Chapter 1 – Time / Hyperhistory

The Three Ages of Human Life

More people are alive today than ever before in the evolution of humanity. And more of us live longer today than ever before. Life expectancy is increasing (Figure 1; see also Figure 18) and poverty is decreasing (Figure 2), even if the degree of global inequality is still scandalous. As a result, disability is becoming the biggest health-related issue for humanity.

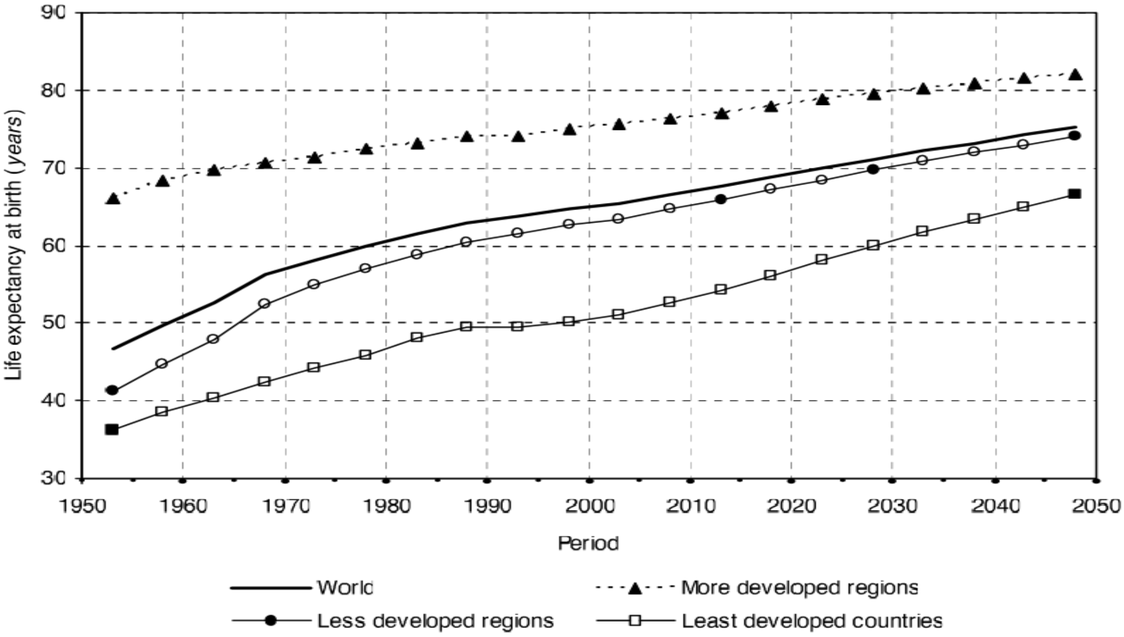

Figure 1. Life expectancy at birth for the world and major development group, 1950-2050. Source: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (2005). World Population Prospects: The 2004 Revision Highlights. New York: United Nations.

To a large measure, the lines on Figure 1 and Figure 2 have been drawn by our technologies, at least insofar as we develop and use them intelligently, peacefully, and sustainably.

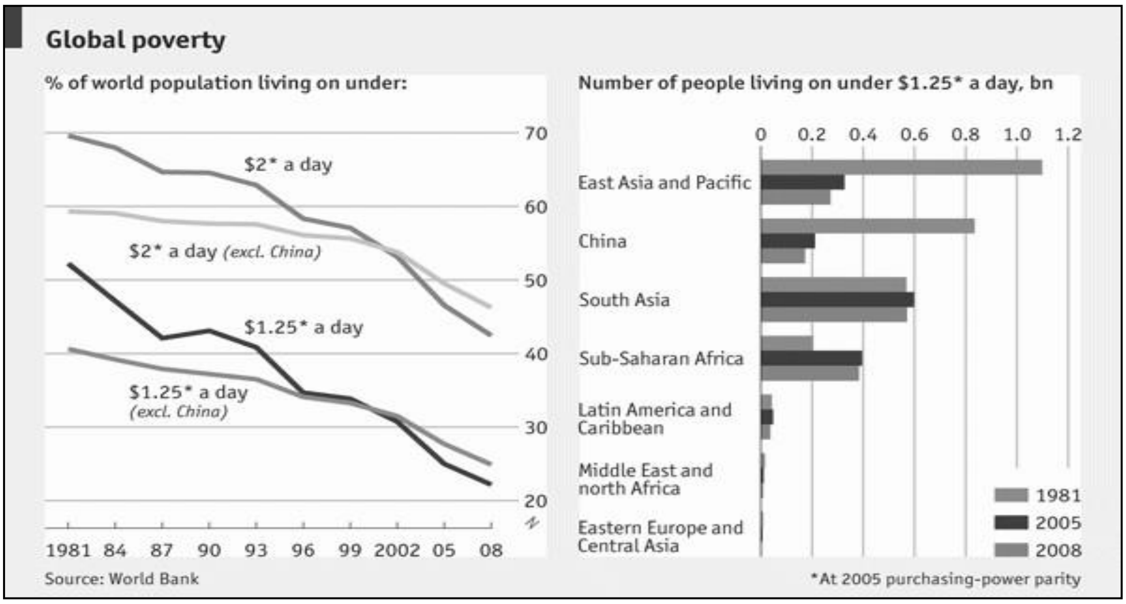

Figure 2. Poverty in the Word defined as the number and share of people living below $1.25 a day (at 2005 prices) in 2005-08. Source: World Bank, picture © The Economist, 29 February, 2012.

Sometimes, we may forget how much we owe to flints and wheels, to sparks and ploughs, to engines and computers. We are reminded of our deep technological debt when we divide human life into *prehistory* and *history*. Such a significant threshold is there to acknowledge that it was the invention and development of information and communication technologies (ICTs) that made all the difference between who we were, who we are, and, as I shall argue in this book, who we could be and become. It is only when systems to record events and hence accumulate and transmit information for future consumption became available that lessons learnt by past generations began to evolve exponentially, in a soft or Lamarckian[2] way, rather than in a hard or Darwinian one, and so humanity entered into history.

No records, no history, so history is actually synonymous with the information age, since prehistory is that age in human development that precedes the availability of ICTs. Such a line of reasoning may suggest that humanity has been living in various kinds of information societies at least since the Bronze Age, the era that marks the invention of writing in Mesopotamia and other regions of the world (4th millennium BC). Indeed, in the 3rd millennium BC, *Ur*, the city state in Sumer (Iraq), represented the most developed and centralized bureaucratic state in the world. So much so that, before the Gulf war (1991) and the Iraq war (2003-2011), we still had a library of more than 500,000 clay tablets; no love letters or holiday stories, but mainly inventories, business transactions, and administration documents. And yet, *Ur* is not what we typically have in mind, when we speak of an information society. There may be many explanations, but one seems more convincing than any other: only very recently has human progress and welfare begun to be not just *related* to, but *mostly dependent* on, the successful and efficient management of the life cycle of information. I shall say more about such a cycle in the rest of this chapter, but, first, let us see why such a dependency has meant that we recently entered into *hyperhistory* (Figure 3).

Prehistory and history work like adverbs: they tell us *how* people live, not *when* or *where* they live. From this perspective, human societies currently stretch across three ages, as ways of living. According to reports about an unspecified number of uncontacted tribes in the Amazonian region[3], at the beginning of the second millennium there were still some societies that may be living prehistorically, without recorded documents. If (or rather when) one day such tribes disappear, the end of the first chapter of our evolutionary book will have been written.

The greatest majority of people today still live historically, in societies that rely on ICTs to record and transmit data of all kinds. In such historical societies, ICTs have not yet overtaken other technologies, especially energy-related ones, in terms of their vital importance. Then, there are some people around the world who are already living hyperhistorically, in societies and environments where ICTs and their data processing capabilities are the necessary condition for the maintenance and any further development of societal welfare, personal well-being, and overall flourishing. For example, all members of the G7 group—namely Canada, France, Germany, Italy, Japan, United Kingdom, and United States of America — qualify as information societies because, in each country, at least 70% of the Gross Domestic Product (GDP) depends on intangible goods, which are information-related, rather than on material goods, which are the physical output of agricultural or manufacturing processes. Their economies heavily rely on information-based assets (knowledge-based economy), information-intensive services (especially business and property services, communications, finance, insurance, and entertainment), and information-oriented public sectors (especially education, public administration, and health care).

[2] According to the French biologist Jean-Baptiste Lamarck (1744–1829), an organism could pass on to its offspring adaptive changes acquired through individual efforts during its lifetime. This pre-Darwinian theory is known as soft inheritance.
[3] Source: Survival for Tribal Peoples report, Uncontacted Amazon Indians face annihilation, 14 February 2011, available online.

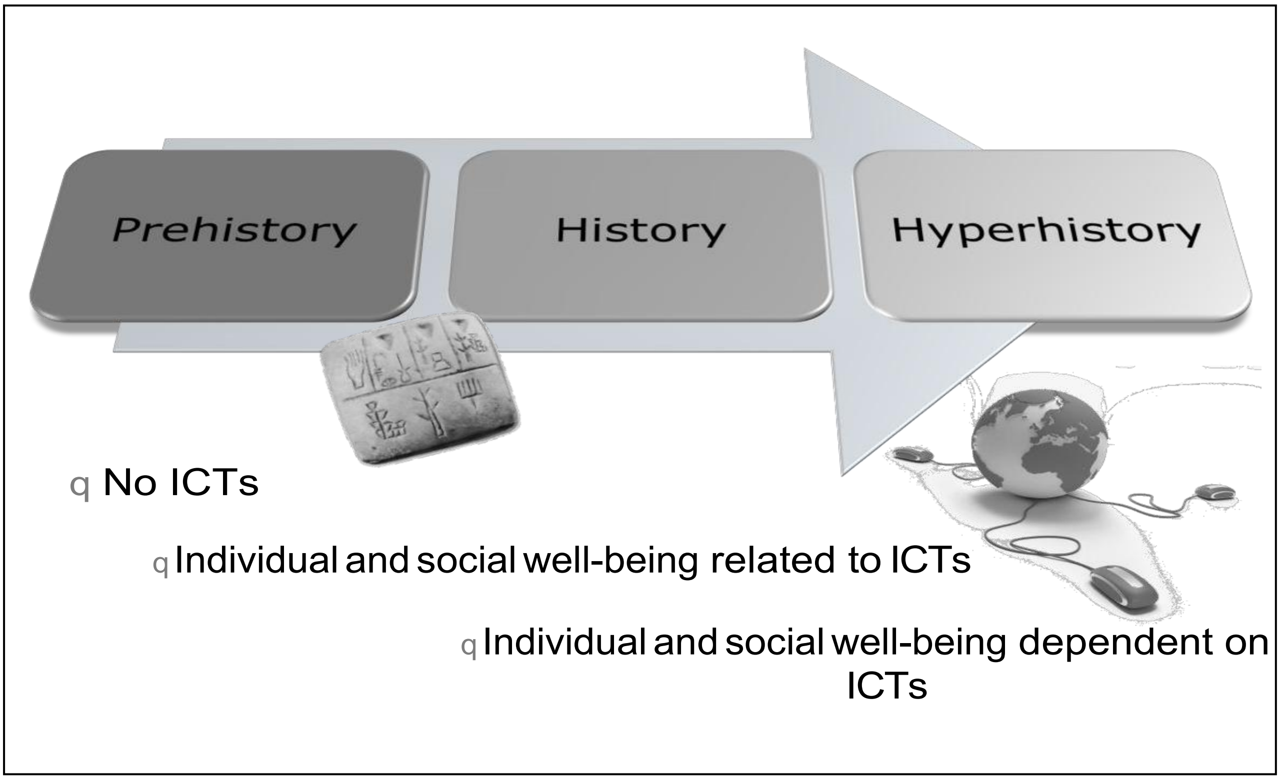

Figure 3. From Prehistory to Hyperhistory

The nature of conflicts provides a sad test for the reliability of this tripartite interpretation of human evolution. Only a society that lives hyperhistorically can be vitally threatened informationally, by a cyber-attack. Only those who live by the digit may die by the digit, as we shall see in chapter six.

Let us return to Ur. The reason why we do not consider Ur an information society is because it was historical but not yet hyperhistorical: it depended more on agricultural technologies, for example, than on clay tablets. Sumerian ICTs provided the recording and transmitting infrastructure that made the escalation of other technologies possible, with the direct consequence of furthering our dependence on more and more layers of technologies. However, ICTs' recording and transmitting facilities evolved into processing capabilities only millennia later, in the few centuries between Johann Gutenberg (ca. 1394/1399–1468) and Alan Turing (1912–1954). Therefore, it is only the present generation that is experiencing the radical transformations, brought about by ICTs, which are drawing the new threshold between history and hyperhistory.

The length of time that the evolution of ICTs has taken to bring about hyperhistorical information societies should not be surprising. The life-cycle of information (see Figure 4) typically includes the following phases: occurrence (discovering, designing, authoring, etc.), recording, transmission (networking, distributing, accessing, retrieving, etc.), processing (collecting, validating, merging, modifying, organising, indexing, classifying, filtering, updating, sorting, storing, etc.) and usage (monitoring, modelling, analysing, explaining, planning, forecasting, decision-making, instructing, educating, learning, playing, etc.). Now, imagine Figure 4 to be like a clock. According to recent estimates, life on Earth will last for another billion years, until it is destroyed by the increase in solar temperature. So, next, imagine an historian writing in the near future, say in a million years. She may consider it normal, and perhaps even elegantly symmetrical, that it took roughly six millennia for the agricultural revolution to produce its full effect, from its beginning in the Neolithic (10th millennium BC), until the Bronze Age, and then another six millennia for the information revolution to bear its main fruit, from the Bronze Age until the end of the 2nd millennium AD. She may find it useful to visualise human evolution as a three-stage rocket: in prehistory, there are no ICTs; in history, there are ICTs, they record and transmit information, but human societies depend mainly on other kinds of technologies concerning primary resources and energy; and in hyperhistory, there are ICTs, they record, transmit and, above all, process information, increasingly autonomously, and human societies become vitally dependent on them and on information as a fundamental resource in order to flourish. Around the beginning of the third millennium, our future historian may conclude, innovation, welfare and added value moved from being ICT-related to being ICT-dependent. She might guess that such a shift required unprecedented levels of processing power and huge quantities of data. And she might suspect that memory and connectivity must have represented some kind of bottlenecks. She would be right on both accounts, as we shall see in the rest of this chapter.

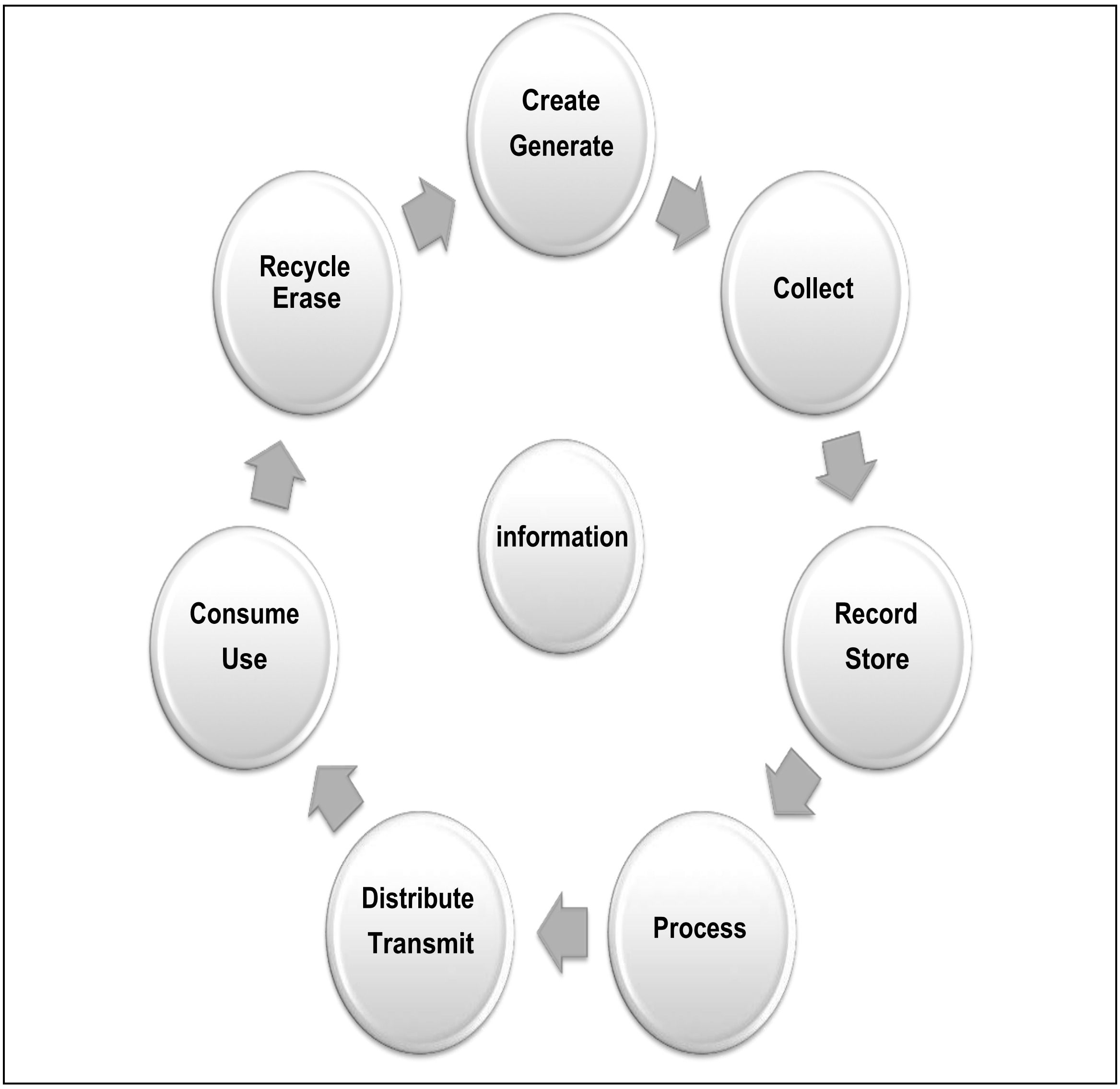

Figure 4. A typical information life cycle

MIPS

Consider the two following diagrams. Figure 5 is famous, almost iconic. It is known as Moore's Law and suggests that, over the development of digital computers, the number of transistors on integrated circuits doubles approximately every two years.

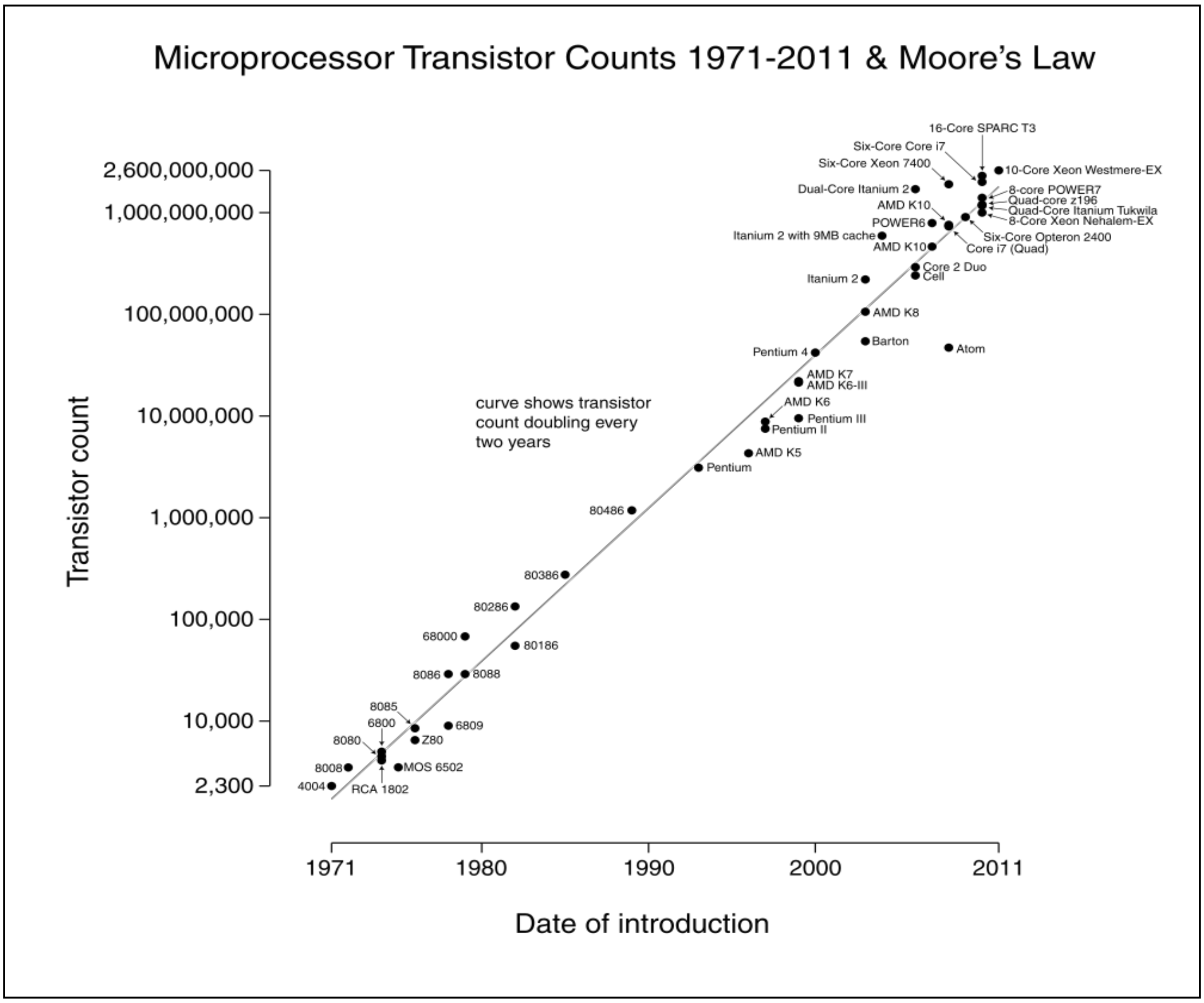

Figure 5. Moore's Law

Figure 6 is less famous but equally astonishing. It tells you a similar story, but in terms of decreasing cost of computational power. In 2010, an iPad2 had enough computing power to process 1600 millions of instructions per second (MIPS). By making the price of such a processing power equal to $100, the graph shows what it would have cost to buy the computing power of an iPad2 in the past decades. Mind that the vertical scale is logarithmic, so it descends by powers of ten as the price of computing power decreases dramatically. All this means that, in the fifties, the 1600 MIPS you hold in your hands—or rather did, in 2010, because three years later the iPad4 already run at 17056 MIPS—would have cost you $100 trillion. This is a number than only bankers and generals understand. So, for a quick comparison, consider Qatar's GDP. In 2010, it was ranked 57th out of 190 countries in the world and it would have been insufficient, for it was a mere $98 trillion.

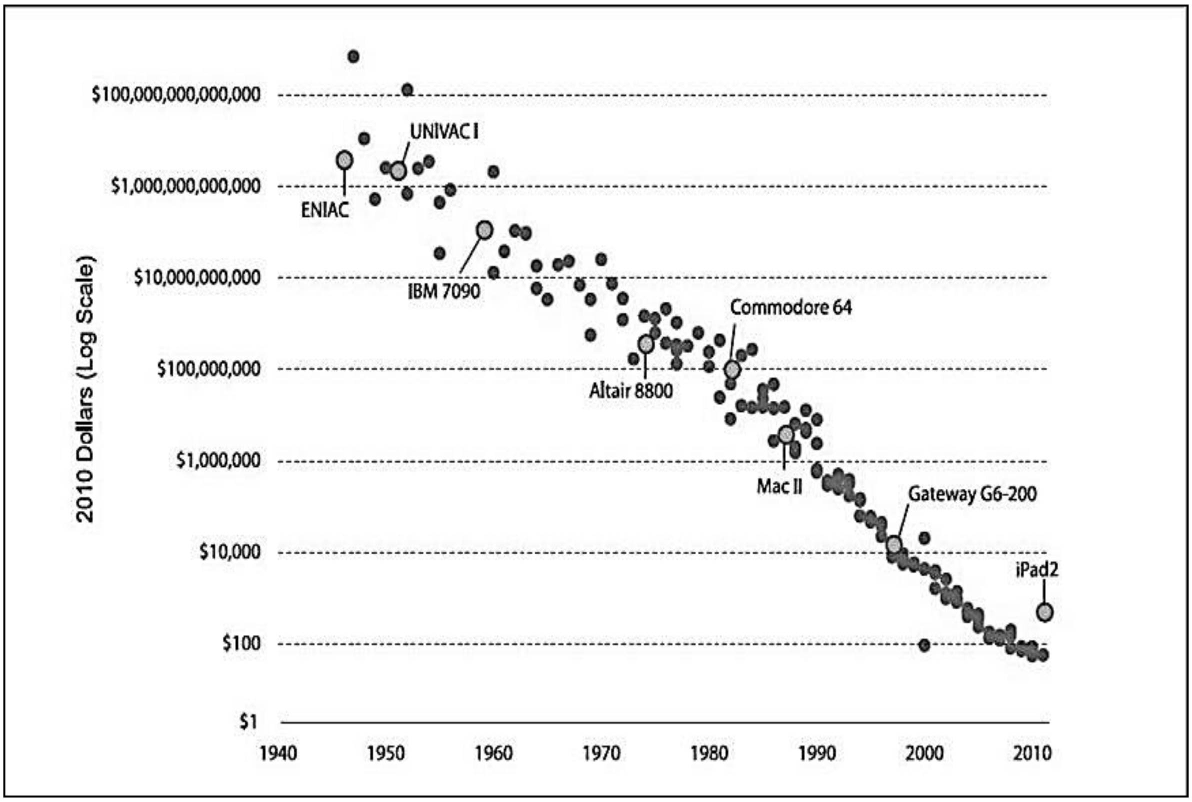

Figure 6. The cost of computing power equal to an iPad2. Source: The Hamilton Project at the Brookings Institution.

Whether you find Figure 5 or Figure 6 more compelling, the conclusion is the same: increasingly more power is available at decreasing costs, in quantities and at a pace that are mindboggling. The limits of computing power seem to be mainly thermodynamical: they concern how well our ICTs can dissipate heat and recover from unavoidable hardware faults while becoming increasingly small. This is the rocket that has made humanity travel from history to hyperhistory, to use a previous analogy. It also explains why ICTs are still disruptive technologies that have not sedimented: new generations keep teaching the old ones how to use them, although they learn from previous generations how to drive.

At this point, the obvious question is where all this computational power goes. It is not that we are regularly putting people on the moon with our smart phones or tablets. The answer is: interactions, both machine-to-machine (M2M) and human-computer ones (HCI).

In machine-to-machine interactions, an ICT system, such as a meter or sensor, monitors and records an event, such as the state of the road surface, and communicates the resulting data through a network to an application, which processes the data and acts on them, e.g. by automatically adapting the braking process of a car, if necessary. You might have heard that there is more computational power in an average, new car today than was available to NASA to send the astronauts to the Moon (Apollo mission, 1969). It is true. There are more than fifty ICT systems in an ordinary car, controlling anything from satellite navigation to hi-fi display, from ABS (anti-locking brakes) to electric locks, from entertainment systems to all the sensors embedded in the engine. It is a growing market in the automobile industry, as Figure 7 illustrates. According to Intel, the connected car is already the third fastest growing technological device after phones and tablets. It is only a matter of (short) time before all new cars will be connected to the Internet and, for example, find a convenient car park space, sense other vehicles, or spot cheaper petrol prices along the journey. Mechanics are becoming computer engineers.

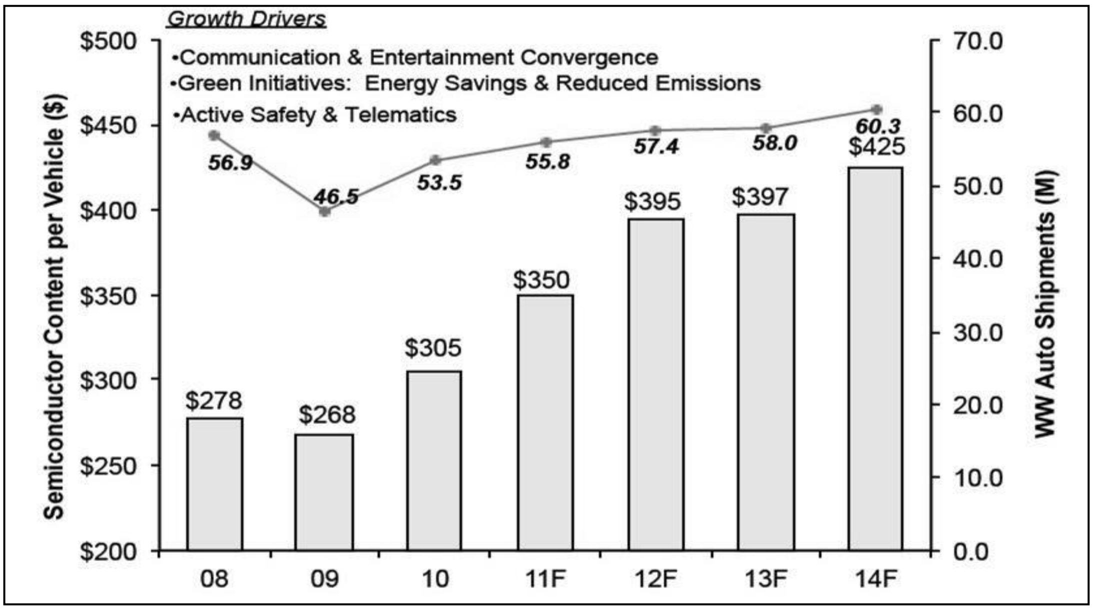

Figure 7. Value in US dollars of the average semiconductor content in automobiles. Source: IC Insights, 2011. Research Bulletin, June 2, 2011. F = forecast. WW = worldwide.

In human-computer interactions, ICTs are used to create, facilitate, and improve communications between human users and computational systems. When talking about ICTs, it is easy to forget that computers do not compute and telephones do not phone, to put it slightly paradoxically. What computers, smart phones, tablets and all the other incarnations of ICTs do is to handle data. We rely on their MIPS much less to add numbers or call our friends, than to update our Facebook status, order and read the latest e-books online, bill someone, buy an airline ticket, scan an electronic boarding pass, watch a movie, monitor the inside of a shop, drive to a place or, indeed, almost anything else. This is why HCI is so important. Indeed, since the mid-1990s, HCI does not even have to involve screens or keyboards, it may be a matter of a neuro-prosthetic device implanted in the brain. Of course, in all human-computer interactions, the better the process, the computationally greedier the ICT in question is likely to be. It takes a lot of MIPS to make things easy. This is the reason why new operating systems can hardly run on old computers.

We know that what our eyes can see in the world—the visible spectrum of the rainbow, so to speak—is but a very small portion of the electromagnetic spectrum, which includes gamma rays, X-rays, ultraviolet, infrared, microwaves, and radio waves. Likewise, the data processing 'spectrum' that we can perceive is almost negligible compared to what is really going on in M2M and HCI interactions. An immense number of ICT applications run an incalculable number of instructions every millisecond of our lives to keep the hyperhistorical information society humming. ICTs consume most of their MIPS to talk to each other, collaborate, and coordinate efforts, and put us as comfortably as possible in or on the loop, or even out of it, when necessary. Almost all MIPS are transparent to us, in the sense that we do not see them, like the oxygen we breathe, but they are becoming almost as vital, and they are growing exponentially. In doing all this, they generate a staggering amount of data, much more data than humanity has ever seen in its entire history (Figure 8). This is the other resource that has made hyperhistory possible: zettabytes.

Zettabytes

Just a few year ago, researchers at Berkeley's School of Information (Lyman & Varian (2003)) estimated that humanity had accumulated approximately 12 exabytes[4] of data in the course of its entire history until the commodification of computers, but that it had already reached 180 exabytes by 2006. According to a more recent study (Gantz & Reinsel (2011)), the total grew to over 1,600 exabytes, between 2006 and 2011, thus passing the zettabyte (1,000 exabytes) barrier. This figure is now expected to grow fourfold approximately

4 One exabyte corresponds to 1018 bytes or a 50,000 year-long video of DVD quality.

every three years, so that we shall have 8 zettabytes of data by 2015. Every day, enough new data is being generated to fill all US libraries eight times over. Of course, trillions of ICT systems are constantly working to keep us afloat and navigate through such an ocean of data. These are all numbers that will keep growing quickly and steadily for the foreseeable future, especially because those very systems are among the greatest sources of further data, which in turn require, or simply make possible, more ICTs. It is a self-reinforcing cycle and it would be unnatural not to feel overwhelmed. It is, or at least should be, a mixed feeling of apprehension for the risks, excitement for the opportunities, and astonishment for the achievements, as we shall see in the following chapters.

Thanks to ICTs, we have entered *the age of the zettabyte*. Our generation is the first to experience a Zettaflood, to introduce a neologism to qualify this tsunami of bytes that is submerging our environments. In other contexts, this is also known as 'big data' (Figure 8).

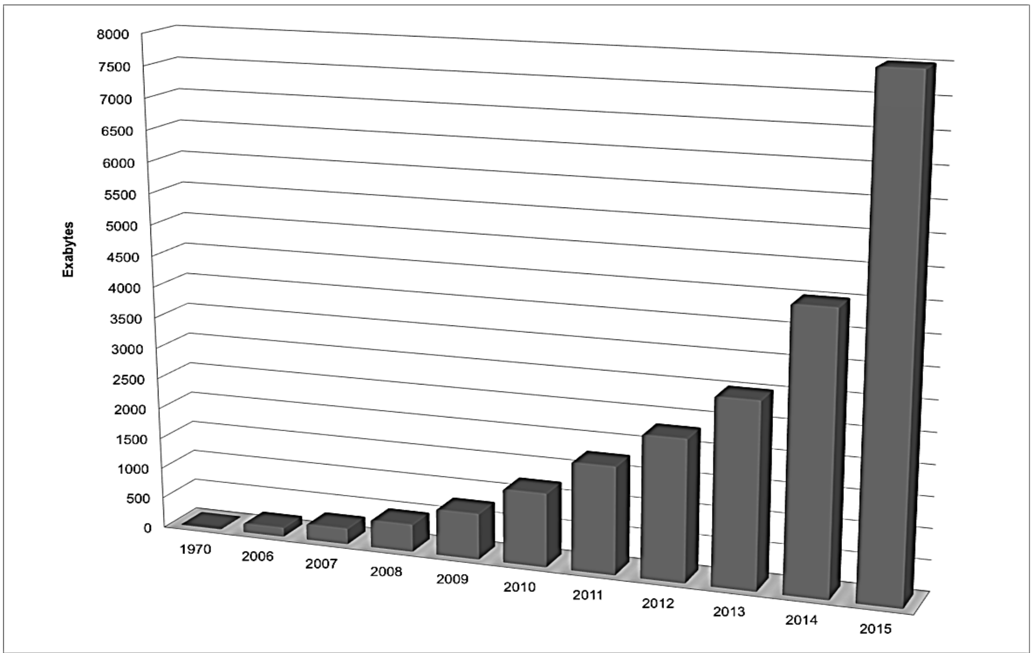

Figure 8 The Growth of Big Data. Based on IDC White Paper, 'The Diverse and Exploding Digital Universe', March 2008 and IDC White Paper 'Worldwide Big Data Technology and Service 2012-2015 Forecast', March 2012.

Despite the importance of the phenomenon, it is unclear what exactly the term 'big data' means and hence refers to. The temptation, in similar cases, is to adopt the approach pioneered by Potter Stewart, United States Supreme Court Justice, when asked to describe pornography: difficult to define, but 'I know when I see it'. Other strategies have been much less successful. For example, in the United States, the National Institutes of Health (NIH) and the National Science Foundation (NSF) have identified big data as a program focus. One of the main NSF- NIH interagency initiatives addresses the need for core techniques and technologies for advancing big data science and engineering. However, the two agencies specify that:

> *The phrase 'big data' in this solicitation refers to large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future. NSF-12-499, available online.*

You do not need to be a logician to find this both obscure and vague. Wikipedia, for once, is also unhelpful. Not because the relevant entry is unreliable, but because it reports the common definition, which is unsatisfactory:

> *a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. Wikipedia, 'Big Data', version dated 17 February 2013.*

Apart from the circular problem of defining 'big' with 'large' (the NSF and NHI also seem to be happy with it), the definition suggests that data are too big or large only in relation to our current computational power. This is misleading. Of course, 'big', as many other terms, is a relational predicate: a pair of shoes is too big for you, but fine for me. It is also trivial to acknowledge that we tend to evaluate things non-relationally, in this case as absolutely big, whenever the frame of reference is obvious enough to be left implicit. A horse is a big animal; no matter what whales may think. Yet these two simple points may give the impression that there is no real trouble with 'big data' being a loosely defined term referring to the fact that our current computers cannot handle so many gazillions of data efficiently. And this is where two confusions seem to creep in. First, that the epistemological (that is, knowledge related) *problem* with big data is that there is too much of it (the *ethical problem* concerns how we use them, more on this presently). And, second, that the *solution* to the epistemological problem is *technological*: more and better techniques and technologies, which will 'shrink' big data back to a manageable size. The epistemological problem is different, and it requires an equally epistemological solution.

Consider the problem first. 'Big data' came to be formulated after other buzz expressions, such as 'infoglut' or 'information overload', began to fade away, yet the idea remains the same. It refers to an overwhelming sense that we have bitten off more than we can chew, that we are being forced-fed like geese, that our intellectual livers are exploding. This is a mistake. Yes, we have seen that there is an obvious exponential growth of data on an ever-larger number of topics, but complaining about such over-abundance would be like complaining about a banquet that offers more than we can ever eat. Data remain an asset, a resource to exploit. Nobody is forcing us to digest every available byte. We are becoming data-richer by the day; this cannot be the fundamental problem.

Since the problem is not the increasing wealth of data that is becoming available, clearly the solution needs to be reconsidered: it cannot be merely how many data we can technologically process. We saw that, if anything, more and better techniques and technologies are only going to generate more data. If the problem were too many data, more ICTs would only exacerbate it. Growing bigger digestive systems, as it were, is not the way forward.

The real, epistemological problem with big data is *small patterns*. Precisely because so many data can now be generated and processed so quickly, so cheaply, and on virtually anything, the pressure both on the data *nouveau riche*, such as Facebook or Walmart, Amazon or Google, and on the data *old money*, such as genetics or medicine, experimental physics or neuroscience, is to be able to spot where the new patterns with real added-value lie in their immense databases, and how they can best be exploited for the creation of wealth, the improvement of human lives, and the advancement of knowledge. This is a problem of brain rather than computational power. Small patterns matter because, in hyperhistory, they represent the new frontier of innovation and competition, from science to business, from governance to social policies, from security to safety. In a Baconian open market of ideas, if someone else can exploit them earlier and more successfully than you do, you might quickly be out of business, miss a fundamental discovery and the corresponding Nobel, or put your country in real danger.

Small patterns may also be risky, because they push the limit of what is predictable, and therefore may be anticipated, about not only nature's, but also people's, behavior. This is an ethical problem. Target, an American retailing company, relies on the analysis of the purchasing patterns of 25 products in order to assign each shopper a 'pregnancy prediction' score, estimate her due date, and send coupons timed to specific stages of her pregnancy. In a notorious case, it caused some serious problems when it sent coupons to a family in which the teenager daughter had not informed her parents about her new status. I shall return to this sort of problems in chapters three and four, when discussing personal identity and privacy.

Unfortunately, small patterns may be significant only if properly aggregated, e.g. in terms of loyalty cards and shopping suggestions, compared, as when banks utilize big data to fight fraudsters, and processed in a timely manner, as in financial markets. And because information is indicative also when it is not there (the lack of some information may also be informative in itself), small patterns can also be significant if they are absent. Sherlock Holmes solves one of its famous cases because of the silence of the dog, which should have barked. If big data are not 'barking' when they should, something is going on, as the financial watchdogs (should) know.

The increasingly valuable undercurrents in the ever-expanding oceans of data are invisible to the computationally-naked eye, so more and better techniques and technologies will help significantly. Yet, by themselves, they will be insufficient. Which is just another way of saying that, because the problem with big data is small patterns, ultimately, the game will be won by those who ‘know how to ask and answer questions’ (Plato, *Cratylus*, 390c), and therefore know which data may be useful and relevant, and hence worth collecting and curating, in order to exploit their valuable patterns. We need more and better techniques and technologies to see the small data patterns, but we need more and better epistemology to sift the valuable ones. So new forms of education are part of the challenge, as we shall see in chapter three.

Big data is here to grow. The only way of tackling it is to know what you are or may be looking for. We do not do science by mere accumulation of data, we should not do business and politics in that way either. At the moment, the required epistemological skills are taught and applied by a black art called analytics. Not exactly your standard degree at the University. Yet so much of our well-being depends on it that it might be time to develop a methodological investigation of it. Who knows, philosophers might not only have something to learn, but also a couple of lessons to teach. Plato would agree. What he might have been disappointed about is the fact that memory is no longer an option. Mere data hoarding, while waiting for more powerful computers, smarter software, and new human skills, will not work, not least because we simply do not have enough storage. Recall our future historian: this is the first bottleneck she identified in the development of hyperhistory, which suffers from digital amnesia.

Digital Amnesia

Hyperhistory depends on big data, but there are two myths about the dependability of digital memory that is important to expose in this first chapter.

The first myth concerns the *quality* of digital memory. ICTs have a kind of forgetful memory: they become quickly obsolete, they are volatile, and they are re-recordable. Old digital documents may no longer be usable because the corresponding technology (floppy drivers, processing software etc.) is no longer available. There are millions of abandoned pages on the Internet (pages that have been created and then not updated or modified). At the beginning of 1998, the average life of a document that had not been abandoned was 75 days. It is now considered to be 45 days. The outcome is that so-called link decay (lost Internet references) is a common experience. On 30 April 1993, the European Organization for Nuclear Research (CERN) announced that the World Wide Web it had created would be free to everyone, with no fees due. Twenty years later, to celebrate the event, a team at CERN had to re-create the first web page (with its original URL etc.), because it no longer existed. Our digital memory seems as volatile as our oral culture was but perhaps even more unstable, because it gives us the opposite impression. This paradox of a digital ‘prehistory’—ICTs are not preserving the past for future consumption because they make us live in a perennial present—will become increasingly pressing in the near future. Memory is not just a question of storage and efficient management; it is also a matter of careful curation of significant differences and hence of the stable sedimentation of the past as an ordered series of changes, two historical processes that are now seriously at risk. Ted Nelson, for example, a pioneer in ICTs who coined the terms ‘hypertext’ and ‘hypermedia’, designed Xanadu so that it would never delete copies of old files. A web site constantly upgraded is a site without memory of its own past, and the same dynamic system that allows one to rewrite a document a thousand times also makes it unlikely that any memory of past versions will survive for future inspection. ‘Save this document’ means ‘replace its old version’, and every digital document of any kind may aspire to such an ahistorical nature. The risk is that differences are erased, alternatives amalgamated, the past constantly re-written, and history reduced to the perennial here and now, *hinc et nunc*. When most of our knowledge is in the hands of this forgetful memory, we may find ourselves imprisoned in a perpetual present. This is why initiatives such as the National Digital Stewardship Alliance (NDSA) and the International Internet Preservation Consortium (IIPC) are vital, in order to preserve our increasingly digital cultural heritage for future generations. The job of information curators is bound to become increasingly important.

There is then the potentially catastrophic risk of immense quantities of data being created simultaneously. We saw that most, indeed almost all our data have been created in a matter of a few years. They are all getting old together, and will reach the threshold of system failure together, like a baby-boom generation retiring at the same time. To understand the problem, recall the old debate about your collection of music CDs and how they would all be unusable within a decade, as opposed to your vinyl records. A MTBF (mean time before failure) figure indicates an estimate of a system's life expectancy. The higher the MTBF, the longer the system should last. CDs used to have a MTBF of about ten years, after which their material degraded too much to guarantee usability. The same applies to our current digital supports, hard disk and memories of various kinds. A MTFB of 50,000 hours (5.7 years) for standard hard disk is rather common. This short life expectancy is already a problem. What makes it even more dangerous is that, contrary to what we experienced in the past, MTBFs are today dangerously synchronized. Big Data will age and become Dead Data together. Clearly, huge quantities of data will need to be re-recorded and transferred to new supports at regular intervals. Indeed, they already are. But which data are going to make it to the other side of any technological transition? For a comparison, think of the transition of silent movies to new kinds of support, or of recorded music from vinyl to the CD. Huge quantities of data were left behind, becoming lost, unavailable, or inaccessible.

According to a 2012 Research Report by IBIS World, the data recovery industry saw its overall revenue over the five years to 2012 fall at an annualized rate of 0.9% to total $1 billion, with a decline of 0.6% in 2012[5]. This may seem counterintuitive. Big Data is growing and so are the problems concerning damaged, corrupted or inaccessible files and storage media. The industry that takes care of such problems should be flourishing. The explanation is that cloud or online storage has expanded the options for data recovery and data loss prevention. If you use Dropbox, for example, and your computer is damaged, the files are still available online and can be easily recovered, so you will not need a data recovery service. Yet, this seems to be just a question of transition and hence time. Cloud computing has put pressure on an industry specialised on computers at a consumer level. The more our gadgets become mere terminals, the less we need to worry ourselves about the data. But the storage of those data still relies on physical infrastructures, and these will need increasing maintenance. The data recovery industry will disappear, but a new industry dedicated to cloud computing failures is already emerging. It is not a matter of relying on the brute force of redundancy (having more than one copy of the same file). This strategy is not available at a global level, because of the second myth about the dependability of digital memory, the one concerning the *quantity* of digital memory.

Since 2007, the world has been producing many more data than available storage (Figure 9). This despite the fact that, according to Kryder's Law (another generalisation), storage density of hard disks is increasing more quickly than Moore's Law, so that it is predicted that, in 2020, a 14 terabytes disk will be 2.5 inches in size and will cost about $40. Unfortunately, this will not be enough, because even the growth forecasted by Kryder's Law is slow when compared to the pace at which we generate new data. Think of your smart phone becoming too full because you took too many pictures, and make it a global problem. In history, the problem was what to save: which laws or names were going to be baked in clay or carved in stone, which texts were going to be handwritten on papyrus or vellum, which news items were worth printing on paper. In hyperhistory, saving is the default option, so the problem becomes what to erase. Since storage is insufficient, something must be deleted, re-written, or never be recorded in the first place. Again, by default the new tends to push out the old, or 'first in first out': updated webpages erase old ones, new pictures make old ones look dispensable, new messages are recorded over old ones, recent emails are kept at the expense of last year's.

[5] Source: IBIS World, Data Recovery Services Market Research Report, July 2012, available online.

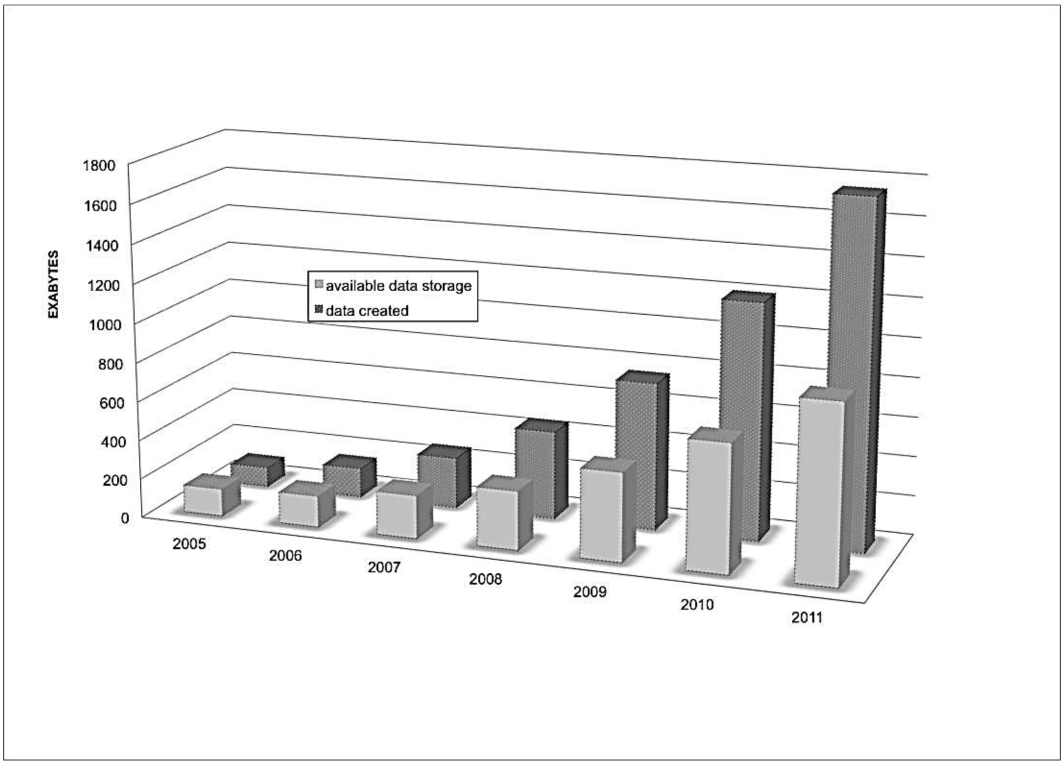

Figure 9. Data vs. Memory. Based on IDC White Paper, 'The Diverse and Exploding Digital Universe', March 2008; IDC White Paper 'Worldwide Big Data Technology and Service 2012- 2015 Forecast', March 2012; and 'Data, data everywhere'; and The Economist, 25 February, 2010.

Hyperhistory ran out of memory space in which to dump its data many years ago. There is no name for this 'law' about the increasing shortage of memory, but it looks like the gap is doubling every year. Barring some significant technological breakthrough in physical storage or software compression, the process will get worst, quantitatively. The good news is that it does not have to be as bad as it looks, qualitatively. Rephrasing a common saying in the advertising world, half of our data is junk, we just do not know which half. You are happy to take ten pictures because you hope one will be decent and the other nine can be discarded. They were never intended to be saved in the first place. This means that we need a much better understanding of which data are worth preserving. This, in turn, is a matter of grasping which questions are, or will be, interesting to ask not only now, but also in the future, as we saw in the previous section. And this leads to a slightly reassuring virtuous circle: we should soon be able to ask big data what data is worth saving. Think of an app in your smart phone not only suggesting which of the ten pictures is worth keeping, but also learning from you, once you have taken a decision (maybe you prefer darker pictures). Then new challenges will concern how we may avoid poor machine-based decisions, improve so-called 'machine learning', or indeed make sure machines re-learn new preferences (later in life you may actually like brighter pictures). More information may help us to decide which information to save and curate. Our future historian may well interpret the zettabyte age of hyperhistory as the time of transition between blind and foresighted big data.

So much for the first bottleneck: shortage of memory. To understand the other, we now need to look at some features of networks.

Hyperconnectivity

Computers may be of limited value if they are not connected to other computers. This was not always obvious, sometimes it is still questionable—as when your computers need to be hacker- proof because they control the launch of nuclear missiles, for example—but, in general, the observation is rather trivial today. When it is, the question becomes: what is the value of connectivity? Many theories and laws have been proposed: Reed's Law, Sarnoff's Law, Beckstrom's Law… but the most famous remains Metcalfe's. Like the laws just mentioned and Moore's Law, it is a generalization ('this is how things tend to go, more or less') rather than a scientific law, but it is, nevertheless, enlightening. It states that the value of a network is proportional to the square of the number of connected nodes of the system (n2). So a network of 2 computers has only a value of 22 = 4, but doubling the number of connected computers already means quadrupling the value of their network to 42 = 16. Figure 10 illustrates what happens after 20 iterations. The idea is simple: the more nodes you have, the more useful it is to be connected and expensive to be unconnected. Indeed, the point to keep in mind is even simpler, for there is an even more inclusive generalization. Any growth bigger than linear (a linear growth is when you multiply x by a fixed number, like your salary by 12 months), e.g. squared, like Metcalfe's, or cubic (n3), or exponential (2x), after a few iterations looks like a straight perpendicular line, like a capital L which has been rotated 180 degrees on its axis: . This 'L law' is all one needs to remember. It is the shape of growth that any business would like to achieve. According to a report by the International Telecoms Union (ITU)[6], in 2013 more than a third of the world population was online. No wonder the value of the network has skyrocketed, straight like a rod. So what is the problem? Any L Law does not really address the communication within the network, but rather the value of its complexity (how many links are possible among how many nodes). Communication requires a link but it comes with a speed. Think of a road, and the difference it makes whether it is a small street or a motorway, with or without traffic. This is the bottleneck our future historian identified. It is known as Nielsen's Law.

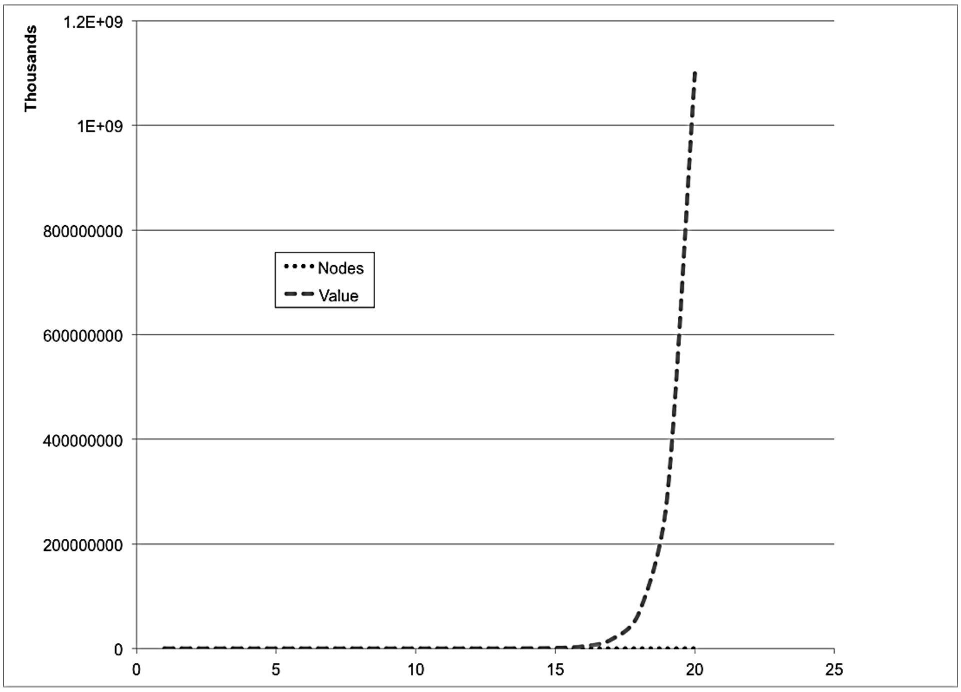

Figure 10. Metcalfe's Law: the value of a network of n nodes = n2.

Some years ago, Jacob Nielsen noticed that, in general, the speed of network connections for home users like you and me increases approximately 50% per year, thus doubling every 21 months or so. This is impressive, but not as impressive as the speed identified by Moore's Law. It is also already insufficient to cope with the faster growing 'weight' (number of bits) of the files we wish to transfer. As a result, for the foreseeable future our online experience will be constrained by our bandwidth.

[6] Source: ICT Data and Statistics Division, Telecommunication Development Bureau, International Telecommunication Union, The World in 2013, ICT Facts and Figures, available online.

Conclusion: Towards an Hyperhistorical Environment

The living generation is experiencing a transition from history to hyperhistory. Advanced information societies are more and more heavily dependent on ICTs for their normal functioning and growth. Processing power will increase, while becoming cheaper. The amount of data will reach unthinkable quantities. And the value of our network will grow almost vertically. However, our storage capacity (space) and the speed of our communications (time) are lagging behind. Hyperhistory is a new era in human evolution, but it does not transcend the spatio- temporal constraints that have always regulated our life on this planet. The question to be addressed next is: given all the variables we have seen in this chapter, what sort of hyperhistorical environment are we building for ourselves and for future generations? The answer is: the infosphere. This is the topic of the next chapter.

Further Readings

Claude Shannon (1916– 2001) is the father of information theory. His seminal work, Shannon & Weaver (1949 rep. 1998), requires a good background in mathematics and probability theory. A very accessible introduction to information theory is still Pierce (1980). I have covered topics discussed in this chapter in Floridi (2010a), where the reader can also find a very simple introduction to information theory. Gleick (2011) is an enjoyable 'story' of information. Ceruzzi (2012) provides a short introduction to the history of computing, from its beginning to the internet. Caldarelli & Catanzaro (2012) give a very short introduction to networks. On big data, a good survey is O'Reilly Media (2012), the Kindle edition is free. On the postmodern society, Lyotard (1984) is an essential reading, philosophically demanding but also rewarding. On the network society, Manuel Castells (2000), the first volume of his trilogy, has shaped the debate. The information society produces much information about itself. Among the many valuable, yearly reports, freely available online, one may consult: *Measuring the Information Society*, which includes the ICT Development Index (IDI) and the ICT Price Basket (IPB), two benchmarks useful for monitoring the development of the information society worldwide; the *Global Information Technology Report*, produced by the World Economic Forum in cooperation with INSEAD, covers 134 economies worldwide and is acknowledged to be the most comprehensive and authoritative international assessment of the impact of ICT on countries' development and competitiveness; the *International Telecommunication Union Statistics*, which collects data about connectivity and availability of telecommunication services worldwide; and the *Digital Planet* report, published by Published by World Information Technology and Service Alliance (WITSA), contains projections on ICT spending. Finally, Brynjolfsson & McAfee (2011) analyse how ICTs affect the job market, transform skills, and reshape the evolution of human labour. They do so from an American perspective, but their insights are universal, the scholarship admirable, and the recommendations on how machines and humans may collaborate quite convincing.

Reference:
Chapter 1 – Time/Hyperhistory. (2016). In L. Floridi, *The Fourth Revolution* (pp. 16-41). Oxford, New York, USA: Oxford University Press . Retrieved February 9, 2018, from http://larvalstage.i2p.xyz/dox/tfr.pdf