

Information Capacity

IJoC

How to Measure the World's Technological Capacity to Communicate, Store, and Compute Information Part II: Measurement Unit and Conclusions¹

MARTIN HILBERT²

University of Southern California
United Nations ECLAC

PRISCILA LÓPEZ³

Open University of Catalonia

This is Part II of a two-part article that reviews methodological and statistical challenges involved in the quantification of humanity's technological capacity to communicate, store, and compute information. In this Part II, we focus on the adequate unit of measurement for quantifying information. Naturally, we do this from the perspective of the results of our recent inventory of 60 technological categories between 1986 and 2007, in which we propose an information theoretic measure that approximates the entropy of the source (which we call "optimally compressed bits"). We explain the interpretation, creation, usage, benefits, and limitations of this unit of measurement. A more coherent understanding of information volumes and magnitudes starts with a

¹ The authors would like to thank Jim Short and Andrew Odlyzko for their detailed comments, and United Nations ECLAC for giving us the opportunity to set the groundwork for many of the ideas developed in this article. We would also like to thank Tom Coughlin from Coughlin Associates, John McCallum, Don Franz from Photofinishing News, Joerg Woerner from Datamath, Manuel Castells and Len Adleman from USC, our research contributors Miguel Gonzalez and Cristian Vasquez, and the statisticians from ITU (International Telecommunications Union) and UPU (Universal Postal Union).

² Provost Fellow, USC Annenberg School for Communication & Journalism; Economic Affairs Officer, UN Economics commission for Latin America and the Caribbean.

³ Research Scholar, Information and Communication Sciences Department.

Martin Hilbert: mhilbert@usc.edu

Priscila López: priscila.lopez@gmail.com

Date submitted: 2012–02–17

thorough understanding of the methodological choices involved in related inventories. In this sense, the methodological challenges outlined in this article represent an open research agenda. We also discuss statistical lessons learned in our exercise (which is informed by more than 1,100 sources) in the roughly 300-page supporting online Appendix that is available at <http://www.martinhilbert.net/WorldInfoCapacity.html>.

Units of Measurement: The Key Methodological Decision

Building on the groundwork that we have laid in Part I of this two-part article, we now review the most essential methodological choice, the unit of measurement. Many units of measurements have been proposed and used to measure the amount of information, including both the number of words or minutes, and the respective hardware capacity. It would be straightforward to translate those indicators into additional ones, such as the corresponding meters or kilograms of books, or the informational equivalent of newspapers or hand-written pages (for some analogies of this kind, see Hilbert, 2011a). Each system of classification inevitably valorizes some point and silences another (Bowker & Star, 2000). The challenge consists in creating a classification system that is as clear as possible to enable the pursuit of a specific question, and as complex as necessary to present the aspects of reality that need to be considered to correctly answer this question. We therefore continue Part II of the article with one of the main insights we drew at the end of Part I: Any methodological choice eventually depends on the question that is on the mind of the researcher.

We review some of the alternatives units of measurement that have been, and can be, used to quantify information processes, focusing on the measurement unit that we used in our recent inventory (Hilbert & López, 2011; also Hilbert, 2011b). The specific question we pursued in this inventory was the following: How did the world's technological capacity to communicate, store, and compute information evolve over the period 1986–2007? To be able to answer this panel-data question (crossing several technologies and several years) we measured the technological capacity to communicate and store information in "optimally compressed bits," and the technological hardware capacity to compute in "instructions per second" (i.e., millions or mega instructions per second, [MIPS]). We explain the theoretical and practical reasons behind this choice, and then end the article with a self-critical discussion of what these indicators can and cannot show.

The Bit: Communication and Storage

We chose the "optimally compressed bit" as the unit of measurement for communication and storage. Our choice was informed by information theory, which is a branch of applied probability theory that, today, is mainly taught in electrical engineering and communication departments. It is the rare breed of a branch of science that can almost exclusively be traced back to one single and epoch-making paper: Claude Shannon's (1948) "A Mathematical Theory of Communication." Shannon's proofs and conceptualization of the bit have revolutionized our world and changed the course of history (for a popular science story on Shannon, see Gleick, 2011). Thanks to the technologies that followed Shannon's ideas and his conceptualization of the bit, information theory is arguably the scientific theory with the most widely felt practical impact on the daily life of people at the dawning of the 21st century (see Pierce, 1980

for an introduction to information theory; for a more formal approach, see Massey's 1998 lecture notes, which might be an easier read than the standard textbook in engineering departments from Cover & Thomas, 2006, which is more complete).

We translate all kinds of information into "optimally compressed bits." This implies two steps: the translation of information 1) into binary digits of hardware capacity, and then 2) into optimally compressed bits. We walk through both steps in the following text.

From Analog to Binary Digits

Being digital is not the "natural" state of information. Information is normally in its analog form when used by humans, but the digital format is more useful for machines. When we talk about "analog" information, we refer to *any form other than "digital,"* while we define *digital* as information in the form of binary symbols (Using the binary choice between 0s and 1s is merely a social convention; any binary choice would suffice, such as yes/no, there/not there, black/white, up/down, left/right, redpill/bluepill, the number 42/everything else, etc.). Digitizing an analog signal means finding out how many binary decisions are necessary to clearly identify a certain analog signal up to a chosen level of accuracy. For example, we have 256 analog symbols, and would like to clearly identify one of them. ASCII (American Standard Code for Information Interchange) uses eight binary decisions (= one byte) to represent $2^8=256$ symbols (26*2 letters [A-Z and a-z] + 10 numerals [0-9] + 194 other characters). The same applies to the 256 color options for each pixel in a picture formatted as a GIF (Graphics Interchange Format).

Now, how many binary digits are needed to encode an analog signal, such as a picture or analog TV signal? One alternative is to look at the informational equivalent of scanning the picture or using a digital camera to film the same movie (i.e., the technique used in the inventory taken by Lyman et al., 2003). The benefit of this approach consists in the fact that one can readily identify possible compression rates (more on compression later), while one of the problems with it is that a scanned picture also includes the information contained when the scanner recognizes the texture and wrinkles of the paper page, or the border of it, etc. While this is surely also information, it is not part of the information transmitted by the text in question. Another approach is to go back to the information theoretic Nyquist-Shannon sampling theorem (Nyquist, 1928; Shannon, 1949; see also Anttalainen, 2003), which—roughly speaking—provides a theoretical minimum of the number of binary decisions required to replicate an analog wave (e.g., for analog telephony and radio). In our inventory, we follow this second approach. In practice, this means that for analog TV, for example, we analyze the resolution of a traditional NTSC/PAL/SECAM display of a television screen and determine how many binary digits we would need to represent the information on each displayed frame. We end up with the uncompressed number of "raw binary decisions" required to replicate the image on the television screen.

From Binary Digits to Optimally Compressed Bits

Confusingly, there are two kinds of “bits” (Laplante, 1999). The first one refers to the representation of data in forms of 0s and 1s. This refers to the hardware capacity to store or communicate binary signals. The 500 GB (or $500 \times 8 \times 1,000,000,000$ bits) hard disk and the 64-bit processor of a PC refer to this metric. We refer to this kind of data in binary form as “binary digits,” even though some authors sloppily refer to them as “bits.” The other kind refers to bits in Shannon’s (1948) sense. Shannon defines *information* as everything that truly reduces the uncertainty of the receiver, and he defines one bit as the amount of information that reduces uncertainty by half (regarding an existing probability space). In this sense, information is defined as the opposite of uncertainty, and uncertainty can be measured in probabilistic terms.⁴ Shannon’s bits also represent a binary choice—the one that reduces uncertainty by half. If the binary choice does not reduce any uncertainty, that is, if it represents something that the receiver (i.e., the receiving machine) can deterministically infer itself, it is mere redundant data, not information.

Of the useful theorems Shannon proved about his definition of information, the one most useful for our purposes is that each stationary source that emits information counts with a given level of uncertainty. He called this measure the “entropy of the source” and measured it in bits: the number of times uncertainty has to be reduced by half in order to convert the given level of uncertainty of the source into certainty (the resulting reduction of uncertainty is what Shannon defines as information). For example, on a coarse-grained level, we could say that there are four kinds of information sources, respectively emitting text, audio, still images, and video. On average, there is a stable average amount of uncertainty (or its opposite, information) emitted from these sources. This is a nice fact to consider when searching for possible information metrics, because the entropy of the source does not change with the particular technology. The amount of uncertainty reduced by a letter of text is the same if it is communicated by a postal letter, a SMS text message or a highly compressed computer Word document. Another important point for our purposes is that the entropy of the source can be approximated by compressing data (binary digits) to its uttermost rate of compression.⁵

Compression is a key concept. Compressing data means to take the redundancy out of the message. Redundant data can be left out without notably reducing the informational content of the message (reducing data without reducing the amount of information contained; or equivalently, reducing data without increasing uncertainty about the content of the message). This is similar to leaving out some of the letters in “ths txt, wthot reducng yr ably” to decode it. The entropy of the source is defined as the most parsimonious representation of the source that (on average) still allows for the unmistakable reproduction of the full original message.

⁴ It could also be measured in algorithmic terms, since Kolmogorov complexity and Shannon entropy approach each other asymptotically (see Zvonkin & Levin, 1970; also Cover & Thomas, 2006, Ch. 14.3; Li & Vitanyi, 1997, p. 187), but this is only holds in the limit and is very difficult to implement practically.

⁵ In this way, Shannon himself (1951) was the first to estimate the entropy (uttermost compression rate) of English text, which he showed to be around one bit per character.

Achievable compression rates depend on the redundancy of the source, and the redundancy depends on the probability distribution of the source (Cover & Thomas, 2006). Most compression algorithms specify a certain standard compression promise for a specific kind of source, such as those that handle text (including letters, numbers, and other signs), audio and sounds (including voice and music), still images (including both black and white and color), and videos (which are basically a sequence of images in time). In general, video is the most compressible kind of content (being partially predictable—or “redundant”—in both space and time), followed by images (space redundant), and audio and text (partially predictable in time). In our inventory, we use these four broad categories to estimate the dominating rates of compression for a given year (see supporting online Appendix B at <http://www.martinhilbert.net/WorldInfoCapacity.html>).

Box 1. Information vs. Data: An Illustrative Example of the Importance of Compression.

The number of (Shannon’s) bits contained in a message can be different from the number of binary 0s and 1s used to encode the message, and can make a large difference when creating time series, such as we have done. This is because compression algorithms have been improved notably during the past two decades. Actually, many new information technologies only became viable thanks to the efficient compression of information, such as mobile telephony. Compression algorithms allow the same hardware (think of a bucket or a tube) to now handle much more information (more filling) because the content is more compressed.

For example, consider a hard disk with a hardware performance of 1 MB of storage of video in 1986, 1993, 2000, and 2007. 1 MB means that this piece of hardware can hold either a 0 or a 1 8 million times. How much information can we store in this hardware? This depends on the compression of the information in question. Let’s consider MPEG-4 as the optimally conceivable “entropic” compression in 2007. Since MPEG-4 is the commonly used standard for video in 2007, the video from 2007 is already at the optimal compression rate and therefore represents 1 MB. Let’s further assume that there was no compression algorithm available in 1986. Without loss of quality, MPEG-4 can compress video to 1.67% of its original file size (compression factor 1:60; see supporting online Appendix B at <http://www.martinhilbert.net/WorldInfoCapacity.html>). This is because videos are usually highly redundant and predictable: If two frames are exactly the same, the second frame does not reduce uncertainty and is therefore redundant. For video, large parts of the same frame are the same (space redundancy), as well as large parts between consecutive frames (time redundancy). Therefore, the uncompressed 1 MB from 1986 is equivalent to 1/60, or 0.0167 “entropic” MB (now counting Shannon bits, not hardware binary digits). We further assume that, in 1993, video was usually compressed with an algorithm called Cinepack (e.g., in Apple Quicktime, in Microsoft Windows, and the game stations of SEGA, Atari, and Panasonic). Cinepack reaches a compression factor of 1:20. Therefore, the 1 MB of video from 1993 represents 20/60, or 0.33 optimally compressed MB (see supporting online Appendix B). For the year 2000, let’s suppose that most videos were compressed with MPEG-1, which achieves a video compression factor of 1:27. This implies that the 1 MB of video from 2000 is equivalent to 27/60, or 0.45 optimally compressed MB (see supporting online Appendix B). In short, 1 MB of hardware capacity used for video from 1986, 1993, 2000, and 2007 translates to 0.0167, 0.33, 0.45, and 1 optimally compressed MBs, respectively, when normalized with regard to what (in 2007) is considered the optimal compression rate. On the contrary, the amount of hardware binary digits stayed the same at 1 MB. It is unfortunate and confusing that both concepts are often referred to as “bits” (or “kb,” “MB,” etc.).

It is important to state that we do not directly calculate the entropy of the source for each kind of content. This would not be practically possible, and those algorithms do not exist in practice (Todorović, 2006), as, for example, each 90-minute movie contains a different amount of information. What is readily available, however, are reports about the average level of compression that can be achieved by a specific compression algorithm for a certain kind of content: text (e.g., with file formats like .zip or .rar), audio (e.g., with MP3 files), images (e.g., with GIF or JPEG files), and video (e.g., with MPEG-4 files). Once we know both the kind of content, and which kind of program is used to handle it, we can infer the level to which this specific content could be compressed if it were compressed with the most efficient compression algorithm available (see Box 1 for an illustrative example).

What Does Normalization of Compression Rates Require in Practice?

A hypothetical example illustrates the underlying logic and facilitates future practical use of it with statistics (see also Hilbert, 2011b). As shown in Figure 1, we suppose the existence of one storage (or communication) device in year_t, which has a hardware capacity of two physical representations (two silicon-based logic "bit-flip" gates to store information, or two communication transmission cables). Half of the information content consists of images (which are not compressed, as in the cases of industrial x-rays or detailed maps) and the other half consists of text, compressed by a factor of 2:1 (using, for example, the Lempel-Ziv-Welch algorithm used in early UNIX systems in the 1980s). This implies a technological capacity to store (or communicate) 3 bits in year_t. In year_{t+1}, investment in infrastructure leads to a duplication of the number of devices and technological progress in hardware leads to a triplication of storage (or communication) units per device. Moore's (1995) famous law, for example, tracks this kind of progress by registering the number of transistors that can be placed on an integrated circuit. Additionally, we suppose that images are now compressed with JPEG (the norm in personal and industrial image handling in 2007), achieving a high-quality compression factor of 11:1, while text is compressed with .zip or .rar, reaching a factor of 5:1. This enables us to store (or communicate) a total of 108 bits in year_{t+1}. The result is a multiplication of the initial amount of information by 36 (108/3, or a growth rate of 3,500%). As visualized by Figure 1, this total of technological change can be traced back to 1) a duplication of infrastructure (growth of 100%), 2) a triplication of hardware performance (growth 200%), and 3) a sextuplication of the software performance for content compression (500%)⁶: $(1+1)(2+1)(5+1)=36$.

⁶ In practice, the contribution of content compression is a combination of the advancement in compression algorithms (software performance) and the general shifts in the kind of content. If more compressible content gains importance, the average technological progress of content compression increases.

The contribution of ever-more powerful compression algorithms for digital content (software performance) can either be calculated as a weighted average of the progress of compression of each kind of content, or as a residuum.⁷ In practice, the latter alternative is more straightforward.

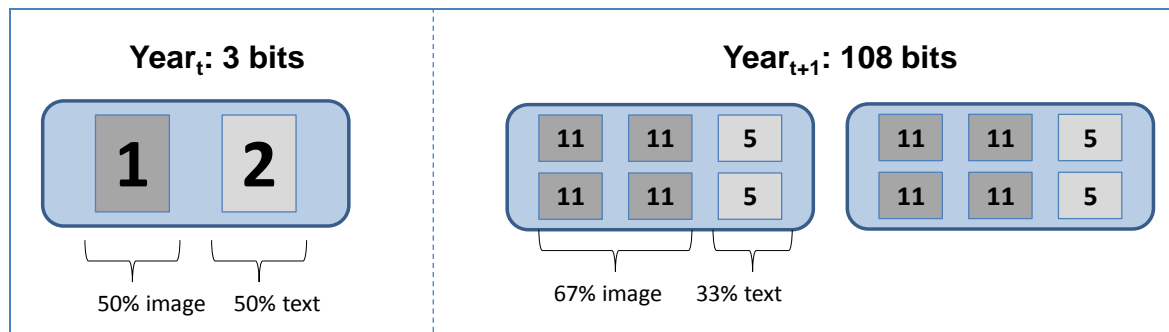


Figure 1. Inside the black box of digital technological progress to store and communicate information. Source. Authors' own elaboration.

As a result, to be able to meaningfully estimate the amount of information stored and communicated in technological devices, we require statistics on:

- (i) the amount of infrastructure and devices,
- (ii) the hardware capacity of each device, and
- (iii) the compression rate with which information is compressed, which itself depends on
 - (iii.a) the type of content compressed in digital technologies (in our level of coarse-graining text, audio, images, and video), and on
 - (iii.b) the dominating compression algorithm used and the optimal level of compression available for this kind of source.

⁷ In the given example, the option of the residuum is straightforward: $[108/3]/2/3 = 6$. More error-prone, but no less correct is the calculation by way of expected value: Year_t: 1/2 of hardware contains image, 1/2 of hardware contains text; Year_{t+1}: 2/3 of hardware contains image, 1/3 of hardware contains text. This means that 1/2 of the hardware stays as image (equal to 1 hardware unit of Year_t), 1/6 of the hardware is converted from text to image (equal to 2/6 hardware units of Year_t), and 1/3 of the hardware stays text (equal to 2/3 hardware units of Year_t). Expressed in bits of Year_t, this is equal to 1 bit staying image (of 1/3 of the bits of Year_t), 4/6 bits being converted from text to image (or 2/9 of the bits of Year_t), and 4/3 bits staying image (or 4/9 of the bits of Year_t). These are the right weights to apply to calculate the weighted average of the contribution of compression: $1/3(11/1)+2/9(11/2)+4/9(5/2) = 6$.

Since these last two statistics are scarce and unreliable, we decided to pick only four representative years over the past two decades to carry out our estimations: 1986 (basically before the digital age), 1993 (the start of the era of the Internet and mobile telephony), 2000 (the height of the financial Internet bubble), and 2007 (the last year for which we obtained reliable global statistics). This choice of four points which are equally distant in time is a compromise. On the one hand, 1986, 1993, 2000, and 2007 are far enough apart to ensure that, in each year different kinds of compression algorithms would have been adopted worldwide. On the other hand, it allows us to measure three equally long periods of growth (1986–1993, 1993–2000, and 2000–2007) for the timeframe for which we have sufficiently reliable statistical sources, which is important because three periods is the minimum requirement to obtain a basic understanding of the overall shape of the growth process: One or two periods might be deceiving when identifying the typical shape of the growth process, which are traditionally either linear, exponential, or logistically S-curve shaped.

What Does "Optimal Compression" Mean for Our Purposes?

There are two additional caveats that have to be addressed when working with any normalization of compression rates: quality and technological progress. Compression algorithms can be "lossless" or "lossy." Lossless compression algorithms only take out those redundancies that do not take away any information from the message, while lossy algorithms reduce the quality of the information content. For example, when compressing a photo to a size that is adequate for upload as an email attachment or a thumbnail on a social networking site, the size of the file is reduced, but the quality of the image often suffers as well. This kind of compression is "lossy." While most of the modern compression algorithms, including JPEG, MP3, and MPEG-4, most commonly use lossy compression, they often also allow the user to choose the level of "loss."⁸ The accompanying manuals and technical reports classify the results of the compression in various groups, which they give names like "very good" or "excellent quality" results.⁹ These latter two categories usually include both products of lossless compression and results of lossy

⁸ In addition to lossy compression, there is also often loss of information during transmission or retrieval. This frequently happens with analog information (such as the notorious snowstorm on analog terrestrial TVs during times of bad reception). The typical degree of interference is usually not reported by equipment producers or service providers. How should we adjust for this potential loss of informational performance? For analog fixed-line telephony, the literature reports so-called signal-to-noise ratios (Anderson & Johannesson, 2005). Luckily, one of the benefits of digital transmission is that error-correction codes reduce information loss to a minimum (see Pierce, 1980). Notwithstanding, it is common that wireless technology (e.g., mobile telephony) is subject to some kind of interference. One can argue that only a part of the transmitted information actually reaches the receiver (depending on the quality of reception). We are not aware of any statistics that report the average interference and reception challenges for such digital technologies.

⁹ For the case of telephony, for example, we use the mean opinion score (MOS) (ITU-T, 1996), which provides a numerical indication of the perceived quality of received media after compression and/or transmission (see supporting online Appendix D at <http://www.martinhilbert.net/WorldInfoCapacity.html>).

compression where the reduction of information cannot be noted by the human observer (i.e., is not possible to receive with the given resolution of the human senses, such as sounds at a very high pitch, or image details too small for the eye to perceive). Additionally, these reported results of tests often also include some additional bits of redundancy that are added by the compression programs to increase the robustness of the content.

This being said, in our inventory, we define the *optimal level of compression* in a given year as *the uttermost level of compression that is achievable with the most powerful existing compression algorithm in this given year, while achieving a level of quality that is indistinguishable by the human observer from lossless compression.*

The foregoing definition of “optimal compression” points toward another qualification to consider: That which was considered optimal compression a decade ago is not the same as what we understand to be optimal compression today. The “optimal” compression algorithm has changed considerably over recent decades, and it is expected to continue to change over time (the exception seems to be for compression of text¹⁰). For some kinds of content, it has only been in recent years that we have made great progress in approximating the probabilistic nature of the source (e.g., with the introduction of so-called turbo codes in 1993), and the search for ever more perfect compression algorithms is still ongoing. It can therefore be ambiguous to declare “optimality” without specifying the particular year in which the algorithm is or was “optimal.” For practical reasons, in our exercise, we normalized the data to what was considered “optimal compression” in the year 2007. This means that one might be able to readjust our exercise in a couple of years and renormalize it to the newly found “optimal compression rates” as coders discover how to exploit yet-unknown structures in the most diverse streams of data.

It is important to underline that the method we chose is, of course, not the only possible way to normalize compression rates across several decades. One might as well allow for more or less information loss, or normalize to some other rate of compression besides the optimally achievable in a given year. For example, we could have normalized to the “most commonly used” level of compression of a given year, or any other characteristic level of compression (e.g., 2000), much like economists normalize to inflation rates of specific years (which is a similarly moving target). Of course, this does not change the validity of the results. However, there are theoretical and practical reasons for opting for the most recent technological frontier.¹¹

¹⁰ Compressing each English character down to 1.21 with an algorithm called DURILCA is within the theoretical limits proposed by Shannon (1951), between 0.6–1.3 bits per character.

¹¹ One theoretical reason is the fact that the most efficient existing compression algorithm “approaches the entropy of the source” the best as possible (engineers refer to this as “entropic compression,” which is naturally redefined with any new and more efficient compression algorithm), while one practical reason is the fact that all other levels of compression in the time series of interest will be at a lower level of compression, and therefore, one does not need to “decompress” the content that is compressed at a higher level than the level chosen for normalization.

Box 2. Thought Experiment: Global Compression as the Ultimate Test for Uniqueness of Information

Being aware of the nature and concepts behind the compression of information, we can now return to the question of how to unambiguously identify the amount of "unique and original information," which we discussed in Part I of this article. What compression algorithms effectively do is take out any duplication (redundancy) of information in a message. In reaching our estimate that the world's global capacity to store information was about 300 optimally compressed exabytes in 2007, we only eliminated the redundancy contained in an archive, not the redundancy between archives. For example, we said that an average analog video contains roughly 98% redundancy, and can therefore be compressed by a factor of 1:60 with MPEG-4 (see Box 1). However, we do not count the compression that would be possible if an algorithm would recognize that the same video is copied twice on the same hard disk. If this would be discovered, an intelligent algorithm would only have to store the first video, plus the instruction "copy the video in case it is necessary" (which is quite bit-efficient). The separate storage of the second video would be "redundant" and would not actually provide any real information. This logic is already applied by intelligent solutions that store different "playlists." A "playlist" does not require a "recopy" of every song, but merely storage of the sequence in which to play the songs. However, these algorithms do not usually look for duplicates of songs automatically. If the program were truly efficient, it would even recognize that parts of an archive (text, audio, image, or video) are in agreement with similar parts of another archive (for example, if a paragraph is copied from one document to another), and would not have to store the entire paragraph again, but only the command "insert paragraph XYZ here." This is usually not what happens on hard disks. Most compression algorithms do not search for redundancy among different files, but only within a given file.¹²

In this sense, we can imagine a thought experiment whereby the world's total amount of information is stored on one giant hard disk. We could then imagine running some large compression algorithm on all the world's archives, having programmed this algorithm to look for similarity among archives, and to identify which part of which archive is truly unique and which part is redundant. It would surely turn out that some parts of a song on the hard disk would be in agreement with other parts of a song on the same hard disk. After crawling through all the world's archives, the algorithm would provide us with the optimal compression among all of the world's information, a total which would be equivalent to the amount of truly unique information. Comparing this with the truly unique information from a previous moment in time would allow us to identify how much truly unique information had been produced in a given year.

Unfortunately, this is only a thought experiment at this moment, and we do not count with such an algorithm; neither do we count with the practical possibility to run it on the world's global stock of information. It is therefore a theoretical concept at this point, but it shows how uniqueness of information could be determined by compression.

¹² Biologists, however, already work with a similar logic (see Christley, Lu, Li, and Xie, 2009).

MIPS: How to Measure Computation?

The choice of a unit of measurement for computation can even be more confusing than the choice for communication and storage. In essence, a computation is some kind of action on a group of input which produces one output. Usually, this transformation of information in time follows some algorithm (a "procedure"). Similar to the cases of information storage and communication, which depend on hardware and compression algorithms, the performance level of a computation depends on (a) hardware (the number of transformations) and (b) software algorithms (the way these transformations are performed).

For hardware performance, the most commonly available performance indicator is the number of MIPS. In our exercise, we use MIPS of Dhrystone 1.1. This indicator is not without criticism, mainly because MIPS performance depends on several conditions, such as the computer's input/output speed and the processor architecture, and can therefore favor or discriminate against certain types of computer design (Lilja, 2000). Other indicators have been proposed by the industry, such as FLOPS (floating points operations per second) and SPEC (system performance evaluation cooperative; see Hennessy & Patterson, 2006). The latter is often updated (e.g., SPEC CPU92, SPEC CPU95, SPEC CPU2000, SPEC CPU2006), and as such, it does not allow for coherent comparisons over time. The former (FLOPS) is the standard to measure the performance of supercomputers (Weicker, 1990), or graphical processing units (GPU). We also use it in our exercise (translating it to MIPS). MIPS, however, is by far the most commonly used hardware performance indicator for the period of the 1980s and 1990s, and statistics tracked in MIPS are widely available (e.g., Longbottom, 2006; McCallum, 2002; Nordhaus, 2007). In this sense, similar to what we often found in Part I of this article, this decision is based on reasons of practical feasibility, and not on any theoretical or conceptual superiority of the indicator.

In theory, the performance of a computer depends on the software. The number of hardware instructions of a computer is roughly equivalent to the number of binary digits that can be found with storage or communication hardware. The kind of software run on this hardware also has become more efficient during recent decades. One can carry out computational tasks with 1 MIPS in 2007 that could not have been executed with the same 1 MIPS in 1986. This is because ingenious software engineers constantly develop better algorithms. Computer scientists are very meticulous about measuring the time-performance of their algorithms, which they do in Θ -notation (or simply O-notation; see Cormen, Leiserson, Rivest, & Stein, 2003). The two main contributors to the speed of algorithms for certain computational tasks are the "leading constants" (which reduce the computational time linearly) and their "polynomial degree" (reducing their computation time exponentially). Faster algorithms allow the execution of the same task much faster with the same amount of hardware instructions per second.¹³

¹³ The journal *Computing in Science & Engineering* lists what it considers to be the top 10 algorithms of the 20th century (Dongarra & Sullivan, 2000). From those selected, one gets the impression that the heyday of the greatest discoveries in algorithms was during the middle of the 20th century and not toward its end. This does not mean that, since then, algorithms have not continued to improved, but that they improved less, mainly by way of improvements in leading constants, not polynomial steps. The importance

To be able to track this progress of software performance for our purposes, one not only needs to know the performance of each algorithm (which might be a feasible task), but also which kinds of algorithms are used by which computational devices, in which intensity, and at which point in time. How frequently does a PC sort information? How often does it optimize something? How frequently does it do which of any kind of matrix transformation? What about a supercomputer or a mobile phone? Unfortunately, these statistics do not exist at this point. They would first have to be created to pursue those questions.

One of the major drawbacks of MIPS as the performance indicator for computer hardware capacity is that MIPS is not directly comparable to the number of bits stored and communication. Short, Bohn, and Baru (2011) therefore estimate the number of bytes processed by enterprise servers per transaction, an estimation executed by several benchmark tests. The result is the hardware capacity in bytes, and therefore presented in the same unit of measurement as storage and communication when measured in binary digits.

Conclusions and Limitations

Let us summarize some of the points made in Part II of this two-part article and discuss the limitations of our approach. In agreement with our conclusions from Part I and in agreement with other researchers in the field, we want to state clearly that “we view this report as a ‘living document’” (Lyman et al., 2003, p. 14). It is our hope that our numbers and methodological decisions can be improved upon in the future (see also our almost-300-page supporting online Appendices A, B, C, D, and E). That being said, we discussed our choice for the unit of measurement in this Part II. Let’s now finish with a critical look at what this measure can and cannot explain.

What Entropic Information Quantity Can Explain

Three of the benefits of the metric of “optimally compressed bits” are that 1) it focuses on a fundamental level, 2) it is a valid measure for time-series, and 3) it provides an objective and non-

of leading constants is not to be underestimated for our practical purposes, since an algorithm that runs twice as fast is equal to roughly 1.5 years of progress in computational hardware (which doubles every +/-18 months; see Hilbert & López, 2011). Important improvements are still being made in the fields of gaming, optimization, and approximation algorithms. The top 10 of the *Computing in Science & Engineering* include the Dantzig simplex method for linear programming in 1947, the Krylov subspace iteration in 1952, the Metropolis Algorithm in 1953, the Fortran I compiler in 1957, the QR algorithm in 1958, the decompositional approach to matrix computation in 1961, a perspective on Quicksort in 1962, the Fast Fourier transform in 1965, the integer relation detection in 1977, and the fast multipole algorithm in 1987. Furthermore, though not published in this list, we might speculate that many computations today are spent on public-key cryptography, like RSA in 1978.

ambiguous logic for how to transform information that resides in different sorts of content (text, images, audio, video) into the same unit.

It is widely accepted in all branches of science that entropy, i.e., Shannon's concept of the bit (entropy), is an indicator on a fundamental level. This approach defines information in Shannon's (1948) sense as the reduction of uncertainty on the syntactic level (given a certain probability space), and we approximate entropy by the uttermost possible compression of a string of data. This measure is not only clearly defined, but also measurable. It can be (and has been) applied to quantify information in all kinds of disciplines and branches of science. Entropy is maybe the most fundamental level there is (to cite the eloquent question of the renowned physicist John Wheeler: "It from bit?" 1990, p. 5; see Zurek, 1990, for the role of information in physics; see Adami, 1997, for its role in biology). While this fundamental measure certainly has limited explanatory power, the technological capacity to handle optimally compressed bits can be used as a basis upon which to build more advanced and specific theories and test other research questions later on. This might include questions like the following: How many bits are actually consumed? How many bits are paid attention to? How do people value different bits, and at which moments?

Second, it is important to remember that optimally compressed bits are a hypothetical measure, in that they do not measure "what is," but "what would be if all data were optimally compressed." This is different from other studies that measure the hardware capacity in (what we call) binary digits (e.g., Gantz et al., 2008). The quantification of hardware capacity might provide important insights, especially when working in the hardware industry or related fields. This measure is therefore not less valid; it simply measures something different than our measure. The number of bits that are represented by one binary digit depends on the chosen compression rate, which depends, in theory, on the probability distribution of the content in question, and in practice, on our knowledge of and technologic capability to exploit the probabilistic nature of the source without reducing the quality of the content (compression). The normalization to specific compression rates changes the results significantly. According to our estimates, advancements in information compression resulted in the fact that a given (hardware) Internet bandwidth in 2007 carried three times more information than the same (hardware) bandwidth in 1986 (see Hilbert, 2011b). In other words, if a 1 Mbps modem in 1986 was able to transmit 1 Mbps, the compression algorithms available in 2007 enabled us to send 3 Mbps of optimally compressed information through the same 1 Mbps hardware channel. Therefore, normalization to compression rates is especially important for the creation of meaningful time series (see Box 1 for an illustrative example).

Once normalized, it is straightforward to meaningfully compare the results in a time series. For example, if one would like to measure the role and impact of "more information" on certain other socioeconomic indicators over time (like economic growth, democratic stability, education, health, etc.), it would be less insightful to test for hardware capacity as the independent variable. For impact, it does not matter if lots of hardware or little hardware is used; what matters is what is on this hardware. Normalization to compression rates makes the actual amount of information on the hardware comparable over time.

A third benefit of this measure is that it allows for the unambiguous measurement of analog and digital content, including a meaningful measure for different kinds of content (text, images, audio, video). This is more challenging when using metrics like minutes, words, or binary digits that host content with different compression rates.

For example, an old adage says that "A picture is worth a thousand words." According to our assumptions, a newspaper image of one cm² is worth between 106 and 213 words. So, turning the ration the other way around, we are able to reconfirm ancient wisdom and conclude that a 6 cm² newspaper image is worth a thousand words (in information theoretic terms).¹⁴ Similarly, comparing the informational magnitudes of written and spoken words, our information theoretic approach shows that a spoken word reduces 750 times more uncertainty than a written word (defining uncertainty in precise probabilistic terms).¹⁵ While this might sound surprising at first, it becomes reasonable when one considers that sound does resolve a considerable additional amount of uncertainty, because there are many choices for how to say the word. For example, when listening to someone speaking, one knows immediately if this person is an adult or child, man or woman. The listener also gets a great deal of information about "how things are said," beyond simply "what is said." Suppose that there are only three levels of pitch (high, medium, low), four speeds at which to say a word (hasty, fast, medium, slow), nine ways to pronounce a word, and seven types of the overall tone. We already have 756 versions of "how to say" the same word ($3 \times 4 \times 9 \times 7$). This means that this kind of spoken words resolves 756 times the uncertainty that the same word does when presented in a written format. Of course, this comparison is not precise in information theoretic terms, but it serves to explain the underlying logic of resolving uncertainty in probabilistic terms.

Something similar applies to the comparison of information contained in videos (moving images), still images, and words. The pioneering Japanese Information Flow Census assumed that one minute of TV broadcasting (moving pictures and audio) is equivalent to 1,320 words (Duff, 2000, p. 79), while Bohn and Short (2009, p. 32) assumed 153 words per minute, or roughly the number of words that can be spoken during one minute. These choices have been justified on the basis of the pursued research question. Our normalized estimates report that one minute of TV broadcasting is the informational

¹⁴ A printed word can be codified with 44 bits (8 ASCII bits/character*5.5 characters/word), which can be compressed to its entropic level of 6.7 bits/word. Compressing each English character down to 1.21 with an algorithm called DURILCA, which is within the limits proposed by Claude Shannon (1951) between 0.6–1.3 bits per character. A 1 cm² low-quality, grayscale newspaper image from 1980 contains 709 bits (compressed by a factor of 6:1), and a 1 cm² normal quality color newspaper image from 2007 contains 1,422 bits (compressed by a factor of 16:1).

¹⁵ We measure 12 optimally compressed kbps for a digital fixed-line phone and 8 optimally compressed kbps for a 2G GSM mobile phone subscription. In terms of words, the Japanese Information Flow Census (Duff, 2000, p. 79) and the UCSD study (Bohn & Short, 2009, p. 32) both assume that 120 words are spoken per minute on the phone. Combined, these numbers result in 6,000 bits per word for our estimation of fixed line telephony, and 4,000 bits per word for our estimation of mobile telephony (let's average to 5,000 bits per spoken word with acceptable quality). 5,000 divided by the 6.7 optimally compressed bits of a written word \approx 750.

equivalent of 7.75 million to 19 million words.¹⁶ This means that, roughly speaking, if one would like to describe everything happening during this one minute of action on the TV screen (every little perceivable detail of movement and color change), one would need the informational equivalent of 12 million words. This is the unambiguous result of applying Shannon's logic: If it reduces the uncertainty of the receiver by half, it is a bit, regardless of whether this reduction in uncertainty is achieved by words, images, sounds, or videos. If all of this detailed information that is displayed by the TV set is also perceived by the viewer, or how much value it adds, is another question, which could be analyzed in a subsequent, additional analysis. It is likely that there are decreasing returns to the amount of informational detail displayed, but at this point, this is merely a hypothesis which will still have to be tested empirically. This brings us to our final point.

What Entropic Information Quantity Cannot Explain: The Value of Information

It is also clear that such fundamental indicators will not be able to explain everything. But it will surely be helpful to explain "something" on a basic and fundamental level, and this can contribute to the isolation of other yet-unknown aspects of our complex reality. As a result, the residuum of the unknown becomes smaller, which is the essence of the entire scientific enterprise.

The quantification of the amount of bits indirectly fosters our understanding of more complex and intangible aspects of communication. For example, if the same number of bits leads to different results (*ceteris paribus*), it can be inferred that the difference is due to yet-unexplored (rather qualitative) aspects of this quantity of information. If the effect changes with the quantity of information bits, we can infer that some correlation does exist, and we can isolate the remaining residuum. While quantity will certainly not explain "everything," it resolves a part of the mystery. Narrowing down the unknown allows us to evolve new and more specific theories, reducing the complexity of reality step by step.

Part of the unknown that we do not yet understand has to do with the value of information. Some of the bits that we handle are certainly more valuable than others, while some might be worthless. For example, instead of an information-rich voice call, maybe the informational needs of the recipient would be just as satisfied when receiving an SMS text message. Maybe it would be enough to tell the user in words that "the stock market closed at 10,000 points," or that "the team won 2-1," instead of having to show them exactly what it looked like. This might be true, but it is a different and additional question. It asks how people value information, which is a more sophisticated question than determining the amount of information, which is the question that we try to answer. The latter is, however, a condition for understanding the former, because the value of information has to be defined as "value per informational unit": US\$/bit, attention/bit, pleasure/bit, happiness/bit, etc., or simply value/bit.

¹⁶ TV broadcasting on an analog black and white NTSC television set receives 0.866 optimally compressed Mbps, while an analog color TV set receives 1.308 optimally compressed Mbps (NTSC), and a digital TV set 2.155 Mbps (downstream and upstream).

This also means that both are independent, and that the variable for "value" can be freely defined. It also means that we will not be able to answer the question of the value of information until we start to quantify the amount of information. The ideas conveyed by a book often seem almost as complete as the ideas conveyed by a movie. Above, though, we found out that one would need the informational equivalent of 12 million words to describe one minute of action on the TV screen. This seems like a lot of words for details that might not be of value, which would propose the following hypothesis: [value/bits in words] > [value/bits in video]. If this hypothesis were always correct, perhaps it would turn out that the media-rich transmissions of audio and video are totally redundant, and that people of the future will exclusively rely on content-succinct texting instead of televisual content. Until now, this does not seem to be the case, and people seem to highly appreciate the additional information provided by media-richness in most situations (and also because video can transmit the same information much faster than words). However, anecdotal experience seems to suggest that there are diminishing returns to media-richness. But again, for now, all of these questions are merely hypotheses that have yet to be tested rigorously. To be able to test hypotheses about information value, we will have to start measuring the quantity of information first.

Our choice of the bit proposes a strategy that suggests setting the focus on basic indicators that reflect essential and irreducible parts of broader theories. In one way or the other, every quantitative theory of information and communication will have to deal with the probabilistic nature of information and the resolution of uncertainty on the syntactic level. Of course, cultural, monetary, semantic, and all kinds of other values can be assigned to bits, and each bit can be given a different weight. But this does not change the fundamental mathematical character of information, which is defined by the number of the binary choices necessary to resolve the respective amount of uncertainty (its number of bits).

The role of information and communication in a society is still not well understood, but we know that, as always in science, measurement and quantification are important ingredients in gaining deeper insight. This process is painstaking and labor-intensive, especially when done for the first time. We nurtured our motivation during the past four years of bit-counting by knowing that we contribute to the sight of a one-eyed researcher-king in the land of the blind.¹⁷ In this sense, our exercise and the exercises of our colleagues leave no doubt that there is much more to discover and understand about the nature and role of information in society.

¹⁷ *In regione caecorum rex est luscus*. [In the land of the blind, the one-eyed man is king]. (32) Desiderius Erasmus of Rotterdam, *Adagia* (2396. III, IV, 96, 1515).

Supporting Online Appendix

We place great emphasis on transparency in outlining the methodological assumptions and sources on the basis of which we elaborated the presented estimates. More than arguing in favor of one specific number, we see the presented estimates as approximations, which could certainly be improved (depending on the available resources). To facilitate future generations of research on this topic, 300 pages of Supporting Appendix that outline the details of the applied methodology, enlisting more than 1,100 distinct sources to include:

Supporting Appendix, Material, and Methods

- A. Statistical Lessons Learned
- B. Compression
- C. Storage
- D. Communication: incl. update for telecommunications (telephony and Internet) until 2010
- E. Computation

This Supporting Appendix can be accessed at <http://www.martinhilbert.net/WorldInfoCapacity.html>.

References

- Adami, C. (1997). *Introduction to artificial life* (Corrected). New York: Springer Verlag.
- Anderson, J. B., & Johannesson, R. (2005). *Understanding information transmission*. Hoboken, NJ: Wiley-IEEE Press.
- Anttalainen, T. (2003). *Introduction to telecommunications network engineering* (2nd ed.). Norwood, MA: Artech House Publishers,
- Bohn, R., & Short, J. (2009). *How much information? 2009 report on American consumers*. San Diego, CA: Global Information Industry Center of University of California, San Diego. Retrieved from <http://hmi.ucsd.edu/howmuchinfo.php>
- Bounie, D. (2003). *The international production and dissemination of information*. Special Project on The Economics of Knowledge, Autorità per le Garanzie nelle Comunicazioni. Paris: École Nationale Supérieure des Télécommunications (ENST). Retrieved from <http://ses.telecom-paristech.fr/bounie/documents/Recherche/Annex.pdf>
- Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: The MIT Press.
- Christley, S., Lu, Y., Li, C., & Xie, X. (2009). Human genomes as email attachments. *Bioinformatics*, 25(2), 274–275. doi:10.1093/bioinformatics/btn582
- Cormen, T., Leiserson, C., Rivest, R., & Stein, C. (2003). *Introduction to algorithms* (2nd ed.). Boston: McGraw-Hill Science/Engineering/Math.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: Wiley-Interscience.
- Dongarra, J., & Sullivan, F. (2000). Guest editors' introduction to the top 10 algorithms. *Computing in Science & Engineering*, 2(1), 22–23. doi:10.1109/MCISE.2000.814652
- Duff, A. S. (2000). *Information society studies*. New York: Psychology Press.
- Gantz, J., Chute, C., Manfrediz, A., Minton, S., Reinsel, D., Schlichting, W. et al. (2008). *The diverse and exploding digital universe: An updated forecast of worldwide information growth through 2011*. Framingham, MA: International Data Corporation, sponsored by EMC. Retrieved from <http://www.emc.com/leadership/digital-universe/expanding-digital-universe.htm>
- Gleick, J. (2011). *The information: A history, a theory, a flood*. New York: Pantheon.

- Hennessy, J. L., & Patterson, D. A. (2006). *Computer architecture: A quantitative approach* (4th ed.). Waltham, MA: Morgan Kaufmann.
- Hilbert, M. (2011a). *That giant sifting sound*. Video presentation at *The Economist Ideas Economy: Information Summit*. *The Economist* and Proof-inc, June 7th-8th 2011, Santa Clara, California. Retrieved from <http://ideas.economist.com/video/giant-sifting-sound-0>
- Hilbert, M. (2011b). Mapping the dimensions and characteristics of the world's technological communication capacity during the period of digitization. Working Paper. Presented at the 9th World Telecommunication/ICT Indicators Meeting. Mauritius: International Telecommunication Union. Retrieved from <http://www.itu.int/ITU-D/ict/wtim11/documents/inf/015INF-E.pdf>
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60–65. doi:10.1126/science.1200970
- ITU-T. (1996). Recommendation P.800 (08/96): *Methods for subjective determination of transmission quality*. Geneva: International Telecommunication Union. Retrieved from <http://www.itu.int/rec/T-REC-P.800-199608-I/en>
- Laplante, P. A. (1999). *Electrical engineering dictionary, CRCnetBASE 1999*. CRC Press LLC, Boca Raton, FL.
- Li, M., & Vitanyi, P. (1997). *An introduction to Kolmogorov complexity and its applications* (2nd ed.). New York: Springer.
- Lilja, D. J. (2000). *Measuring computer performance: A practitioner's guide*. New York: Cambridge University Press,
- Longbottom, R. (2006). *Computer Speed Claims 1980 to 1996*. Roy Longbottom's PC Benchmark collection. Retrieved from <http://www.roylongbottom.org.uk/mips.htm>
- Lyman, P., Varian, H., Swearingen, K., Charles, P., Good, N., Jordan, L. et al. (2003). *How much information? 2003*. University of California, Berkeley. Retrieved from <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003>
- Massey, J. (1998). *Applied digital information theory: Lecture notes by Prof. em. J. L. Massey*. Zurich: Swiss Federal Institute of Technology. Retrieved from http://www.isiweb.ee.ethz.ch/archive/massey_scr
- McCallum, J. (2002). Price-Performance of Computer Technology. In V. G. Oklobdzija (Ed.), *The computer engineering handbook* (pp. 136–153). Boca Rotan, FL: CRC Press.

- Moore, G. E. (1995). Lithography and the future of Moore's law. *Proceedings of SPIE* (International Society for Optics and Photonics) (pp. 2–17). Presented at the Integrated Circuit Metrology, Inspection, and Process Control IX, Santa Clara, CA. doi:10.1117/12.209195
- Nordhaus, W. D. (2007). Two centuries of productivity growth in computing. *The Journal of Economic History*, 67(1), 128–159. doi:10.1017/S0022050707000058
- Nyquist, H. (1928). Certain topics in telegraph transmission theory. *AIEE Transactions*, 47, 617–644.
- Pierce, J. R. (1980). *An introduction to information theory* (2nd rev. ed.). New York: Dover Publications.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656. doi:10.1145/584091.584093
- Shannon, C. (1949). Communication in the presence of noise. *Proc. Institute of Radio Engineers*, 37(1), 10–21. Reprint as classic paper in *Proc. IEEE*, 86(2), Feb. 1998. Available at <http://www.stanford.edu/class/ee104/shannonpaper.pdf>
- Shannon, C. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, 30, 50–64.
- Short, J., Bohn, R., & Baru, C. (2011). *How much information? 2010 report on enterprise server information*. San Diego, CA: Global Information Industry Center at the School of International Relations and Pacific Studies, University of California, San Diego. Retrieved from http://hmi.ucsd.edu/howmuchinfo_research_report_consum_2010.php
- Todorović, A. L. (2006). *Television technology demystified: A non-technical guide*. Oxford: Focal Press.
- Weicker, R. P. (1990). An overview of common benchmarks. *Computer*, 23(12), 65–75.
- Wheeler, J. (1990). Information, physics, quantum: The search for links. In W. H. Zurek (Ed.), *Complexity, entropy and the physics of information* (pp. 3–28). Oxford: Westview Press.
- Zurek, W. H. (1990). *Complexity, entropy and the physics of information*. Oxford: Westview Press,
- Zvonkin, A. K., & Levin, L. A. (1970). The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematics Surveys (Uspekhi Mat. Nauk)*, 25(6), 83–124.