflare, and Flashpoint, quickly combined forces to take it down.

Botnets tend to be seen only as providers of massive, and free, Internet access, e.g., for DDoS attacks. What is often ignored is that bots, especially IoT bots, spend almost all of their cycles in idle loops. Those idle cycles can be used by botnets to mine cryptocurrencies or train Machine Learning systems. System on a Chip (SOC) bots often have full Linux operating systems, including SSD memory that could contribute free distributed storage as well. And the largest cost in cryptomining is the cost of the electricity to power the mining computers. Using IoT bots makes the owners of the IoT devices pay the electric bills. The situation with Machine Learning is similar.

IoT devices are being inserted willy-nilly into hundreds of different products. Relatively few purchasers are aware that the products contain a general-purpose computer that can easily be hacked let alone how to secure the devices. Home hubs and routers should, in any case, insulate them from the Internet at large. Kaspersky labs has begun looking at such issues in the IoT. Their findings are startling, especially given Kaspersky's relationship with Russian Intelligence. What they are finding about American IoT vulnerabilities would be quite useful in a Russian cyber-attack on US elections, electric power infrastructure, or commerce. And in the 2016 Presidential election, that possibility became a reality.

Best practices for end users to reduce the risk of being recruited into a botnet include: regularly updating devices with the latest firmware, changing devices' default credentials, using intrusion detection and prevention systems, and being wary of known attack vectors, such as unsolicited emails. Needless to say, naive digitally illiterate consumers do not follow such guidelines. Therein lies an interesting opportunity for makers of IoT devices -- IoT as a service! Manufacturers of IoT devices could take on the responsibility for keeping the devices safe, and in return, use the idle time of the devices for money-making computing services such as cryptocurrency mining. And they could rebate to the consumers a portion of the profits. This business model would benefit everyone. The "first movers" into the business model might very well become the giant corporations of the future cyber world. A similar business model is already provided by WinMiner. To play the role proposed here, WinMiner would only need to take logical custody of your IoT devices via the Internet, manage the Internet security of those devices, and ensure that they play their designated IoT role when needed.

Discussion

Cyberspace began in 1989 for the purpose of sharing academic research. There were no commercial applications, nor was there even thought of such applications for several years. Growth of the web and of its users was organic, driven by the fascination of its users. People built and browsed web pages for personal reasons. Even Google began because its founders were fascinated by the challenge of users to find web pages to suit their interest among tens of thousands of pages irrelevant to the person searching. There were no ads in cyberspace and for the most part, commercial companies were slow to let the Internet inside their walls. Cyberspace, such as it was, lived primarily in universities.

As web-savvy university graduates were hired by businesses, they complained to their employer about their inability to browse the web. One company after another grudgingly relented – and discovered the seductiveness of the web. To "knowledge workers" the web was addictive. Corporate minds began to wake up. After all, anything that seductive is bound to have commercial value.

Data Barons and the Economics of the Digital World

The earliest efforts to monetize the web attempted to get people to subscribe to or otherwise pay directly for web content. Almost all such attempts failed. But advertisers were used to paying for eyeballs in other media and so it wasn't a big stretch for them to pay to put ads on websites. The first click-through web ad appeared in 1996[87]. A business model was born. Google was incorporated in 1998. As a privately held company, it did not disclose earnings. The world did not recognize how lucrative web advertising could be until Google's IPO in 2004. The realization that Google was highly profitable triggered the commercial gold rush that has since transformed cyberspace, and the modern world.

The commercial dynamic was further transformed by Google's demonstration that huge data farms and data mining can convert masses of data into great wealth. But such wealth takes very large investments in gathering the data, building the data farms and providing the megawatts to power them, not to mention developing the software to extract the value. One result of this transformation was that the capital required to build and run data farms cemented cyberspace to the techniques and goals of large corporate players at the expense of individual concerns such as personal privacy.

The technology required by data farms in turn depends the economics of the huge factories that build the various integrated circuits required by the CPUs, data storage devices and the many types of support chips. These factories are known in the trade as chip fabs. They too cost upwards of a billion dollars. And they take many years of running full tilt to recover the cost of their construction. So, here again, we see economics that inherently demand very large corporations.

The "black magic" alliance between by Big Data and huge semiconductor fabs has a positive feedback component as well. Advances in data mining software require ever more compute power and ever larger storage capabilities. The constant advances of Moore's Law that drive requirements for chip fabs always hold the promise of just such increasing compute power and storage capacity at the next smaller increment of line width. The IEEE Spectrum article on the economics of semiconductor manufacturing

[87] Briggs, Rex; Hollis, Nigel, Advertising on the Web: Is there Response Before Clickthrough? Journal of Advertising Research, March–April 1997, pg 33-45

explains:

> “More than anything else, Moore's Law has been responsible for the gigantic costs. It takes huge amounts of capital to support the incessant cycles of investment and obsolescence that keep Moore's Law on the march. That rapid cycling explains why a company's shining jewels can turn into white elephants in just five years.”

The economics that drive both semiconductor technology and Big Data farms preclude any return to the the “wild west” days of the early web. “Big Business” agendas are in the saddle. Quaint concerns about individual privacy get short shrift in American, Chinese, Korean, and Tiawanese executive boardrooms. Only Germany, with its much more recent experience of intrusive surveillance by the East German Stassi stands in the way of increasing focus on surveillance.

However, one might ask how long the rate of increase of data from human activity on the web will continue to grow. …

Gartner (January 2014) estimates that "Samsung Electronics and Apple have topped the semiconductor consumption table for three years running, with their share of the design total available market (TAM) rising from 12 per cent in 2011, to 17 per cent in 2013". That is due to the popularity of smartphones But that growth is slowing. Gartner asserts that: "Wearable devices as well as 'Internet of Things' (IoT) devices are the next growth drivers for hardware vendors that can offer the required energy saving technologies and energy harvesting technologies to realise these applications,"

The corporate agendas for mining money from users of the web require new sources of data and the IoT seems to be the potential new source of data. But the IoT does not seem seductive or magical, just intrusive and creepy. Corporate efforts to convince us that RFID will make our lives wonderful seem to fall flat. Most people other than the surveillance community are at best little interested in ever present RFID and at worst put off by it. Thus conflict looms and it will be fought out in the political realm where the money from corporate lobbiests will push for more and more access to data while the public pushes for some shreds of rights to privacy.

Billion dollar data centers are one end product of cheap hardware. But when you have Big data centers (as Google, Microsoft, Facebook, Apple, IBM and many others do), the economics of scale require endless new sources of data. Both more and more powerful smartphones and the Internet of Things promise to proved almost endless new data to store and mine in the data centers which are enabled by endless new chips with new functions which are required to make the economics of building chips work. And all this new hardware provides a fertile field for more and more software magic. A classic positive-feedback emergent phenomenon.

RFID space continues to expand as well. Stanford University and UC Berkeley researchers have created prototype ant-sized radio-on-a-chip devices powered by ambient radio waves. “Comprising receiving and transmitting antennas and a central processor, the completely self-contained ant-sized devices are very cheap to manufacture, don't require batteries to run and could give the 'Internet of Things' (IoT) a serious kick start.” They envision trillions of these devices connected together and to the Internet. What will they do? And say to each other? That is not the concern of the hardware researches.

The Evolution of Magic in the Digital World

Humans love magic and have wished for a long time to escape from the constraints of the real world. We would rather be able to soar through the air with the greatest of ease, to be forever young, to transmute lead into gold, to teleport, or read minds. Or, since wizards are seen as cool and powerful, at the very least we can do is to festoon ourselves with magical devices (wearable tech) to give the impression we are like the wizards. People wear sports jerseys or dress like the stars as style statements.

Innovations in the digital world are further complicating the interactions between physical and virtual/digital worlds. The nascent Internet of Things already has added "real world" effects into otherwise virtual worlds, and vice versa. A well organized hacking group appears to be targeting industrial control software (e.g., power plants) with the probable intent of being able to shut them down. Social networking brings digital actions into the social/cultural real world where they can be seen and manipulated (sharing a news feed with a few Facebook friends for example). Researchers at Facebook report that "For one week in 2012, Facebook skewed nearly 700,000 users' news feeds to either be happier or sadder than normal. … after the experiment was over users' tended to post positive or negative comments according to the skew that was given to their newsfeed."[88] And 3-D printing (or additive manufacturing) which turns digital models of 3-D objects into real-world objects has generated the kind of seductiveness and enthusiasm that the web did. The variety of objects, tools, products, and fanciful artworks that can be built with 3-D printing is already large, growing rapidly, and many of the models are open-source. For example, Makerbot has open source models for prosthetic hands that can be customized and built on relatively cheap (about $2000) MakerBot 2 Replicator (see example here).

The digital world of social networking also spills out into public consequences via the temptation to seek short term digital "fame" in the twittersphere by opining on some popular hashtag. There are long-term consequences of the permanence of such communications and the ultimate access by corporations and governments to intrude into or exploit what seem to be private actions and communications, but are not. Edward Snowden has made visible many surveillance tactics of governments. Google, Microsoft, Amazon, and a host of others use data they collect or buy from data vendors. But casual Twitter users act as if their tweets are accessible only to those who follow them whereas in fact, all tweets not explicitly made private are fully public. "They can be read almost instantly by anyone with an internet connection on the planet Earth. This is not a bug in Twitter; it is a feature. Twitter is a thing that allows you to publish things, quickly, to the public." They not only can be read as soon as they are sent, but also a decade later. Everything most people post on Twitter is both public and permanent.

Out of Control Wizards

Digital magic in the hands of digital wizards affects the physical world (unlike the spells and gestures of medieval wizards). And muggles seem already to prefer superstition to the effort of learning and understanding the lawfulness of the real world. As magic has expanded in the world, disdain for

[88] "Experimental evidence of massive-scale emotional contagion through social networks". Adam D. I. Kramer, Jamie E. Guillory, & Jeffrey T. Hancock, Proc. Natl. Academy of Science, vol. 111 no. 24, June 17th, 2014. [I can't help but comment here that Orwell's 1984 is only thirty years late. Now we learn that Big Brother's name is Zuckerberg.]

science has grown, especially in the US where the Internet was birthed. The notion that data is good – all data, the more the better – has become the wave of the future. Why? Because with enough data presented to us in the right way, we are like putty in the hands of those who can put carefully crafted commercial or political messages before our eyes. Scientific marketing via the medium of TV ads became so effective in telling us what to buy or who to vote for that the candidate or car or breakfast cereal with the best marketing and the most money for ads was most likely to win.

However, on the Web, audiences are segmented more finely and so much more is known about each and every web viewer that Web advertising can be targeted more accurately. We now are headed toward being viewed as little more than clusters of data-points in Big Data farms. What will be the consequences of that viewpoint for our society? What will it mean to be human? It has already been noted that social media such as Facebook and Twitter are addictive and, as with all addictions, they can turn sour, leaving the addicts depressed and adrift. Have we overestimated the Digital World as a good place to inhabit in the long term? And what choice do we have?

It is also worth considering the degree to which the evolution of digital money provides lessons for us all. To the extent that varieties of Big Data are the focus of much of Corporate America these days, we can confidently predict that there will be a boom and bust in Big Data exploitation. Can we go further to get any insight into which of the assumptions behind this boom will turn out to be the flaws that lead to the bust? And what of monopoly power? Google already acts monopolistically, as does Facebook and Amazon. One wild card danger to their monopolies is that their unholy alliance with NSA and the British GCHQ has generates increasing distrust in US allies around the world, and increasingly within the US as well. Will governments' insatiable lust for data, data, and more data kill the golden goose? Or will the flood of new data in Internet of Things do it for them?

When a person pulls out the smartphone, how many lurking big data gatherers will be watching? And how many times can each person (or consumer) be sold? When some Big Data company such as Axciom and Rocket Fuel merges enough “views” of people into one large context graph, what's left for others to do? Will the devil take the hindmost? Remembering the dot.com market bust, will there be a Big Data bust?

Digital Magic in Society

As people embrace digital magic, do they lose touch (literally) with the physical world?

We have discussed the impact of digital data predators on individuals (loss of privacy and anonymity, subjugation to corporate agendas, etc.). But what about the other way around, or better put, the positive feedback of bi-directional influence. Social Psychologists have found that people absorb a certain persona appropriate to their experiences. Experiences that we choose in the digital world can have impacts as real as those we choose in the physical and social worlds. Wealth gained by exploiting tricks of digital finance such as high frequency trading is wealth nonetheless. Not only does the wealthy lifestyle it allows change those who live it, but the wonder of becoming wildly rich by use of magic may be even more powerful than working in manufacturing or commerce. Mastery of some fantasy first person shooter game is mastery nonetheless. “Fame” and “influence” gained by having lots of followers of our tweets is fame and influence nonetheless. Feelings of inferiority and depression because everyone else's Facebook page seems to us to be cooler than our own, is inferiority nonetheless. And it seems almost certain that we absorb those feelings and they are reflected in our personal lives. Some studies show that Facebook and Instagram engender depression, loneliness, and low self esteem due a bias people have to burnish their image on their postings. One insightful article, titled "I Facebook therefore I am" discusses the positive feedback loop as people exaggerate their life to compete with the exaggerations of their "friends" which becomes a kind of self image arms race.

Human societies, and languages, are adapted to thousands of years of experience, mostly superstitiously driven, nonetheless well adapted. We do not know and cannot know which portions of that adaptation can be rapidly changed without serious consequences and which cannot. We are getting some hints, though, as so many people become immersed in Cyberspace.

- **Reading Human Emotion –** "UCLA scientists found that sixth-graders who went five days without even glancing at a smartphone, television or other digital screen did substantially better at reading human emotions than sixth-graders from the same school who continued to spend hours each day looking at their electronic devices."
- **Social Isolation –** The more Facebook and Twitter, the less personal interaction? We see people in restaurants sitting at the table with another person where both ignore each other and focus on their smartphones. But also, the comments about the danger of texting or interacting with their smartphone while walking were focused on the external consequences of that behavior. What isn't as obvious is that many of those walkers also have earbuds in their ears. They are isolating themselves from almost all external physical world and social interaction in favor of their choice of digital interaction. That isolation has consequences itself, whether or not the digital addicts are in physical danger.
- **Depression and Low Self Esteem –** The more Facebook, the lower self esteem. Social media such as Facebook can magnify negative psychological problems. A large Swedish study (1000 subjects) shows that low educated groups and low income groups who spend more time on Facebook also report feeling less happy and less content with their lives. It also shows the following: the average user spends 75 minutes per day on Facebook, and logs on to Facebook 6.1 times per day, 70 percent log in every time they start their computer or web reader, and 26 percent feel ill at ease if they do not get to log in regularly.
- **Game Avatars –** Games with avatars have social impacts. Psychological research shows that the choice of avatar affects how people behave in the digital game world and to some extent outside it.
- **Smartphone Addiction –** Smartphones themselves are addictive. One study shows that the more college students used smartphones, the more anxious they became when they were prevented from using them. Interestingly the study also showed that those who were allowed to keep possession of the phones but keep them turned off were not as anxious as those who were made to surrender the phones. Thus mere possession of the phone allayed some of the anxiety! But perhaps more importantly, smartphones enable or exacerbate other addictions because someone who shares your addiction is only a text message, tweet or phone call away..
- **Social Balkanization –** With the help of Google, one can make contact with kindred spirits no matter how unusual the worldview. Every political, cultural, religious, economic, hobby, sport, craft, and sexual preference can become a nucleus around which like-minded people gather. Virtual discussion becomes easier because you can ensure that your virtual "friends" will share your views. That tends to reinforce people's views. Thus there is less honest casual real-world

> conversation because the likelihood that someone nearby will share your true views decreases. A balkanized world is a polarized world. Politics becomes a “discussion” between people who already agree with each other that features rants against all those evil folks who disagree. Those interested in politics fragment along an increasing number of fault-lines. Gardening groups fragment according to whether they prefer growing orchids or marijuana. Sports fans can ignore those pesky fans of the wrong team or the wrong sport, or the wrong player. Thus we withdraw further into our specialized digital worlds and barely speak to those who have different views.

Morality, Privacy, and Insecurity

All bits are not created equal and just because they can be collected does not give the collector the rights to use them in arbitrary (hence magical) ways. But EULAs pretend otherwise, and both society and government are asleep at the switch. It is not clear that legal spells can corral digital spells, especially given that some digital spells (code for example) are given the protection of legally supported secrecy for “intellectual property” (IP) reasons.

The physical world is becoming confounded with IP claims as well. All sorts of information is increasing vulnerability to having claims staked in this gold rush. The stakes potentially can be claimed by cameras (including infrared cameras), microphones, accelerometers, radio signals (RFID and cell phone SS7 signals), chemical sensors, vibration and presure sensors, and many other sensors and actuators. Google Glass is only the most visible issue today. CCTV surveillance perhaps ought to be considered a bigger issue, but most of us aren't conscious of its pervasiveness and power. Google Glass is more unusual, more visible in social situations, and more threatening, not to mention just weird! It makes more clear and obvious the violation of our social expectations.

Legal casuistry claims that digital information in your social media communications is like any other publicly available information, but there is no other publicly available information even remotely like it, certainly not in breadth and volume. Whatever the legality, the morality of collecting any and all “public” information about people is questionable.

Spyware can turn on the computer's video camera or microphone, steal passwords and credit card numbers, monitor the GPS or accelerometer of a smartphone to see where you are or what you are doing (including what you are typing), cause your smartphone to call pay numbers, etc. Moreover, spyware or other viruses, worms, etc., may be installed on your computer from anywhere on the planet and monitored or exploited similarly. Rootkits may be installed into your boot ROM so that they are essentially undetectable by the computer itself. Detection requires searching the ROM before boot and even if found, some are almost impossible to remove.

Microsoft, Apple and Linux operating systems contain mechanisms for semi-automatically installing software “updates”, e.g., Microsoft Update. Some of that software reports some of your behavior back to the manufacturer. Even if the company is trustworthy, others can exploit the update software. The infamous Flame virus can infect even secure PCs by tricking them into believing its malicious payload is actually an update from Microsoft.

Microsoft's Windows 10 Preview EULA grants permission to watch your every move. Some excerpts:

> *"Microsoft collects information about you, your devices, applications and networks, and your*

> *use of those devices, applications and networks.* ***Examples of*** *data we collect include your name, email address, preferences and interests; browsing, search and file history; phone call and SMS data; device configuration and sensor data; and application usage."*
>
> *"We may collect information about your device and applications and use it for purposes* ***such as*** *determining or improving compatibility" and "use voice input features like speech-to-text, we may collect voice information and use it for purposes* ***such as*** *improving speech processing."*
>
> *"If you open a file, we may collect information about the file, the application used to open the file, and how long it takes any use [of] it for purposes* ***such as*** *improving performance, or [if you] enter text, we may collect typed characters, we may collect typed characters and use them for purposes* ***such as*** *improving autocomplete and spellcheck features."*

Note that phrases in bold seem to make their data collection benign, but those phrases do not actually limit Microsoft to those purposes; any and all other purposed are fair game.

Conclusions

Arthur C. Clarke's Third Law: Any sufficiently advanced technology is indistinguishable from magic. And in the early days of writing, anything involving literacy was considered magic hence the notion that spells and runes and special marks on magical circles, etc., were seen as magic by the illiterate (which made up a huge proportion of the population for centuries). The human experience with the new technologies of writing (cuniform marks in clay, pens and inks and papyrus and paper) should show that such new technologies impact the society over the long term and outright resistance is futile. Yet as monetizing all available digital data is claiming more victims and the Internet of Things is fundamentally foreign to most of us. Many in the computing industry are aware of these issues but they are insiders to the system and hope to capture some of the crumbs of the monetization windfalls themselves, so they find it difficult to acknowledge that the dumb bit gathering machinery is black magic, not white.

We face a willful effort to ignore or gloss over the degree to which digital magic already impacts our lives and our culture. And anyone who looks objectively at the challenges of taming the Internet of Things before it ensnares us should be willing to admit that we do not know how to tame it.

It would be comforting to think that even magic can be countered by human law, but our experiences with EULAs aren't very positive examples of such an ability. Society is not unfamiliar with acts that are done just because they can be but are nonetheless against the laws of the society. Yet in the case of who owns your data, society so far has wimped out. It is widely assumed in the computing community that any bits you can get hold of without outright armed robbery can be used in any way one wants. That is clearly false. Bits, no matter how obtained, cannot legally be used against a person for straightforward extortion or blackmail. Yet we pretend that those bits in some Big Data store can be “anonymized” and used for any purpose whatever even though with enough other big data they can easily be de-anonymized. As long as we reason with such blinders on, we cannot expect to be happy with the results.

We must work to teach the next generation how to avoid being overwhelmed, both morally and in everyday practice. How should we engage in the digital world without becoming androids or bots ourselves?

JR Hennessey's column in The Guardian (July 21, 2014) ends with:

> “A simple fact remains: there is something intrinsically repellent about a world in which our food, jobs and personal relationships are replaced by digital proxies in the name of ultra-efficient disruption. The geeks, with their ready willingness to abandon social norms, are pulling us toward a utopia nobody wants.”

That captures only some of the dynamics at play. While the geeks may want a geeky utopia that no one else wants, they are successful because many of us *do* want magic: not only the magic that Silicon valley offers, but also what Digital Wall Street offers (high-tech gambling). And proponents of the Internet of Things offer a digital our “things” take care of all our physical wants. As Harry Potter showed us, magic is seductive.

The Data Barons at huge companies like Google, Facebook, Acxiom, and RocketFuel, work to extend their monopolies of data, our data, that they claim to own or control with at best questionable legal basis (vague and impenetrable EULAs) and at worst outright stealth and theft (e.g., pixel tags hidden on Web pages to track us in cyberspace, or RFID tags hidden in all manner of consumer goods such as clothing, food containers, pharmaceutical containers...or the pills themselves) to track everything we do when we aren't in cyberspace. But muggles and not a few wizards know little about those issues and seem to passively accept them uncritically because they cannot foresee the long-term consequences.

Each major new technology takes mankind a long time to digest and master, and "master" may be too strong a word. We have had centuries to "master" guns, and wars still happen. We seem to have mastered ships and cars and jets, yet thousands die each year in transportation accidents. And recently a very new Boeing 737 Max crashed in Indonesia due to a software flaw killing all aboard.

There is no reason to suppose that the Data Barons will go quietly into the night. But we cannot and will not forego the new capabilities offered by digital computing. And in the wings is Artificial General Intelligence. It will take many years to bear fruit, but it also could enslave us if we let it.

Appendix: Civil Law and “Ownership” in Cyberspace

Digital magic obeys the laws of Physics to the extent they apply, but not the laws of man. Transmission of bits is constrained by the speed of light and of the somewhat slower speed of electrical signals in conductors. Constraints due to the electronic limits of storage and computational speed exist as well[89] But we are so far away from those limits that they are largely irrelevant today. The physics of digital input and output devices continues to advance: input devices, screens, microphones, speakers, GPS receivers, accelerometers, radio senders and receivers, chemical sensors and many other devices become more capable, sensitive and precise each year. Yet what matters most in developing digital magic are the computational algorithms and heuristics that make use of data and input/output (I/O) devices. Those capabilities are effectively limited only by the imaginations of programmers and hardware engineers.

Human law is as much magic as is digital magic, albeit of a different sort. The law is made of words, not bits, but laws are as arbitrary as bits. For example, in England, it is illegal to die in the Houses of Parliament, the head of any dead whale found on the British Coast automatically becomes the property of the King and the tail goes to the Queen, and it is Illegal to Consume Mince Pie on Christmas[90]. In Florida, unmarried women who parachute on Sundays can be jailed. In Vermont, women must obtain written permission from their husbands to wear false teeth. In Providence, Rhode Island, it is still illegal for shop owners to sell toothpaste and toothbrushes to the same customer on a Sunday. In Alexandria, Minnesota, it is still illegal for a man who has garlic, onions or sardines on his breath to have sex with his wife. And so forth.

History of Civil Law in Cyberspace

Man-made laws intended to constrain what is legal in Cyberspace met their match long ago. Lawrence Lessig, currently a Harvard Law professor, published a landmark book in 1999 titled “Code and Other Laws of Cyberspace[91]” in which he pointed out that the Internet allows one to easily ignore Copyright law by simply sending copies of digital materials via email to others. But the Internet was young then[92], computers were bulky and slow, the Web was rather static and primitive and Google had just 8 employees[93]. Now, almost twenty years after Lessig published that book, “Cyberspace” is far larger, far more diverse and dynamic, and far more thoroughly woven into the physical and social worlds. In addition to Google, untold other Data Barons or wannabes traverse the entire accessible Web, copying it, analyzing it, and monetizing it *without any regard for any civil Law*. But Larry Lessig points out that: “*A culture without property, or in which creators can't get paid, is anarchy, not freedom.”*

Intellectual property law covering copyright, patents, and trademarks became entwined with programming constructs almost as soon as software became salable[94]. Enforcement of such laws has

[89] E.g., the limits of Moore's Law, which continue to recede as new techniques are found.

[90] See http://list25.com/the-25-craziest-laws-ever/

[91] See summary at http://code-is-law.org/magnant_sum.html

[92] It became available to commercial users in 1995, See http://en.wikipedia.org/wiki/History_of_the_Internet

[93] http://www.google.com/about/company/history/

[94] http://digital-law-online.info/lpdi1.0/treatise17.html

been difficult from the beginning when shrinkwrap software came on sealed floppy disks. Enforcement becomes even more problematic when the end user license agreement is a “click through” agreement[95]. The result is that intellectual property law has a very uncomfortable relationship with what is now called cyberspace. In cyberspace, code does what it does without regard for what laws say it may or may not do, and typically what it does is obscure. Companies such as Google and Facebook and various government sponsored groups collect and save data on all interactions with users without regard for the legality of doing so and without any possibility of users knowing what is being collected about them, how long such data is kept (mostly forever) and what the groups involved do with such data (essentially whatever they want). Today, cyberspace is like the American Wild West; there is a polite pretense that laws keep things under control, at least when the Sheriff is nearby. Worse, in the case of Cyber Warfare, there is no sheriff. A few nation-states are already presumed to have the ability to disable nation-wide power grids and other services vital to nations[96]. And in the next wave of cyberspace featuring ever present surveillance and the Internet of Things, our lives seemingly will be tracked and measured without even a wink and a nod toward laws or our rights to privacy. *If we let it be so.*

Every nook and cranny of the digital world is governed by the code that manages that nook or cranny. Operating systems such as Windows, MacOS and Linux typically involve tens of millions of lines of code, browsers over 10 million, and the size of the source code for Facebook is comparable to that of Windows[97]. There are billions of lines of html and javascript code operating in the estimated 5 billion web pages in the world[98]. There are many more lines in the code that runs all the web-crawlers, web-servers, and commerce sites such as eBay and Amazon, the browser clients, the mobile apps in smart phones, laptops, and the cloud servers such as Amazon's ec2. And we mustn't forget the code in the hidden malware or spyware crafted by cyber-criminals and governments, most notably in China, Russia, the US, and the UK. Whenever each of those many billions of lines of code is executed, it is a law unto itself in its particular niche[99]. Nearly all of these billions of “laws” expressed in code are invisible to all but a tiny minority of humans who in theory have both the access to the code and the skill to read and understand it in the context of today’s Internet. However it would take many lifetimes for such a wizard to even read, let alone understand a billion lines of code. Thus I would update Lessig's assertion to say: in practice, *there is NO LAW in Cyberspace*. Cyberspace is essentially an arbitrary place much like Merlin's, Gandalf's, or Dumbledore's prescientific world of magic where the right incantation could accomplish anything at all – including things never before imagined.

Individual hackers and spammers exploit the impotence of the law. Data Barons take even larger liberties with with the Law. As Snowden's disclosures have shown, the NSA and the British GCHQ act as though they are completely above the law[100] The World Wide Web Foundation, founded by Web

[95] http://en.wikipedia.org/wiki/End-user_license_agreement

[96] http://www.offthegridnews.com/2014/11/21/is-china-preparing-to-take-down-the-u-s-power-grid/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+offthegridnewscom+%28Off+The+Grid+News%29

[97] http://www.wired.com/2013/04/facebook-windows/

[98] http://www.worldwidewebsize.com/

[99] It is perhaps an exaggeration to say every *line of code* is a law unto itself. The number of lines of code is a very rough measure. To be a bit more precise, every machine language conditional branch is a law unto itself, as is every store into a variable that is anywhere tested to determine a conditional branch. That includes each external event caught by the operating system, each IP routing choice made in the Internet or LAN, etc.

[100] http://opencanada.org/features/the-think-tank/essays/snowden-the-state-and-the-future-of-internet-governance/

inventor Tim Berners-Lee, said the alleged hacking by the National Security Agency and its British counterpart, GCHQ, was "another worrying sign that these agencies think they are above the law."[101] On the rare occasions when the law does threaten their operations, they browbeat Congress and/or Parliment into changing the law in their favor. “Put simply, for governments, which are primarily charged with protecting the security of their populations, the ends (improved security) justify the means (online surveillance)”[102]

This paper discusses many of these issues but focuses primarily upon the commercial exploitation of the extremely weak constraints placed upon them by the Law. The reason for that focus is that the growing class of Data Barons has become so wealthy, hence powerful, that their lobbying efforts now influence strongly the Law for future generations.

Legal Chicanery: EULAs, “Privacy Policies” and Wiretapping

Bill Gates championed the notion of copyrighting binary software and protecting it via a “shrink-wrap“ license[103], At that time, in the 1980s, the newly invented theory that copyright applied to binary software and that shrink-wrap licenses were enforceable was wholly untested. Now, the descendants of such licenses, known as End User License Agreements (or EULAs), are ever present and considered by some to be binding even for websites (e.g., Facebook's) that provide services based upon data contributed by its users. Relying upon their wealth, legal, and lobbying power, they assert (couched in impenetrable legalese) that if you use their website, you agree to permit whatever they wish to do with the data they collect from you. The legal basis is sketchy but who wants to fight it out in court with a behemoth like Google or Facebook. What EULAs are really about is protecting the company's “right” to collect and keep forever whatever information you send their way and to use if for whatever purposes they find lucrative.

The combination of digital magic with the legal magic is synergistic. Very few technologists understand legal spells and very few lawyers understand digital magic. If you pay an attorney to vet a contract someone is asking you to sign, the attorney will (or should) make it very clear that non lawyers are not competent to assess the meaning of the contract because of the prevalence of what are called “terms of art.” Such terms may (and often do) mean the opposite of what non-lawyers would reasonably think they mean. Yet, with a straight face, they assert that it is responsibility of the computer user to read and divine the magical meaning of long and involved EULAs. In private moments these lawyers will admit that most EULAs are virtually impossible for the user to interpret because these disclaimers are about magical digital functions. The lawyers will admit that even they are incapable of understanding the issues. At a 2007 American Bar Association meeting[104],

> “...attended by prominent intellectual property lawyers and law professors, a loaded question was posed to the audience: 'By a show of hands — and be honest, now — how many of you read the terms and conditions presented in an end-user license agreement? Of the nearly 100 people in the auditorium, not a single hand was raised.”

More importantly, and with even less legitimacy, these legal disclaimers are taken to apply to

[101] http://www.philly.com/philly/news/nation_world/20150220_ap_77153be38af84c2fb005a518bec91547.html
[102] http://opencanada.org/features/the-think-tank/essays/snowden-the-state-and-the-future-of-internet-governance/
[103] http://www.computerhistory.org/atchm/microsoft-ms-dos-early-source-code/
[104] http://apps.americanbar.org/buslaw/blt/2007-01-02/kahana.shtml

unspecified deep magic technologies e.g., face recognition, geotagging and big data analytics that often are provided by third party companies (to be discussed below). Lawyers and the general public have virtually no understanding of what these technologies might be or do. Even if we were all minor wizards who did know something about these unspecified technologies today, the data lives forever and no one can know what violations of expected privacy future technologies might enable.

Google's “Privacy” policy[105], which is one of the more readable ones, defines some obscure terms, one being “pixel tags”. The Google definition reads: “*A pixel tag is a type of technology placed on a website or within the body of an email for the purpose of tracking activity on websites, or when emails are opened or accessed, and is often used in combination with cookies.*” The central obfuscation here is the phrase “tracking activity on websites.” Unless you already know what pixel tags are and do, this gobbledygook is meaningless. Wikipedia provides a more complete definition[106] that sounds considerably less benign, especially for the case of email. What such magical incantations actually do, how they do it, and what business models they enable are in general not understandable to non-wizards. The fact is that “Privacy policies” are not designed to protect your privacy, they protect the wizard's company from any liability for subverting your privacy, and to give a fig leaf of legal cover to the the partners in magic that derive value from the data collected by the EULA-protected first-level magic companies.

Those who *send email* to people with gmail, hotmail or yahoo freemail addresses are not even given the opportunity to assent to an unintelligible EULA. The freemail vendor simply makes whatever use the email they wish. Email metadata (the email addresses in the to:, cc:, and bcc: lines provide valuable social graph information that the sender may consider private. The title line provides a user generated “purpose” or abstract of the body contents that provides valuable hints for text mining. Attachments may contain private photos or documents (e.g., contracts, financial information, legal documents such as real-estate transfer papers, divorce papers, subpoenas, and the like) that obviously are, or should be private. No matter, they are all considered to be fodder for the Data Barons to mine.

Google, by far the largest email provider, was sued in 2013 for data mining students' emails[107]. The suit was based on the claim that data mining gmail violated state and federal wiretapping laws. Google responded, “We've permanently removed all ads scanning in Gmail for Apps for Education, which means Google cannot collect or use student data in Apps for Education services for advertising purposes.[108]” This affects some 30 million students, teachers and school administrators. But note Google's careful wording; their comforting claims apply only to users of Apps for Education. The remaining half billion or so gmail users and an untold number of people who send mail to gmail users are still subject to Google's data mining. Moreover, Google claimed only to have disabled “ads scanning” for Apps for Education users. That does not rule out scanning for other purposes such as social network data, face recognition in photos, and any location data from smartphone users. Wiretapping is still wiretapping even when the data isn't used for targeting ads. The suit may still be ongoing[109].

Government and law enforcement restrictions on reading email predate most public use of the Internet

[105] www.google.com/intl/en/policies/privacy/key-terms/#toc-terms-unique-device-id
[106] http://en.wikipedia.org/wiki/Web_bug
[107] http://www.edweek.org/ew/articles/2014/03/13/26google.h33.html
[108] http://nakedsecurity.sophos.com/2014/04/30/google-stops-data-mining-students-email/
[109] http://www.bloomberg.com/news/2014-01-28/google-loses-bid-for-appeals-review-of-gmail-wiretap-suit.html

and therefore need strengthening.

> “Alan Butler, appellate advocacy counsel for the Electronic Privacy Information Center, explained[110] that right now law enforcement officials only need to obtain a warrant to access someone’s email if it is stored on a remote server, unread and less than 180 days old. Anything on a server that has been read or is older than the 180 days is open to government investigation without a warrant.
>
> Virtually all email is kept on remote servers by email service providers.
>
> "If you kept letters at home, law enforcement would have to get a warrant," Butler said. "The ECPA was established to extend that protection to electronic data, but the law was passed in the earliest phases of the Internet."
>
> "At the time, because of the costs associated with storage, no one expected that historical communications would be stored on a remote server for any significant length of time."

Other privacy issues need attention as well. Examples include: the Electronic Communications Privacy Act (ECPA), the Cyber Intelligence Sharing and Protection Act (CISPA), the Computer Fraud and Abuse Act (CFAA), and the Trans Pacific-Partnership Agreement (TPP)[111]. These acts tend to reflect old notions about how the Internet is or should be used.

[110] http://america.aljazeera.com/articles/2014/5/1/white-house-electronicprivacy.html
[111] http://www.networkworld.com/article/2164315/lan-wan/4-internet-privacy-laws-you-should-know-about.html