Information content of human intelligence and life

Warren D. Smith*
warren.wds@gmail.com

June 5, 2006

Abstract — **Regard various standard approximate biochemical assumptions as correct. Then we show that the proteome of human cells constitutes an "information bottleneck." We use this to estimate the information content of a "blueprint" sufficient for efficiently building a human intelligence (as opposed to some unintelligent animal). We find it is only about 6 or 81 Kbytes (in two oppositely-extreme models), with 4 times these figures being proposed as a high confidence upper bound. This seems surprisingly small compared to many software attempts to create artificial intelligences.**

However, in the companion paper we create a "mathematical definition of intelligence," use it to prove a theorem that a "universally asymptotically competitive intelligence" (UACI) exists, and show how to implement it as a remarkably short program. (Although mathematically a success, this UACI is not a success in any engineering sense because it will take a very long time before doing anything interesting.) Also Fong's PAPPI human language principles package and Baum & Durdanovic's "Hayek" AI-via-artificial-economy code both lie between these two numbers. Hence this *perhaps* is not really a paradox despite initial appearances.

The tasks of creating an AI and of creating artificial life may share the similarity of requiring a very large search over configurations to find a good one.

1 Overview

We begin in §2 by looking at the "information content" of the genomes of various lifeforms, noting in passing the "creation problem" that the simplest life we know of is far more complicated than the first life must have been. In §3 we focus in on *Homo sapiens* and especially on its mind and brain. There is an "information pipeline" – or perhaps a better analogy is "bucket brigade" – which starts at your DNA, moves through transfer RNAs to messenger RNAs, moves through proteins, and further on that information is used to direct the cellular and multicellular development of the human form, and to guide the development of the brain, which then transmits, receives, stores, and processes electrical information, and creates the audio and written communications and physical acts and creations that constitute human culture – which is the final stage in the information pipeline.

We can (and do) estimate the information content – bit counts – at each particular stage in this information pipeline. Each of these estimates is necessarily an *over*estimate of the true information content. However, the *bottleneck* in the information pipeline – the bucket with the *smallest* raw bit count – gives us the least overestimation and hence the best estimate.

Most of these bit counts arise easily from published literature and simple arithmetic. But one exception is the number of bits in human memory – a poorly understood and badly measured quantity that we review in §4. That review ceases to affect our main argument as soon as it establishes that human memory has far larger information capacity than the bottleneck. But our review is fairly extensive nevertheless, because we wish to make clear the extremely poorly done nature of human memory research and some easy ways to improve it tremendously.

Our main purpose is to make it clear that the human information bottleneck is the *proteome*. Under two oppositely-extreme models, this bottleneck constitutes either **2.4 or 32 megabits of raw information** (both numbers are estimates believed to be within a factor of 3).

Now only a small fraction of this information – we estimate 2% – constitutes the "blueprint for human intelligence" that causes humans to be "intelligent" as opposed to other "unintelligent" animals.

The result is an estimated upper bound on **the Kolmogorov complexity of human intelligence: either 6 or 81 Kbytes** (in the two models, each estimate accurate to a multiplicative factor of $\lesssim 4$).

This may seem impressively small. In §5 we compare these numbers with the known code sizes of a list of computer programs – including some designed to have especially small code-lengths. All these programs of course would seem to accomplish far less than human intelligence – most observers feel that no computer has yet achieved "intelligence."

However, in a companion paper [37] we provide a "mathematical definition of intelligence" and show how to construct "universally intelligent" programs, called "UACIs," with very small code-lengths. Similar ideas had been invented independently a few years earlier by Marcus Hutter [17]. Those UACIs would produce interesting outputs only after an enormous delay and hence would be too slow to to be practically useful – but they would meet the technical definition of "intelligence." The companion [37] also outlines an engineering-research program aiming to produce more-practical universal intelligences, and it is extremely unclear how difficult that

*Non-electronic mail to: 21 Shore Oaks Drive, Stony Brook NY 11790.

improvement task will be. That is where the numerical estimates of this paper come in: they provide at least some rough bounds. This entire paper is focused simply on producing the numbers in bold print above.[1]

2 The information content of life

lifeform	#bits	#genes
Psilotum nudum (whisk fern plant)	5×10^{11}	
Lily plant	2.4×10^{11}	
Lungfish	2×10^{11}	
Wheat plant	3×10^{10}	60000?
Homo Sapiens (Human)	6.5×10^{9}	23000
Mus Musculus (Mouse)	6.0×10^{9}	27000
Maize plant	5×10^{9}	
Danio rerio (Zebrafish)	3.4×10^{9}	25000
Xenopus Tropicalis (Frog)	3.0×10^{9}	20000
Gallus Gallus (Chicken)	2.1×10^{9}	19000
Tomato plant	1.9×10^{9}	
Strongylocentrotus purpuratus (Purple Sea Urchin)	1.7×10^{9}	?
Takifugu rubripes (pufferfish)	7.8×10^{8}	23000
Apis mellifera (Honeybee)	4.5×10^{8}	14000
Drosophila Melanogaster (Fruit fly)	2.6×10^{8}	14000
Arabidopsis thaliana (flowering weed)*	2.5×10^{8}	25500
Caenorhabditis Elegans (tiny transparent roundworm)	2.01×10^{8}	20000
Saccharomyces Cerevisiae (Budding Yeast, eukaryote)	2.4×10^{7}	7000
Plasmodium falciparum (Malaria)	4.6×10^{6}	5300
Escherichia coli (bacterium)	9.4×10^{6}	3200
Synechococcus WH8102(2) (motile photosynthesizing plankton) [28]	4.8×10^{6}	2050
Prochlorococcus SS120 (smaller non-motile phytoplankton)	3.5×10^{6}	1694
Mycoplasma genitalium (bacterium)	1160148	500 ± 30
Nanoarchaeum equitans (archaeon microbe)*	981770	544 ± 8
Coronavirus	59502	
DNA viruses generally	10000 to 500000	10 to 900
RNA viruses generally	3400 to 120000	1 to 25
L1 retrotransposon inside human genome	13000	3
ΦX-174 (bacteriophage infecting E.Coli)	10772	10
Hepatitis B virus (single & double-standed DNA mix)	4800	4
Alu replicating element inside human genome	600	0
Viroids, satellite viruses, and virusoids generally	500 to 750	0 or 1
Spiegelman's RNA "monster"	440	0 or 1
More-evolved Spiegelman monster [10]	108+96=204	0 or 1

Figure 2.1. Genome sizes. #bits(= 2×#base-pairs) given; #genes are usually estimates from `www.ensembl.org`. The wheat gene count is a crude extrapolation based on partial sequence data collected so far. * denotes current record-holder for smallest-known genome among plants, free-living microbes. Many organisms got on this list because they have atypically short genomes. Prochlorococcus is by far the most abundant self-sufficient organism ($\approx 95\%$) and Synechococcus is the second most abundant. Both are cyanobacteria. Hepatitis B is an unusually small virus with a DNA-loop that is partially single stranded and partally double stranded; it codes for 4 genes, all of which overlap on the DNA to save space. L1 is a "selfish DNA" element capable over evolutionary time of inserting further copies of itself into the human genome (which currently contains about 90 copies). (Also with help from retroviruses such selfish DNA elements can spread to other species.) Alu is a much simpler stretch of selfish DNA which "parasites" off L1's genes to copy itself at a considerably faster rate than L1, but which could not "continue to reproduce" without L1. There are over 1 million Alu copies in the human genome representing about 11% of our total DNA. I speculated elsewhere [36] that the vast differences in repetitive "junk DNA" quantity in the genomes of various apparently similar organisms is caused by runaway dynamics among, and DNA "garbage" produced as a side effect of, such "replicons." "Satellite viruses" are viruses that can only reproduce in the presence of an associated larger genuine virus. A "viroid" is a "small infectious pathogen composed entirely of a low molecular weight naked RNA molecule." They don't act as messenger RNAs, i.e. don't make the cell synthesize enzymes: they rely completely on pre-existing enzymes

[1] Our techniques can also be used to get small upper bounds on the description-complexity of the proteomic-metabolic dynamical systems of simple *bacteria*. However, this is not of great interest for casting light on the "creation problem," because for that purpose what matters is not only the description length of the abstract dynamical system, but *also* the description length of the physical machinery (i.e. protein sequences) needed to implement it. In contrast our work *does* cast useful light on the nature of intelligence, because there it really is an interesting question to determine the descriptive complexity of an *abstract* system to instantiate an "intelligent being."

in the host for reproduction. Spiegelman monsters are RNA molecules that replicate quickly in vitro in the presence of a certain viral RNA polymerase and xTP. ▲

Table 2.1 lists the genome sizes of several lifeforms. The current smallest-genome record holders among free-living organisms are *Mycoplasma genitalium* and *Nanoarchaeum equitans* each at[2] about 10Mbits.

But this is 4640 times larger than the maximum conceivable irreducible description length for the first living organism on Earth (250 bits) in view of probability, the $<$ 1Gyear time available, and the number of atoms on Earth, and avoiding any supernatural explanation[3] Even if the relevant numbers instead are regarded as the atom-count and age of the entire observable *universe*, still at most 360 bits of description length would seem allowable!

That is an immense paradox that anybody must tackle to explain the creation of life. There will be no fully satisfactory resolution until ultrasimple life is created in the lab or discovered.

The presumed (and hence unsatisfactory) explanation is that any initial small-description-length life might have had very slow reproduction rates (10000 years?) or in some other ways have had very poor Darwinian fitness compared to later large-description-length life. Hence the later life killed it. (And even if some still exists, it might be very hard to notice.)

But in suitably "more-friendly environments" than just naked exposure to the elements on Earth, life with much shorter description lengths is known to be possible. ΦX-174 phage "lives" in the "environment" consisting of *E.Coli* host organisms and their surrounding medium. Its raw genome length is only 10772 bits, and assuming the same data-compression factors discussed in footnote 2 are possible would reduce this to 7750 or perhaps 2940 bits, i.e. "only" 31 or perhaps 12 times larger than the 250-bit-threshold.

This can be regarded as "cheating,"[4] though, because phages cannot survive in the absence of host organisms; they take over their preexisting reproductive and synthetic machinery for their own purposes. For this reason it is dubious to try to use phages as models for Earth's initial life forms. Other things which "cheat" to an even larger degree can be even smaller, e.g. "satellite viruses" which are parasites on viruses which are parasites on cells.

3 The proteomic "information bottleneck"

Consider the development of a human being starting from a single cell as an "information pipeline" or "bucket brigade." We are going to number the buckets in the chain 1 to 10.

Initially (almost) all the information is contained in (1) our DNA molecules. The stages in the information pipeline are as follows: starting from (1) DNA, information flows through (2) transfer RNAs into (3) messenger RNAs into (4) ribosomes which synthesize proteins.

Now (5) proteins interact with both each other, the DNA and RNA, and all metabolic chemicals and raw materials (such as ions, sugars, lipids, vitamins) and the cell's physical structure (membranes and transport networks) to self-regulate and direct the growth of the cell. We call this large network of chemical interactions the "proteome." It can be viewed mathematically as a large-dimensional time-evolving dynamical system of *nonlinear differential equations*; its variables are the concentrations of each particular kind of chemical[5] The system also has inputs corresponding to chemical concentations and other conditions external to the cell.

Now (6) in the development of a multicellular organism like a human, the initial pluripotent cells develop into different kinds of cells such as muscle cells, fat cells, neurons, transparent eyeball cells, etc. Why do these cells (all of which have identical DNA) know what to be? The answer is that their proteomes move to different attractor-regions in the high-dimensional space in which the dynamical system lives. It appears that different cells to some degree "know" their physical locations and orientations inside the organism, and that to some degree cells can communicate with their neighbors and also in some cases to faraway cells. All this is not well understood (but rapidly becoming better understood) by current science, but anyhow all such cellular "self-knowledge" must be caused by

[2] The *M.genitalium* genome really has smaller information content than the naive value of 1160148 bits because its DNA sequences are amenable to "data compression." One reason is that it has 32% GC content. Another reason is that arguably only the codon *triplets* within *gene-coding regions* matter. Actually, promoter and repressor gene-prefix regions also matter, and RNA-coding-genes matter as well as just protein genes – but if we simplistically ignore those facts: If only the *amino*-sequences of the proteins coded for by the apparent genes in *M.genitalium* (in a standard 20-character alphabet, separated by extra 21st characters) are considered, then its genome is 177708 characters long and the data compression tool `gzip` encodes this in 832520 bits (1.39× smaller than the naive genome bit count). This is still 3330 times larger than 250 bits so the paradox remains immense. The TIGR scientists who sequenced *M.genitalium* tried systematically destroying single genes in an effort to determine a small gene-subset essential for survival. They concluded that only 56-78% of them were essential for living. The lower scale factor would reduce the bit count to only about 466000 bits, i.e. reducing the 3330-gap-factor to "only" 1864. Perhaps an appropriate kind of *lossy* data compressor could reduce this without its errors affecting viability of the organism. For example a well known approximate conception of proteins is that the 20 amino acids can each be modeled as only two types – hydro-phobic and philic – with the exception of a few aminos per protein which have to be got right because they form an "active site," as opposed to mere "scaffolding." Even assuming optimistically that this can be done would only get the description length down to around 170000 bits – not counting the length of the decompression instructions – which is still an immense paradox.

[3] That is, there are 10^{50} atoms in the planet Earth, and assuming new configurations were being tried every nanosecond (which is very generous since chemical reactions happen much more slowly on average) for 10^9 years, that is $< 3 \times 10^{75} \approx 2^{250}$ configurations tried in all. Hence if the simplest life required more than 250 bits to describe, it would have been unlikely to occur.

[4] However, mammals also could not survive in the absence of other forms of life, and hence too are "parasitic" and not "true" life, by that reasoning.

[5] We regard methylation of DNA, and the different rates of synthesis of different proteins as regulated by whatever is sticking onto the DNA at that time, both as just part of the proteome dynamical system. And this system is really a system of "*stochastic* differential equations" because there are considerable noise and randomized discreteness effects. This differential equation system model is a standard idea. For some representative component differential equations, see any book on enzyme kinetics. Recently some extremely ambitious projects [1][31] have started with the goal of producing just such a computerized differential-equation model of the entire *E.Coli* organism.

different proteomic attractor states and all such "communication" between cells must be caused by chemical secretions or receptions – which can be viewed as additional inputs and outputs of the dynamical system modeling one cell's proteome.

Then (7) the human brain and neural system somehow self-develops via such processes, and then additional *electrical* communication between neurons can and does occur. This electrical communication is believed (8) somehow to feed back to affect the growth of "synapses" between neurons, allowing the brain to remember information, as well as to self-alter and self-develop in unknown ways. The ultimate result of this plus sensory inputs is the development (9) of an "intelligent conscious being." Now (10) this human can contribute to human knowledge stored in libraries, or in other ways affect his external environment.

Observe that this pipeline features almost entirely *one-way* information flow[6] except in its central step 5: the proteome interacts with almost everything.

Now let us consider the number of bits of *information* in each bucket in the chain; see table 3.1. The important thing to note is that *although the initial and final buckets in the chain are very large, there is a far smaller "bottleneck bucket" in the middle describing the proteomic state.*

If you want to understand the true information content of an organism, then I recommend, to get an upper bound as tight as possible, to measure the information content *at that bottleneck*.

"information pipeline stage," a.k.a. "bucket name"	#bits info
Human DNA	5.8×10^9
Human "non-junk" (regulatory & coding) DNA	10^8
Proteomic state (A)	3×10^6 (range: times 3, or divided by 3)
Proteomic state (B)	4×10^7 (range: times 3, or divided by 3)
Proteomic state (Baum estim. [4] p.57 as corrected)	2×10^6
Number of synapses in human brain	2×10^{14}
Number of cells in human body	10^{13}-10^{14}
Visual sensory information input & processed during human lifetime	4×10^{15}
Amount of information output by human by writing 100 books	10^9
Amount of information output by vocal human by talking during life	10^{13}
Information storable in a human's memory	10^{10}-10^{13}
Information stored in books in world libraries	10^{13}

Figure 3.1. Estimates of (uncompressed format) information at each stage in the pipeline; the minimal "bottleneck bucket" is in **bold**.

Notes: Reasons to believe human DNA is 98% "junk" include the apparent absence of any coding or regulatory function, the fact much of the junk consists of short repetitive sequences, the fact that closely related species can have vastly different amounts of junk (and hence total) DNA, and the fact that Arabidopsis thaliana has 23 times less DNA than humans despite being apparently comparable in genuine information content to plants with much more DNA than humans (plants generally speaking have more DNA and more genes than animals), thus "proving" that at least 95% of our DNA is not really needed. (Similarly, the zebrafish has vastly more junk DNA than the pufferfish.)

Data compression software will reduce books by about a factor of 2. Certain prolific authors have written over 100 books. Each synapse stores at least one bit of information by its very existence or nonexistence. The estimate of 2×10^{14} synapses comes from [27]. The fact that "20 questions" is a reasonable game suggests that humans remember and label $2^{20} \approx 10^6$ concepts. But a far stronger lower bound estimate of $\approx 10^{10}$ bits is got by more sophisticated measurements ([20] with corrections), and the feats of memory prodigy "Elizabeth" make even 10^{12} or 10^{13} bits seem plausible (we discuss these poorly understood subjects in §4). "Visual information" assumes 3 Mbit/sec input rate [43] for 16 hours/day for 70 years. (Total lifetime sensory input is at most about twice visual.) "Vocal information" is for somebody who speaks for 5 hours per day for 70 years assuming MPEG audio compression to 20000 bits/second. (But speech can still be intelligible if compressed 10× smaller than that.) For estimates about the information content of the proteomic/metabolic state, see text; A is assuming introns and exons can be regarded as an earlier stage in the information pipeline and should not be reckoned as part of the "proteome"; B is assuming the contrary, resulting in a much "larger" proteome; C is an independent estimate by Baum which we shall also discuss. ▲

The important thing about the proteome dynamical system is that it, and essentially it alone, suffices to understand the cell and whatever multicellular development and all the rest of it happens. In other words, if we knew enough information to *write down* the system of differential equations known as the human proteome & metabolism, and knew the initial state-vector for the initial zygote cell, then we would in principle be able, by mechanical computation, to predict the entire development of the human body, brain, and intelligence, etc.

So: how much "information" *is* that? We now estimate that. Essentially there are not very many kinds of biologically-important molecules in humans. In a *bacterium*, 23000 genes would mean only 23000 protein and 23000 RNA types that matter, and furthermore most of the RNAs would *not* really matter in the sense that (it is generally believed) they do not

[6] Actually, many exceptions to unidirectionality are known, such as RNA "reverse transcriptase," RNA "editing," DNA manipulations occuring inside our immune systems, and humans deciding who to mate with based on sensory and mental information. However, we shall follow the usual practice of regarding such exceptions as comparatively unimportant and negligible.

interact with anything besides ribosomes in any interesting way, and hence should not be counted (or at least, the count should be negligibly small) as part of the information bottleneck because they really are an entirely separate (earlier) bucket in the information bucket brigade.

However, *eukaryote* genes have *introns* and *exons*, which increases the number of possibly-important RNA types (in the case of the human genome [33]) by a factor of 16.6. It has been speculated that the intron, exon, and associated RNA-splicing machinery may play an important regulatory role. But on the other hand, they might have essentially zero regulatory role and hence almost no impact on the bit-count in the information bottleneck, and instead are there as a defense against viruses or to increase "evolvability" and facilitate diversification[7] At present, the evidence to resolve this question is murky, unclear, and lacking. (Similar remarks can be made about the purpose of the nuclear membrane in eukaryotic cells.)

So we are going to proceed simultaneously considering both extreme hypotheses:

Bacterial model: Introns, Exons, and messenger RNA generally, should not be counted when reckoning the description bit length of the proteome dynamical system.

Exon-dominance model: Count them all.

Finally, we suppose there are also 10000 other important chemicals. That means the proteome is a ($D = 414800$)-dimensional dynamical system – but only ($D = 33000$)-dimensional under the bacterial model.

Now suppose a typical chemical's concentration time-derivative is directly affected by 2 other "random" chemicals (we do not count ubiquitous chemicals such as water, ions, and ATP as "random" – those interactions have a much smaller information content thanks to the high data compression properties Huffman coding can achieve for high-frequency letters; also undoubtably some chemicals are directly affected by many more than 2 others but only the *average* matters)[8] That means that describing the directed-network structure of the interactions takes about $D \lg((D-1)D/2)$ bits. (Available wallcharts of metabolism & etc networks roughly support these estimates.)

We also need to associate some numerical coefficients with each arc in the network, and we assume that typically, it suffices to get each coefficient right to within a factor of 11% accuracy (because we assume cells are robust enough to survive 24% changes) on a factor-1000-wide logscale [i.e. each coefficient can be stored in 5 bits, plus a sign bit], and that for each network arc there are two coefficients. Plus add 4 more coefficients per node to describe interactions with ubiquitous chemicals, and another two coefficients for self-regulation (8 in all), for a total of $48 = 8 \times 6$ bits per chemical (node). Finally, to describe the initial log(concentrations) of all chemicals sufficiently accurately, again suppose 5 bits per chemical suffice. We conclude that

Bacterial model: With 23000 genes, the description length of the human proteome is 0.66×10^6 bits to describe the network structure, plus 1.58×10^6 bits to describe the coefficients, plus 0.16×10^6 bits to describe the initial state, for a total of 2.40×10^6 bits.

Exon-dominance model: With 23000 genes, and 16.6 introns and exons (combined) per gene, the description length of the human proteome is 10.4×10^6 bits to describe the network structure, plus 19.9×10^6 bits to describe the coefficients, plus 2.1×10^6 bits to describe the initial state, for a total of 32.4×10^6 bits.

There is evidence [9] that biochemical dynamics are *extremely* robust even to *extremely* large numerical perturbations. In that case, these bit count estimates might be decreasable by a factor of 3. On the other hand more interaction types (i.e. a more complicated network) might be needed than we estimated, perhaps increasing the bit counts by a factor of 3.

Now if we only consider the *mental* part of the human proteome, i.e. consisting of those genes that concern themselves with the development of human brain and intelligence, and crudely[9] model the non-mental genes and proteome dynamics as "identical" for humans and other mammals commonly regarded as "unintelligent," then (in view of the known "98% similarity" between human and chimp genomes[10]) these estimates shrink by a factor of 50. We conclude:

The "blueprint" for building a human-like brain and intelligence, as opposed to some unintelligent animal, fits in

Bacterial model: *6 Kbytes*

Exon-dominance model: *81 Kbytes*

(central estimates; ranges are /4 or ×4); 1 Kbyte=8192 bits.

We now describe an independent estimate by Baum ([4] box on p.57, section 2.2). Baum's argument (but with altered numbers that we consider to be accuracy-improvements) is this. The human genome contains about 23000 genes[11] Assume approximately that each codes for a protein with average length 300 aminos. Since there are 20 amino acids, defining each protein sequence takes $\lg(20^{300}) = 1297$ bits per protein. Multiplying by 23341 we find that a total of 30 Mbits describes them all. Multiply by 2 as a "fudge factor" to account also for regulatory information. The result is 60 Mbits as the descriptive complexity of a human.

Now Baum insightfully points out that Keefe & Szostak [18] found that 1 in 10^{11} random 80-amino polypeptides had a

[7] It is known different races of humans have vastly different vulnerability to smallpox, measles, and yellow fever viruses, and in the former two cases, it is known that those genetic differences must have developed during the last 5000 years. Had the human species been incapable of evolving that quickly, it would have been a much less-fit species.

[8] Actually, these biochemical networks appear to have "power law valence scaling" [2] i.e. the probability that a node has valence v is proportional to a fixed power of v. The approximation we have been describing is a crude attempt to estimate the entropy of this.

[9] This should still yield a valid *upper bound* because not all chimp-human differences are mental.

[10] From an "NIH News" press release 31 Aug. 2005 based on the 1 Sept *Nature* paper [6] on the chimp genome: "The consortium found that the chimp and human genomes are very similar and encode very similar proteins. The DNA sequence that can be directly compared between the two genomes is almost 99% identical. When DNA insertions and deletions are taken into account, humans and chimps still share 96% of their sequence. At the protein level, 29% of genes code for the same amino sequences in chimps and humans. In fact, the typical human protein has accumulated just one unique change since chimps and humans diverged from a common ancestor about 6 million years ago."

[11] As of March 2006, the Ensembl genome-annotation system estimates there are 23341 human genes plus 719 "pseudogenes."

particular chemical functionality (namely, binding to ATP).[12] Assume that this is true for *any* given chemical functionality. (This is a rather dramatic extrapolation from one data point! – But that is somewhat ameliorated by the fact that we shall take the logarithm.) Hence *really* a protein *described by its function* perhaps is describable in only $\lg(10^{11}) = 37$ bits. In that case,[13] again putting in a "fudge factor" of 2 to account for regulation and fine-tuning, the descriptive complexity of a human really would be only 1.7 Mbits.

This corrected-Baum number coincides quite closely with our bacterial-model estimate of 2.4 Mbits. Baum goes on to observe that this is far smaller than the number of Darwinian evolutionary experiments that went into finding these 1.7 Mbits – in [37] we provide an estimate of "between 2.8×10^{42} and 3.6×10^{44} not counting viruses, and these numbers would be increased by a factor between 16 and 251 if we also count viruses" for the number of lifeforms throughout Earth history.

Finally, it also is possible to get an intelligence-blueprint upper bound by yet another method. The chimp-genome paper [6] says that the human-chimp difference consists of "35 million single-nucleotide changes, 5 million insertion/deletion events, and various chromosomal rearrangements." Multiply by 2% to only consider the changes in the non-junk and figure each change can be described with (on average) 10 bits, to get 8 Mbits as an upper bound on the description complexity of the human-chimp difference. If only exonic changes are considered interesting, this can be halved. Also, it is legitimate to instead reckon the difference between the human and X genes, where X is the most recent human-chimp common ancestor, as half the human-chimp difference. That also leads to a halving. The result is either 4 or 2 Mbits as an upper bound on the description complexity of a human-intelligence blueprint. We can further reduce these numbers by considering the fact that *genomic* information is larger than the proteomic *bottleneck*; the results of applying that final reduction (in our two models) are quite close to ours.

4 Human memory and memory prodigies

We now review the history of quantitative human memory research, which unfortunately has been a comedy of errors.

Landauer's estimates about typical human memory. For our purposes, Landauer [20] is as far above most psychologists studying memory, as a dog is above an ameoba, because he actually has heard of the concept of "a bit" and uses at least approximately-valid information-theoretic techniques [13][7] in his experimental design and analysis, whereas the others would seem (at least from their writings) entirely blissfully ignorant of all of that. I strongly recommend they adopt Landauer's techniques – or preferably better ones, because I chose the word "dog" advisedly; even Landauer still fails to use fully correct information theory (consequently making major errors), thus bringing the net number of memory-psychologists ever to do so, to a grand total of zero.

And unfortunately, Landauer only tested normal people and made no effort to find super-talented memorists. The latter may be more interesting as giving a better estimate of true human "hardware capabilities," under the view that most of us do not reach our ultimate limits.

As a simple example of his techniques: Landauer notes that in prose passages with random words omitted, people can guess the omitted words about 50% of the time. This, he argues, proves the information content of prose per word averages about 1 bit. The most important remark about that is that this is far less than might be naively supposed based on the fact that number of words in dictionaries far exceeds 2. Now actually, Landauer's 1 bit/word deduction is false (albeit far closer to being true than anything the other psychological authors say) because, if, say, half the words were deducible and the other half were 64-way uncertain, it would really be 3 bits per word. A second reason is that to make this argument valid, really the deduction of the word needs to be based on the *preceding* prose only. There are in fact techniques based upon "gambling" (where people make real-money wagers about the next word) to estimate the true entropy of English prose, but Landauer did not use these more sophisticated techniques. Also guessing the next *letter* is superior to guessing *words*, and also non-gambling techniques superior to Landauer's already had been given by Shannon [35] at the dawn of information theory.[14] The Cover-King gambling estimate [8] finds that the entropy of English text is 1.3 bits per symbol, closely coinciding with Shannon's estimates, both of which (note) are about *7 times larger* than Landauer's (crude and wrong) estimate! And furthermore, Cover, King and Shannon apparently used a comparatively low-entropy text (the book "Jefferson the Virginian") in their experiments [24], so really the correction-factor for Landauer probably should not be 7 but in fact *10*.

As a more sophisticated example, Landauer found that people guessing random missing words in a prose passage they had never seen before, could do so 48% of the time, but if the prose passage had been *read before* in its entirety (i.e. with all words) then their success rate was 63%. Landauer therefore estimated that the amount of extra information about the prose that the humans had stored in long-term memory during their previous reading, was $\log_2(63/48) \approx 0.4$ bits per word. Using their measured reading speed, Landauer then found their memorization bit rate was 1.2 bits per second. Again, this estimate is not really valid and would have been improved via gambling-based techniques to measure true entropy: probably a better estimate than Landauer's (which I got by crude scaling of the Cover-King estimate) is about $(63-48) \times 10/100 \approx 1.5$ bit per word, i.e. 4.5 bits per second; and considering the harder-to-guess words are probably predominantly longer ones, 6 bits per second is probably more

[12] This is Baum's analogue of our "proteome differential equation system" idea. In our approach, we argue all that matters is the "differential equations." In Baum's approach all that matters is the "chemical functionality." Both are in essence the same idea.

[13] Precisely: $2 \times 23341 \times \lg(10^{11}) \approx 1705817$. Baum made an arithmetic mistake here and got a result over 4 times larger.

[14] Shannon's procedure was for a human to guess the next letter and keep guessing until they succeeded. He gave two formulas for producing lower and upper bounds on the entropy of text as functions of the number of guesses required. Shannon thought his upper bound formula was probably far closer to the truth because humans are good guessers. Better analytical methods than Shannon's, resulting in tighter entropy bounds at the cost of more computation, later were devised by Levitan & Reingold [22].

appropriate, i.e. over an order of magnitude higher than Landauer's flawed estimate. Nevertheless, even Landauer's highly flawed estimate is far superior to the other psychologists who had no information-theoretic estimates whatever.

Also note that the humans could have been memorizing bits encoded very inefficiently, e.g. actually *storing* 10 bits per every actual bit of true information useful for this word-guessing task. Landauer's experiments give no insight into this encoding-efficiency question. That is important since for our purposes in this paper, what seems to matter more than "true information content" is in fact the (larger) number of "raw" bits stored in the underlying hardware. Finally, note that Landauer's measurements concerned the information absorbed during normal albeit concentrated, reading. At least for me, reading is an effortless act and garners some but not all available information (which I might describe as a compressed version of the "meaning") – whereas intentionally trying to memorize a prose passage verbatim is an unnatural and effortful act. Because Landauer's measurements were focused more on normal human reading and picture viewing as opposed to abnormal rote-memorizing acts, they probably give a substantially better picture of true average human information storage.

Now in some related (and probably also flawed) ways [20] Landauer was able to estimate people's true memorization bit rates from pictures (for picture-recognition tasks), finding 2.3 bits per second. Interestingly (according to Landauer), it appears that several different kinds of input types all yield roughly the *same* number, namely 1-3 bits/second, for the memorization uptake rates of average humans.

Then Landauer under the assumption a typical human learns 2 bits/second all the time for 70 years of 16-hour days, deduces human memory is 2×10^9 bits lifetime, albeit due to Landauer's flawed techniques this number really should be both less certain than he said and probably larger, e.g. 10^{10} bits.

However, this all is under the (of course wrong) assumption that no bits are lost during those 70 years. Landauer also was able to use his techniques to estimate true-information loss rates in humans, finding an estimate (and he admits discomfort with its crudity and large suspected error) of 10^{-9} bits/bit/second. If that number is correct, then Landauer finds that his human-lifetime memory estimate of 2×10^9 bits needs to be decreased by a factor of 2, and about the same factor should pertain to our corrected larger Landauer estimate.

So Landauer's final estimate was that humans in their lifetime typically store about 10^9 bits of true information, which we correct to about 10^{10}, which suggests (in view of the number of synapses in the human brain, which is 4-5 orders of magnitude larger) a very inefficient coding scheme. This corresponds to a net rate of about 1 truly-useful bit per second memorized during waking hours, where this net figure includes both gain from memorizing and loss from forgetting; but this should be corrected to more like 5 bits per second. This same corrected figure is what also would result from Reed & Durlach's [30] table of human information input and output bit rates from various modalities (braille, morse code, visual reading, audio communication, court stenography machine, etc; the fastest input bit rates found were for visual reading and listening to speech at 30-60 and 25-60 bits/sec respectively) by multiplying them by $(63 - 48 = 15)\%$.

It would be a good idea to redo all of Landauer's experiments this time correctly; our preliminary estimate is human memory is 10 times larger than his estimate![15]

Rajan Srinivasan Mahadevan once memorized 31811 digits of π to get listed in the Guinness Book of World Records; he recited the first 10,000 digits to his inviligators at a mean rate of 4.9 digits per second. This is a recall bit rate of 23 bits per second, although Rajan gradually slowed down during the next 21811 digits, in his final stages achieving only 2.8 digits/sec. Rajan could memorize matrices of random digits at a rate of 1 digit per 1.5 to 5.8 seconds depending on the matrix size from 5×5 to 20×20 with larger matrices taking longer per digit. (Error rate$<$ 1%.) Note that a "digit" is $\log_2(10) \approx 3.32$ "bits," so these are memorization bit rates of 0.6 to 2.2 bits per second. In contrast, normal humans had 4-5 times slower memorization rates to get error rates ranging from slightly to 10 times higher [41].

Shereshevsky was a Russian professional mnemonist studied as "S" by A.R.Luria [23]. S could remember sequences (of apparently unboundedly long length) of letters, numbers, words, or nonsense words, provided the items were read out to him slowly and distinctly at a rate of about one per 3-4 seconds (or presented in writing) without distracting influences such as noise or other people talking. S would then (inside his mind) go over the sequence once, after which he could recall it either forwards or backwards, usually with no mistakes, even many years later. Assuming a million candidate nonsense-words, this is a memorization bit rate of 7-10 bits per second, not counting the time for the mental revisit/check.

Note that Rajan fell below, while S exceeded Landauer's (corrected) bit rate estimates for typical humans, albeit their feats pertain to a more difficult memorization task than Landauer's since Landauer had made absolutely no demand for high accuracy and verbatim recall.

"Elizabeth," a 23-year old teacher and artist [39][40], could gaze at an image of a million random dots with her right eye for three 3-minute periods. Four *hours* later she could look at another such image with the other eye. The randomness was correlated between the two images to form a "random dot sterogram"; anybody looking simultaneously at both images, one to each eye via special optics, will see a simple 3D figure, although either picture alone is just a random dot cloud. Elizabeth could merge the memorized image with the new one to see the 3D figure, which appeared by parts in her mind over a period of 10 seconds. She then would describe the 3D figure. Other feats by Elizabeth:

1. With a 10^4-dot random stereogram, 1 minute of gazing at the first picture sufficed, and then she could see the 3D figure immediately when the second picture was presented and viewed with her left eye. She could do this anywhere from 10 seconds to 72 hours (or presumably more) later. Some of these feats were done un-

[15] Landauer was also criticized heavily by L.Hunter, whose paper [16] opens "Landauer's estimate is deeply flawed." However this intense piece of criticism was itself of very poor quality, as is indicated by the fact that it totally neglected to mention all *our* Landauer criticisms.

der double-blinded conditions where not even the experimenter knew the 3D image during the experiment. (Normal people cannot do this even if the delay between images is made as short as 150mSec, although with the two images flashed alternately to each eye at > 20Hz, almost everybody can see the 3D figure.)
2. She could selectively recall any desired one of four such memorized 10^4-dot pictures.
3. She could match a memorized R-picture with more than one L-picture to see entirely different 3D images.
4. She could remember images with either eye eye-specifically.
5. She could recall a poem in a foreign language as an image, years later, and write it down rapidly.
6. She claimed once to have recalled a 10-second interval of a Laurel and Hardy *movie*, eidetically one week after viewing.

The "Shass Pollak" were a group of (mostly Polish) Jews who carried on a religious tradition of memorizing the Babylonian Talmud [38]. This tradition was speculated to have started in the days when books were very expensive and it was deemed a religious duty to preserve the Talmud inside human memory. To appreciate their feats one must realize that the Babylonian Talmud consists of 12 large volumes in total 5422 pages long. All printed editions have the same number of pages and in all editions each page has the same number of words and the same initial and final word. Thus one could ask one of the Shass Pollak "what is the 4th word on the 8th line of the Tractate Berakhot volume?" and then confirm his answer by consulting the book. The Shass Pollak evidently had in their minds, essentially a photograph of each Talmud page.

A 34-year old woman named "AJ" claimed that since age 11 she'd had an extraordinary memory for daily events; she could remember instantly and vividly details of her past, without having ever devoted any particular effort (such as mnemonics or rehearsal) to doing so. She first became aware of her extraordinary abilities at age 12 and by age 14 they had improved and become automatic. If any date in her high-remembrance period (age 14 onward) were stated to her, she could immediately tell you what day of the week it was, what news events or events from favorite television shows happened on that day, what the weather was, what she was doing that day ("shopping with my mother"), etc. [29][19].

Bit counts for Elizabeth. Assuming Elizabeth's eidetic images were stored permanently (or at least in non-overlapping fashion) and that she could in principle accomplish her feat every 15 minutes per 13-hour day for 50 years, and assuming the image storage requires 10 bits per dot, we find that Elizabeth has a 10^{13}-bit memory. This is interesting because it is actually nearing the number of synapses in a human brain, which has been estimated to contain 2×10^{14} synapses by Pakkenberg et al [27] and for nearby estimates see [5].

Caveats. However, this all is not completely clear because
(a) Elizabeth was tested on random dot stereograms of *simple* 3D figures such as T-shapes and raised squares; hence she might, in principle, have only memorized the locations of, say, 1% of the 10^6 dots and would still be able to describe the 3D figure. Had she been tested on a *non*stereographic random-dot image pair whose superposition was a relatively complicated and unpredictable figure, and had she been asked to draw that figure with a pen on the later random dot sheet, then if the result was accurate we could be fairly certain she actually memorized all the dots – but this experiment unfortunately was never performed.

However, if Elizabeth really only memorized 1% of the dot-locations, then we must ask how she picked out the particular 1% subset, and how she memorized those particular locations so accurately without remembering the others as "landmarks." Also, it seems plausible that Elizabeth must have been able to store these images in a highly uncompressed "raw" form in order to be able to successfully export the data to her mental stereoopsis mechanism. That mechanism evidently expects image data in a raw form (as opposed to some form highly compressed by "feature recognition" to a small list of "image features") which seems proven by the very fact that random dot stereograms – or successful stereoscopic depth detection when viewing otherwise featureless surfaces with fine surface roughness – work at all. This reasoning all makes it plausible that Elizabeth really did memorize the locations of all 10^6 dots in about 10 minutes.

(b) More than one type of random dot stereograms were used including a type with a square grid of black-or-white pixels 50% of which were black. To store this type, only *one* bit per pixel suffices, whereas for the type with randomly located sparse circular dots, many more bits per dot are required – e.g. about 10 bits per dot for dots of sparsity 1/1024 and integer coordinates; or for dots of sparsity 1/64 with location accuracy of 1/4 of a dot-width. And unfortunately Strohmeyer [39][40] *did not say which* type of stereogram was used in the most impressive (10^6-dot) experiments!

Conclusions about Elizabeth. It seems clear that Elizabeth when memorizing eidetic images memorized *at least* 20 bits/second and quite possibly as many as 16,000 bits/second (the vast gap between these two estimates is a sad artifact of the incredibly poor experimental design and reported information in [40] and the complete lack of attention they paid to bit count estimates). That is much more impressive than the other memorists.

Bit counts for the Shass Pollak. To memorize 5422 pages with 2800 characters per page requires (at 5 bits per character) 10^8 bits of memory. Assuming the memorization was of bits *compressed* to the Shannon-Cover-King entropy estimate of 1.3 bits per symbol, this would be only 2×10^7 bits. *We can take that as a confident lower bound on human memory.* On the other hand if Shass Pollak memorization was *photographic* – as *uncompressed images* not words or characters – then the bit count is much larger, $\approx 10^{10}$. Unfortunately nobody mentioned in [38] ever *asked* the Shass Pollak how they remembered the data (as characters, words, or photographic image) nor ever tested them to try to find out which it was (e.g. could they remember printing flaws?).

My guess would be that it was as an image, for the following reasons which fall far short of certainty. If I were simply remembering text, then I would not know which was the 10th word on the 20th line. E.g. Rajan when memorizing π (and also A.C.Aitken) used "auditory" methods sometimes combined in Rajan's case with links to perverse cues (such as

"111" was "Nelson" because "Admiral Nelson had one leg, one arm, and one eye") while other π memorizers used mnemonics and number-to-syllable verbal codes – hence none of them would have been able to tell you the 15793rd digit of π without performing a long count. The printed editions of the Talmud do not have numbered lines. Conceivably, though, the Shass Pollak could have regarded the Talmud as a text string *with artificially added line numbers* in which case mental "photography" would not have been necessary (and similarly Rajan probably could have modified his π memorization methods to permit him to quickly retrieve requested digits, if he had had that goal in mind). But even then, any memorizer regarding the Talmud as text would probably as a side effect be able to interpret it and would have thought a good deal about it, whereas, some or all Shass Pollak apparently felt unable to interpret it.

Eidetic memory in infants? Strohmeyer speculated that perhaps eidetic memory was more common in children than in adults. Indeed, one could even make the (wild) speculation that perhaps *all* or many children in some age range that is a subinterval of 0-3 years old, have eidetic memory (and then almost all lose it). This speculation might not be incompatible with the available data because of "infantile amnesia," i.e. the fact that all or most people do not remember anything that happened to them before age 3. If true, this wild speculation would of course be tremendously important for the light it would shine on children's mental development. It could be experimentally tested even for pre-verbal infants by use of (nowadays common) eye-tracking devices. Specifically: show an infant some random dot images, some of them magically correlated to yield interesting combined images, and some not. (There is no need for stereo for this purpose.) If the eyetracker shows the infants look at the magic images longer, that (and some trivial statistical-significance analysis) would prove the speculation. Unfortunately, this experiment has never been done.

Bit counts for AJ. Unfortunately none of the extensive experiments on AJ attempted to estimate a bit size of her memory.[16]

It would have been easy enough to do. Let me explain one approach. Show AJ a movie she has not seen before but thinks would be interesting (since AJ's memory works best on things she's interested in). After some delay, ask her prepared questions about the movie, e.g. who entered the scene, who left, what was sitting on the shelf, what color was somebody's shirt etc, and/or have her just describe random subintervals of the movie in detail. The point is, we can estimate how many bits it would take to be able to answer all possible questions of the same ilks as the ones asked, and we know what percentages AJ is able to answer – and with secret preparation of questions intended to act like a random sample from questions of those same ilks, there is no way for AJ to prepare for them. Then by arithmetic we find a lower bound on how many bits AJ remembers about the movie. Another approach (following Landauer [20]) would be to have other control humans view the movie up to a certain point immediately before some event, then be required to guess what that event would be – and even better (following Cover & King [8]) would be to use the "gambling" methodology of converting those guesses into entropy estimates. The differences between the success rates before and after seeing the rest of the movie would be useable to estimate the information gain got from any previous viewing by either AJ or anyone else. Now *scale* the duration of the movie up to AJ's entire 26-year period of high-performance memory – and what is the resulting bit count? (The same experiment could also be done for normal humans of course.) It is not even necessary to show AJ a movie; this also could be done simply on some movie she is known to have seen previously and liked, and by doing this for several movies she saw at various dates further and further into the past (easy since AJ remembers all dates) we can determine her "forgetting curves."

In the absence of such purpose-designed experiments, we must estimate AJ's bit count from what was published in [29] undoubtably yielding a gross underestimate. Assuming AJ remembers 5 events per day, each event stored in 200 bits, for 28 years, that is 10^7 bits.

5 Comparison with some computer programs

Recall from §3 our central estimates (Bacterial & Exon-dominance models: 6 & 81 Kbytes) for the size of a "blueprint" for building a human-like brain and intelligence, as opposed to some unintelligent animal. (Ranges of uncertainty in these numbers: /4 or ×4.)

That is amazingly small! First, it is smaller than the bit counts for even the smallest known bacterial genomes from table 2.1. Second, many a computer programmer, including me, has written a program much longer than 7 Kbytes without creating anything near a human-like artificial intelligence. And those programs that *have* been intended to try to create anything reasonably near an artificial human-like intelligence, usually have been much larger than 7 Kbytes.

Examples are listed in table 6.1.

(Possible) Lessons. It is interesting that both "Hayek" [4] and PAPPI's [11] principles file (both in bold print in the table) – although *not* the PAPPI natural language system as a whole, which is far larger – lie between our estimated upper bounds for the human intelligence blueprint. This *perhaps* indicates that the underlying conceptions of their progenitors – Eric B. Baum & Igor Durdanovic and Noam Chomsky & Sandiway Fong respectively – are not too far off the mark as theories of how human intelligence works. The fact that they exceed our smaller (7 Kbyte) bound perhaps indicates that the "bacterial model" is wrong and the "exon dominance model" correct, or perhaps merely indicates the inability or unwillingness of these authors plus the gzip compressor to code these things concisely. It also perhaps indicates that most of the infrastructure for both "intelligence" and "innate linguistic principles" was already built into the brains of many very-*pre*human animals – since that would might give us 5-15 or more times more "elbow room" than if we assume this all

[16] Also not published were "forgetting curves." I am unable to explain the incredible allergy to bit counts among psychologists researching memory. Reading that literature it often seems as though those authors are engaged in an intentional conspiracy to try hard to avoid measuring the single most important number in their field.

arose between chimps and humans: we are basing everything on the widely quoted "98% similarity" between the chimp and human genomes, as opposed to the "70-90% similarity" between the mouse and human genomes.

The "principles and parameters" linguist community that Chomsky & Fong associate themselves with seems remarkably uninterested in the competing "link grammar" framework invented by computer scientists [14]. Their link parser system also is electronically available to everybody, works fairly well, seems (at least to me) more easily understandable, assimilable, and incorporatable into other programs, and it is of comparable size: about 800 Kbytes uncompressed source code, plus about 1 Mbyte of data (dictionaries, word properties, etc) including about 20 Kbytes of "knowledge." (Because of this size-similarity, only Fong's system is listed in table 6.1.)

6 The dominance of culture

Some of the 10 stages in our "information pipeline" are different in the sense that some of their information arises not only from DNA but also from the external environment – in particular, from reading books and listening to other humans. Indeed, Baum [4] and van Schaik [34] have argued that a great deal of human intelligence is transmitted *not* genetically but in fact through human culture. Humans have a great advantage over other animals because they have spoken and written language, which enables the easy transmission and storage of "computer programs." Other animals without language still can (and do) learn by imitation,[17] but that only enables transmission of a restricted class of computer programs; those whose actions are largely *externally visible*. But for future *AIs* (computerized artificial intelligences) transmission and storage of computer programs is expected to be a triviality so they in turn will have a large advantage over humans in this respect.

Is Baum's "culture dominates" claim justified? Yes. We can tell that because using any of our three upper bounds on the bottleneck, and considering from §4 the fact that humans learn 5 bits/sec from reading or listening, we can see that only a few days to months of reading and listening will suffice to get more cultural information than your entire amount of genetic information transmitted through the bottleneck.

If cultural information dominates our intelligence, then what is the meaning of our assertion that a "blueprint" for human intelligence would fit in 81 Kbytes? It is a matter of definition. Humans have genetically what it takes to build and maintain a culture to amplify their intelligence and chimps and other animals do not — and our blueprint-size bound is a valid upper estimate of the information-content responsible for that *difference*.

Baum ([4] chapters 12-13) gives an interesting discussion to show that "intelligence" varies smoothly from *E.Coli* to honeybees to dogs to humans. In the event that the reader considers some creature X other than chimps on this spectrum "unintelligent" and humans "intelligent," then he can produce an appropriate altered version of our numbers by scaling the human-chimp genome similarity of 98% to whatever the human-X genomic similarity is.

References

[1] Kazuharu Arakawa, Yohei Yamada, Kosaku Shinoda, Yoichi Nakayama, Masaru Tomita: GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes, BMC Bioinformatics 7 (2006) 168.

[2] A-L. Barabasi: Linked: The New Science of Networks, Perseus Publishing, Cambridge 2002.

[3] C.B.Barber, D.P.Dobkin, H.T.Huhdanpaa: Quickhull algorithm for convex hulls, ACM Trans. Math'l Software 22,4 (1996) 469-483.

[4] Eric B. Baum: What is thought?, MIT Press 2004.

[5] V.Braitenberg & A.Schüz: Anatomy of the cortex, Springer 1991.

[6] Chimpanzee Sequencing and Analysis Consortium: Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437 (1 September 2005) 69-87.

[7] T.M. Cover & J.A. Thomas: Elements of Information Theory, Wiley NY 1991.

[8] Thomas M. Cover & Roger C. King: A Convergent Gambling Estimate of the Entropy of English, IEEE Transactions on Information Theory 24,4 (July 1978) 413-421.

[9] G.von Dassow, E. Meir, E. Munro, G. M. Odell: The segment polarity network is a robust developmental module, Nature 406 (2000) 188-192.

[10] M. Eigen & F. Oehlenschlager: 30 years later - a new approach to Sol Spiegelman's and Leslie Orgel's in vitro evolutionary studies: dedicated to Leslie Orgel on the occasion of his 70th birthday, Orig. Life Evol. Biosph. 5-6 (1997) 437-457.

[11] Sandiway Fong: PAPPI natural-multilanguage parsing system available for download (along with many related papers) from his web page `http://dingo.sbs.arizona.edu/~sandiway/`.

[12] C.M.Fraser et al.: The minimal gene complement of Mycoplasma genitalium, Science 270 (1995) 397-403.

[13] R.G. Gallager: Information Theory and Reliable Communications, Wiley 1968.

[14] Dennis Grinberg, John Lafferty, Daniel Sleator, Davy Temperley: Link grammar (software system and technical reports) `http://www.link.cs.cmu.edu/link/`.

[15] M.D.Hauser: Wild Minds, Henry Holt NY 2000.

[16] L.Hunter: Estimating human cognitive abilities, a response to Landauer, Cognitive Sci. 12,2 (1988) 287-291.

[17] Marcus Hutter: Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability, Springer, 300 pages, Berlin 2004. ISBN=3-540-22139-5.

[18] A.D. Keefe & J.D. Szostak: Functional proteins from a random sequence library, Nature 410 (April 2001) 715-718.

[17] This includes fish. Many mammal and bird examples are found in [4][15]. One famous example is blue tits in 1930s England. After many years of milk delivery in foil-lidded milk bottles, one blue tit learned to open the bottles to drink the cream. The practice then spread rapidly in expanding circles about the first observation, forcing the dairies to alter the bottles. Another example [34] is orangutan culture. Orangutans live at extremely high density in Suaq, a swamp forest in northern Sumatra, and presumably due to their high population density have become more sociable there with each other than is usual for orangutans. They employ tools in numerous ways. The fact that that tool use is socially transmitted is shown by the fact that the orangutans in the big swamp on one side of a river use tools, but those in the smaller swamp on the other side, do not (even though the same opportunities and conditions are available there).

[19] Susan Kruglinski: Some people really do remember everything, an exclusive interview with a woman who can't forget, Discover Magazine 4 April 2006.

[20] Thomas K. Landauer: How much do people remember? Some estimates of the quantity of learned information in long-term memory, Cognitive Science 10 (1986) 477-493.

[21] D.B.Lenat: CYC, A large-scale investment in knowledge infrastructure, Communications of the ACM 38,11 (1995) 33-38.

[22] Lev Levitan & Zeev Reingold: Entropy of natural language – theory and practice, Caho Solitons fractals 4,5 (May 1994) 709-743.

[23] A.R. Luria: The mind of a Mnemonist, Basic Books NY 1968.

[24] H.Moradi, J.W.Gryzymala-Busse, J.A.Roberts: Entropy of English texts, experiments with humans..., Information Sciences 104, 1-2 (1998) 31-47.

[25] Atul Narkhede & Dinesh Manocha: Graphics Gems 5 (Ed. Alan Paeth) Academic Press 1995.

[26] Peter Norvig: Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp, Morgan Kauffman.

[27] B.Pakkenberg & 6 others: Aging and the human neocortex, Exp'l. Gerontology 38 (2003) 95-99; B.Pakkenberg & H.J.G.Gundersen: Neocortical neuron number in humans: effect of sex and age, J. Comparative Neurology 384 (1997) 312-320.

[28] B.Palenik & 14 others: The genome of a motile marine Synechococcus, Nature 424, 6952 (28 Aug 2003) 1001-1002.

[29] E.S. Parker, L. Cahill, J.L. McGaugh: A case of unusual autobiographical remembering, Neurocase 12,1 (Feb.2006) 35-49.

[30] C.M.Reed & N.I.Durlach: Note on information transfer rates in human communication, Presence – teleoperators and virtual environments 7,5 (oct. 1998) 509-518.

[31] J.L. Reed, T.D. Vo, C.H. Schilling, P.O. Palsson: An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR), Genome Biol. 4,9 (2003) R54.

[32] Stuart Russell & Peter Norvig: Artificial Intelligence: A Modern Approach (2nd Edition), Prentice Hall 2002.

[33] M.K. Sakharkar, V.T.K. Chow, P.Kangueane: Distributions of exons and introns in the human genome, In Silico Biology 4 (2004) 0032.

[34] Carel van Schaik & Perry van Duijnhoven: Among Orangutans: Red Apes and the Rise of Human Culture, Belknap Press 2004.

[35] Claude E. Shannon: Entropy of printed English, Bell System Tech. J. 30 (1951) 54-65; also Collected Papers, IEEE Press 1993, now reprinted by John Wiley & sons.

[36] W.D.Smith: DNA computers in vitro and vivo, DIMACS Series in Discrete Mathematics and Theoretical Computer Science (Amer. Math'l Soc.) 27 (1996) 121-185 and `http://math.temple.edu/∼wds/homepage/works.html` #29.

[37] Warren D. Smith: Mathematical definition of 'intelligence' (and consequences), `http://math.temple.edu/∼wds/homepage/works.html` #93.

[38] George M. Stratton: The mnemonic feat of the 'Shass Pollak', Psychological Review 24 (1917) 244-247.

[39] Charles F. Strohmeyer III: Eidetikers, Psychology Today 4,6 (Nov. 1970) 46-50.

[40] C.F. Strohmeyer III & J. Psotka: The detailed texture of eidetic images, Nature 225 (1970) 346-349.

[41] C.P.Thompson & 4 others: Rajan: study of a memorist, J.Memory & Language 30 (1991) 702-724.

[42] Alan M. Turing: Some calculations of the Riemann Zeta function, Proc. London Math'l Soc. 3,3 (1953) 99-117.

[43] D. Warland, P. Reinagel, M. Meister: Decoding visual information from a population of retinal ganglion cells, J. Neurophysiology 78 (1997) 2336-2350.

[44] The Mathematica 5.2 Book Online `http://documents.wolfram.com/mathematica/book`.

size	program
10^8 axioms	D.Lenat's planned (but incomplete) CYC system [21] encoding "commonsense knowledge" which he argued was an essential prerequisite for intelligent behavior.
6×10^7 lines of code	Debian GNU/Linux operating system version 2.2, mostly C. But only 390K lines of assembly-language code and "kernel" is only 1.8 million lines. (All counts exclude comments.)
3×10^7 lines of code	Microsoft windows 2000 operating system.
157 Mbytes	The NAG Fortran library (April 2006).
688K lines of code	"GNAT" Ada compiler (2000).
630K lines of code	"Emacs" text-editor (2000).
"Typically several megabytes"	Executable code for `Mathematica` symbolic manipulation program (says §2.14.4 of [44]).
3.2 Mbytes	C source code for "Logistello" world champion othello game playing program by Michael Buro (source code publicized 2002), not including either opening book or eval function tables. "Zebra," an also-strong freeware othello program by Gunnar Andersson & Lars Ivansson, consists of only about 51K lines of C code.
1.8 Mbytes	bzip-compressed Java source code of ALICE (2006 version of chatterbot by R.Wallace, N.Bush, et al; won 2000, 2001, & 2004 Loebner prizes for best Turing test contestant).
210K lines of code	LEDA `C++` algorithms library v4.0 (line count includes comments).
125K lines of code	"Booch components" ADA library 1987 (line count omits comments), but line count shrunk to only **15K** in the `C++` release 2.0 (with increased functionality!).
450 Kbytes	Qhull 2003.1 multidimensional convex-hull-finding program [3], C source code (& documentation included, which is about 50%).
440 Kbytes in 78 files	The LISP source code files included on the diskette accompanying Norvig's AI book [26].
229 Kbytes	Gzip-compressed C source code for gnuchess-5.07 (opening book not included), a comparatively weak free chess program (low master strength).
188 Kbytes (gzipped)	LISP source files included with Russell & Norvig's AI book [32]; 983 Kbytes uncompressed.
96 Kbytes	Freeware "first person shooter" video game with very impressive graphics.
75 Kbytes	"**Hayek**" attempted artificial intelligence via an "artificial economy" [4] (`C++` code by I.Durdanovic); add 14 Kbytes to make it try (unsuccessfully) to learn to solve Rubik's cube.
16 Kbytes	gzipped C source code for **Seidel polygon-triangulation** algorithm [25].
1440 bytes	Turing's program [42] to rigorously confirm/deny the Riemann Hypothesis about the zeta function in given intervals along the critical strip. (Turing read numbers in radix-32 in a strange alphabet with same ease nowadays people read decimal, so no need to do 32→10 radix conversion; also Turing had to write everything, like basic I/O and $\sqrt{x}$ subroutines, unlike now where programmers use libraries.) Size grows to 2080 bytes if also count numerical tables entered by Turing.
1663 bytes (gzipped)	Patrice Bellard's "Tiny C compiler" (winner of 2002 international obfuscated C code contest) which is able to compile itself. Subset of C general enough to write a small C compiler, then extended subset until reached 2048 bytes of C source excluding ';', '{', '}' & space characters.
1683 bytes (gzipped)	Oscar Toledo's tiny chess program (a winner of 2005 IOCC contest).
1000 instructions	LISP interpreter by Richard Stallman.
874 characters of `C++`	Thistlewaite's nested-subgroup algorithm to input and solve Rubik's cube configurations in about 30 moves average (authored by Tomas Sirgedas of Ann Arbor MI to win a programming contest; the shortest successful cube-solving program entered was only 528 characters of Perl).
625 Kbytes	PAPPI "principles and parameters" human-language parsing program by Sandiway Fong (Prolog source code as compressed with gzip tool; omits lexicons, linguistic principles & parameters, and dynamic libraries).
22 Kbytes	The universal linguistic **principles used by PAPPI** (gzip-compressed).
2426 bytes	The parameters for English language (gzip-compressed).

Figure 6.1. Sizes of some computer programs. Most programs today are written totally ignoring the size of the source code; the rare exceptionally small programs have historically been (a) programs written for early machines with tiny memories, and (b) bizarre programs written to win programming contests. We include those here, as well as (c) more conventional and familiar programs and (d) programs intended to be (or to move toward being) "Artificial intelligences." Programs falling between our two models' size estimates for human intelligence, are in **bold** print. ▲