Mathematical definition of "intelligence" (and consequences)

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

June 18, 2006

Abstract — **In §9 we propose an abstract mathematical definition of, and practical way to measure, "intelligence." Before that is much motivating discussion and arguments why it is a good definition, and after it we deduce several important consequences – fundamental theorems about intelligence. The most important (theorem 5 of §12) is our construction of an algorithm that implements an "asymptotically uniformly competitive intelligence" (UACI). Although our definition of intelligence initially seems "multidimensional" – two entities would seem capable of being relatively more or less intelligent independently in each of an infinite number of "dimensions" of intelligence – the UACI is an intelligent entity that is simultaneously as intelligent as any other entity (asymptotically) in every dimension simultaneously. This in a considerable sense makes intelligence "one dimensional" again and presumably explains why "IQ" is a useful quantity.**

Unfortunately the obvious UACI implementations are useless for practical purposes because of enormous constants of inefficiency. There are many obvious and nonobvious ways to try to get more practicality and efficiency, and it is entirely unclear how far and fast that can be pushed (§15 & 22).

In §16-20 we examine the four most important bodies of experimental facts about human intelligence and find that all four are *predicted* by the hypothesis that human intelligence works similarly to our mathematical construction of a UACI: (1) the Spearman positive correlation and g principles (which we shall see are less supported by evidence than is generally claimed, but probably still roughly correct), (2) the findings of Jean Piaget and successors about the time-development of human intelligence, (3) forgetfulness, and (4) time-consumption behavior.

To a large extent this all is a rediscovery of recent work by Marcus Hutter; we survey that in §24. Although we believe our definition of "intelligence" largely demystifies that concept, it is more mysterious what a "consciousness" is – although we propose a tentative definition which if correct would trivialize that issue. Finally, to further the development and sanity of AI, we recommend that an annual intelligence contest be held. We explain how to do that in §23; both humans and computers could enter the contest.

Contents

1 Preliminaries 2
2 Motivation: human vs. computer comparison 2
3 The Turing Test 4
4 What is intelligence? Three preliminary thoughts 4
5 An intelligence test – initial version 5
6 Discussion – in which we recognize NP 6
7 Should we try to go beyond NP? No. 8
8 Three more wrong roads 9
9 Formal statement of definition of intelligence 10
10 Short historical rant 11
11 What should AIers do now? 11
12 Theorems about intelligence 11
13 Related previous Universality Results 14
14 Important intelligence-related computational complexity classes 15
15 Faster than brute force search 17
16 Human intelligence vis-a-vis our definition 19
17 Spearman's g and human IQ tests 20
18 Piaget's observations – lessons learned from children 31
19 Forgetfulness 33
20 Time-consumption behavior 35
21 Consciousness – still a mystery? 37
22 How can we build an intelligence? 40
23 Hold contests! (and why that will work) 46
24 About previous work, especially by Hutter 46
25 Multiresearcher Consensus 51
References 51

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

1 Preliminaries

We shall assume the reader is familiar with computational complexity theory [53][129][148][187] to the extent of knowing about NP-completeness and the P=NP question [61] (and occasionally some related matters such as PSPACE and APX [6]) and knowing about polynomial equivalence. We also in some parts will assume familiarity with linear algebra [65][77]. This paper will reach a mathematical level of precision and rigor throughout sections 9, 12-15 but in most sections will not.

Most sections begin with a short "precis" laying out their main accomplishments, and the impatient reader can read only those, delving into the full details only for the sections in which they are interested.[1]

Acronyms (for PG, SC, and ET, see table in §9, while see §14 for computational complexity classes NP, P, PH, #P, BPP, PSPACE etc.

- AES: Advanced Encryption Standard, a secret-key cryptosystem [40] commonly regarded as effectively unbreakable.
- AI: Artificial Intelligence.
- HUH: Human UACI Hypothesis (that the human mind works similarly to §12's mathematical construction of a UACI); discussed in §16-21.
- IQ: Intelligence Quotient (the psychometricians generally define IQ to have mean 100 and standard deviation 15).
- UACI: Universal Asymptotically Competitive Intelligence, defined in §12.

This paper was written in early 2006 but then was found to be, to a considerable extent, a *rediscovery* of ideas by Marcus Hutter during 2000-2006 which he had published both in several papers [80] and in a 2004 book [79]. Hutter's ideas in turn developed from work by Ray Solomonoff [191][192][193][194][195] and D.G.Willis [224] during 1960-2005, some of which are also discussed in [107]. After I realized that, I showed Hutter and Solomonoff a preliminary draft of this paper and we corresponded. I then incorporated the results of that correspondence into a revised draft. The bulk of the present work is almost unaffected by either that correspondence or Hutter's work, with the exception that §24 is extremely affected by it – in fact it is a *survey* of the relations (and differences) between this and Hutter's work, while §25 describes the fate of my subsequent attempts to reach a multi-researcher consensus.

Although there is an eerie degree of similarity between my and Hutter's developments, I believe that my rediscovery has independent value both because (a) it arose from rather different soil and our works complement each other synergistically, (b) the fact that it happened constitutes evidence that we are both on the right road, and (c) also because it then turned out that Hutter and I had some disagreements. Indeed, I believe that Hutter made at least one important mistake (see §24). I believe that anybody who wants to learn about this subject will be best off reading *both* Hutter's and my work in order to get the benefits of both points of view. I do not at all claim to be better than Hutter in every respect, but believe the reverse inequality is also invalid. Finally, there are certain topics which each of us has investigated that the other has essentially not examined at all.

2 Motivation: human vs. computer comparison

Precis. We motivate studying "what is intelligence." A numerical comparison of the crudest possible upper bounds on the raw information processing speed of human brains and 2005-era computers shows the latter are superior. We list notable machine-intelligence accomplishments and failures thus far.

It seems worthwhile to revisit the question of "what is intelligence" because

(a) People want to build computerized artificial intelligences,

(b) Key questions about the interpretation of Quantum Mechanics depend on notions of "intelligent conscious observers," e.g. the validity of the "many worlds interpretation" rests on unproven speculations about how such observers "feel" in quantum scenarios. Without a definition of "intelligent conscious observer" there surely is no hope to make that rigorous.

(c) [2] We have now reached the point where, at least according to the crudest estimates of hardware speed, there seems no inherent reason why computers cannot equal or surpass human intelligence.

Numerical comparison. Intel corporation's 3.6GHz Xeon processor chip introduced in mid-february 2005 had 286 million transistors (and consumes about 100 watts). The human brain has been estimated to contain 10^{12} to 10^{15} synapses, e.g. specifically 2×10^{14} by Pakkenberg et al [147] and for nearby estimates see [25]. So at the crudest estimate of raw processing power, regarding a CMOS transistor-*pair* as roughly comparable to a synapse and assuming rather generously[3] that synapses can continuously operate at 400 Hz, we find that a 3.6 GHz Xeon has processing power $0.5 \times 3.6 \times 286 \times 10^{9+6} = 5.1 \times 10^{17}$ bit-ops/second, whereas a human brain has clearly inferior raw power $10^{12\text{-}15} \times 400 = 4 \times 10^{14\text{-}17}$ bit-ops/second. Plus the Xeon bit-ops are better understood and probably more reliable[4] than the human bit-ops. These bit-op/sec estimates are of course really merely upper bounds on what is achievable in practice; it would be impossible for all neurons in your brain simultaneously really to pulse at 400 Hz because

[1] Our maximally assiduous readers will go further and actually read the footnotes.

[2] A fourth reason that might be claimed is the Search for Extraterrestrial Intelligence (SETI). However our work seems almost irrelevant to SETI because for physical reasons SETI searches necessarily must employ extremely crude tests. To a good approximation, any radio observer of the planet Earth would see an unexpectedly high amount of radio noise modulated in a fairly reproducible manner with a period of about 24 hours. Oddly enough, this fact is not employed in SETI searches. The "`seti@home`" Arecibo piggyback receiver and signal processing project does four tests (all extremely crude compared to the ideas in the present work): (1) searching for spikes in power spectra, (2) searching for Gaussian rises and falls in transmission power, possibly representing the radio telescope beam's main lobe passing over a radio source, (3) searching for "triplets" (three power spikes in a row), and (4) searching for pulses possibly representing a narrowband digital-style transmission.

[3] See [94] p.143 and p.155 for reasons to believe 400 Hz is an upper bound on the sustainable throughput of a typical synapse.

[4] For information about synapse unreliability, see p.90 of [94].

the heat would kill you[5] – and the same is probably also true for the Xeon.

Despite this apparent numerical superiority of Xeons, the fact remains that computers in the eyes of most viewers have *not* succeeded in becoming intelligent, and before now there seemed no hope of them doing so in the forseeable future because

1. nobody seems to have a clear enough idea either of what an intelligence is, or of
2. how to build one.

Indeed (for two simpler and more clearly defined goals) there seems no hope in the forseeable future for computers to become superior to humans at visual recognition of common objects, or for computers to surpass human ability at playing the Oriental board game "Go" [24].

The record shows that all the "famous founding fathers of artificial intelligence" all made far too optimistic predictions about AI progress.[6]

On the other hand, computers *have* succeeded in equalling or surpassing the ability of the top humans in the following areas that once seemed solely owned by humans:

1. Seeking closed form solutions of differential equations (ODE-solving package now available with the symbolic manipulation program MAPLE; uses "Lie group" methods [81][145])
2. Playing tournament chess,[7] as well as certain other games like checkers, gomoku=5-in-a-row, reversi disk-flipping game, backgammon, and scrabble crossword games.
3. Rapid search in large literature databases to find key words and phrases ("Google")
4. Landing and controlling aircraft.
5. Deducing DNA sequences from numerous substring sequences.
6. Deducing molecular structures from atomic composition, charge/mass ratio data for molecular fragments got from a mass-spectrometer, and forbidden substructure claims (Lindsay, Feigenbaum, et al's DENDRAL system)
7. Diagnosis of (and recommending therapy for) human blood and meningital diseases from lab results (E.H.Shortliffe's MYCIN system, shown by a double-blinded study to be superior to Stanford medical professors given the same input data).

In at least the first two of these cases, the methods used by the computers are quite different than, and much more "brute force" in character (although the necessary brute-force methods have their own kind of elegance), than the methods used by humans. Specifically the methods used by the top chess programs all are much closer to "Shannon type A" (brute force search) than the "type B" (highly selective) searches used by humans (humans search quite unexhaustively even at 2 ply depth, whereas every top computer is exhaustive down to depth 4 at the very least) – although it also is true that in order to reach grandmaster strength a substantial amount of

[5] For information about energy consumed by the brain and neuronal processes, see [4][5]. The human brain consumes 13 to 15 watts, with about half of that consumption estimated to be by molecular pumps that maintain membrane potentials. Experimental measurements [5] show that rodents that increase their cortical activity by 1 pulse per neuron per second raise their O_2 consumption by 145 $m\ell$/100g grey matter/hour. Assume approximately the same number is true for humans. A resting human emits about 100 watts of thermal energy – equivalent in view of the 5648kJ/mol heat of combustion of sugar $C_{12}H_{22}O_{11}$ to combusting about 17 liters of O_2 per hour – of which 3.0 liter per hour are devoted to powering the brain (which for typical adult humans weighs 1.5kg, most of which is the cortex). We conclude from this that the cortical activity of resting humans is at most $3.0/(15 \times 0.145) = 1.38$ pulses per neuron per second – and if only half of resting mental energy consumption is devoted to neuronal signalling (the rest being for other purposes) then this should be divided by 2. So if all neurons really were pulsing at 400 Hz, the human brain instead would emit at least 2000 watts of heat, which would quickly kill you.

[6] A.Turing [210] in 1950: "I believe that in about 50 years' time it will be possible to programme computers with a storage capacity of about 10^9, to make them play the imitation game [Turing's test] so well that an average interrogator will not have more than 70% chance of making the right identification after five minutes of questioning." Turing's prediction about growth in typical memory capacities proved exactly correct, but his prediction about Turing test competency totally wrong. The "Loebner prize" is a $2000 prize awarded to the most human-seeming "chatterbot" each year. It is immediately clear from examining transcripts of the winners' conversations that they would fail a true Turing test after only a few sentences – which is why the larger Loebner prizes have never been awarded. M.Minsky in his classic 1967 textbook ([129] page 2): "Within a generation, the problem of creating 'artificial intelligence' will be substantially solved," undeterred by the fact that Herbert Simon in 1960 had written "Duplicating the problem-solving and information-handling capabilities of the brain is not far off; it would be surprising if it were not accomplished within the next decade." Wrong. John McCarthy (later joined by various other big names in AI such as Donald Michie and Seymour Papert) during the period 1968-78-84 lost two famous big-money bets with chessmaster David Levy about when a computer would be able to beat Levy in a chess match. (Herbert Simon also wrongly predicted, in 1957, that a computer would be able to beat the world's best chess player within 10 years, although Simon did not participate in the Levy wager.) Ultimately, however, McCarthy et al's views did pan out when in 1997 IBM corporation's "Deep Blue" hardware-software system beat human world champ Gary Kasparov 3.5-2.5 in a match – it just took 3-5 times longer than predicted. The United States Department of Defense sponsored [42] three major AI research projects during the early 1980s – to develop autonomous land vehicles, an "expert system" for "sea battle management," and a system that would help Air Force pilots by communicating with them via natural spoken language – because they had been assured by leading AI experts that these three goals would, if funded, be achieved within 10 years. All failed despite $600 million in DoD funding. However, ultimately the experts' views did prove somewhat justified in *one* of these three cases in the sense that in October 2005, five robotic vehicles successfully navigated a 132-mile course, with Stanford's winning vehicle doing it in 7 hours to win a DARPA competition with a $2 million prize. These experiences suggest that optimistic predictions by AI pioneers cannot *always* be dismissed, but in the cases when they prove correct, the time required will exceed the prediction by a factor 3-5 or more. Extrapolation of Minsky and Turing based on this exaggeration factor of 3-5 (or more) then yields the new predictions that an AI will first be developed in 2050-2200 (or later). Although Minsky and McCarthy have largely admitted failure (e.g. Minsky by 1982 had changed his tune to "the AI problem is one of the hardest science has ever undertaken"), some, such as Ray Kurzweil and Edward Feigenbaum, continue apparently entirely undeterred, to make grandiose optimistic predictions, some book-length. For example there was Feigenbaum's 1983 book [57] on Japan's "fifth generation computer project" (Feigenbaum was proved wrong when the project's impact proved negligible and its goals failures), and there are more recent Kurzweil books [100] predicting a technological "singularity" caused by the development of superintelligences. Kurzweil relies heavily on extrapolations of "Moore's law" of historical exponential growth in computer capacity and predicted human-level AI in about the year 2020.

[7] According to the February 2006 "Swedish Rating List," the commercial chess program "Hiarcs 10" running on a "1200 MHz Athlon with 256 Mbytes" has a chess rating of 2845 ± 33; Garry Kasparov is the only human ever to have exceeded 2810 and as of April 2006, the top-rated human is Veselin Topalov with 2804.

human-like chess "knowledge" (although far less than grandmasters have) and several non-obvious search-algorithm tricks both seem required. The methods used by humans to solve ODEs are a mixture of a large "bag of tricks" [91][229] plus imagination and search; in contrast the computer techniques use a systematic "Lie group" analysis which had been dismissed at the time of their original invention by Sophus Lie (1842-1899) as requiring too much work for people to use, but which is now within the grasp of computers and seems far more powerful than any "bag of tricks" since those tricks almost all are merely special cases of the Lie Group framework.

Although general purpose mathematical theorem proving and counterexample-finding tools usually are far below human strength, in certain comparatively rarely-visited subareas of mathematics, they occasionally can impressively exceed human capabilities.

Furthermore, human mental abilities in some areas are *ridiculously* poor compared to computers, such as speeds of arithmetic operations (10^{11} times slower), and short term memory – if humans see a 3×4 array of letters rapidly flashed in front of them, they can then write down a subsequently-randomly-chosen row or column of digits, but cannot write down the entire array, even though the entire array must have been in their minds to enable recall of random rows and columns [198]; for other results indicating the pathetically limited nature of human short term memory see [127][130] and for the poor performance of human long term memory see [7][110][112].

3 The Turing Test

Precis. We explain the "Turing test" for intelligence and why it is inadequate.

Alan Turing proposed [210] a legendary "test" for intelligence, designed to be applicable to nonhuman machines.

Briefly, the "Turing test" (which he called the "imitation game") is this. A panel of inquisitors carry on conversations (conducted over teleprinters) with both the machine under test, and a human being. Both the machine and the human are sealed in separate rooms and their only connections to the outside world are their teleprinter links to the inquisitors. After some time, the inquisitors provide their opinions on the question of which teleprinter is connected to the human, and which to the machine. If statistically these votes are indistinguishable from tossing a coin, or if the computer seems more human than the human (and assuming humans are "intelligent"!), then Turing proclaims that the machine also is "intelligent."

Unfortunately, the Turing test has many disadvantages.

1. The Turing test can prove intelligence, but not *lack* of intelligence. It can say "yes" but not "no." It is a "sufficient but *not* a necessary condition" for intelligence.
2. For example, a human from a foreign country not speaking the same language as the inquisitors could be intelligent, but probably would fail the Turing test.
3. The Turing test can provide a yes answer to "is it intelligent?" but does not provide a numerical quantitative *measurement* of intelligence, nor even a $<$, $=$, or $>$ *comparison* between two intelligences. Indeed, a machine substantially *more* intelligent than a human quite plausibly could still "fail" Turing's test.
4. The Turing test demands simulation of a human. But surely considerable intelligence is possible in principle even without being able to simulate a human?
5. The Turing test is extremely inefficient and it requires intelligent entities – the inquisitors and the "control" human – in order to conduct the test. It would be better if the test could be administered and judged mechanically, preferably with extremely high communication bit rates.
6. For many of the reasons above, the Turing test is not useful (or at least is extremely uneconomical) for practical purposes of attempted software development of an artificial intelligence.
7. The requirement that the machine be isolated in a room is necessary to prevent a "cheating" strategy where the machine phones somebody and says "I offer you a million dollars to spend the next hour answering questions on a Turing test." (And then the human must similarly be isolated for reasons of fairness.) However, surely one's effective intelligence is increased by the ability to seek outside help or do library research. Humans are handicapped by isolation and presumably machines might also be, and if we are interested in the limits on building intelligence then we will be interested in unhandicapped intelligence.[8]
8. The Turing test does not provide a satisfactory definition or understanding of what "intelligence" really is. By which I mean, it is not a mathematical definition, it is not a definition that has any intrinsic meaning in the absence of humans, and it has little or no predictive power about the nature of intelligent entities.

Our purpose here is to try to devise a better test, and a better "definition of intelligence" that overcomes these disadvantages. (Its formal statement in in §9.) One reason this seems important is that it seems an essential prerequisite, in order to have any progress in the field of "artificial intelligence," first to *define* intelligence.

4 What is intelligence? Three preliminary thoughts

1. Generality: It seems to me that the, or a, primary feature of intelligent entities is a willingness to investigate *any* kind of mental problem, and an ability to solve, or make progress on, some of them (the more progress on the more of them at more speed, the "more intelligent" they are). Certain monkeys, for example, exhibit some kinds of tool-using or language-using behavior, but they have never been observed to investigate "what makes stars shine?" or "what are the effects of fire on iron ore?" or "does white have a forced win in chess?" or

[8] Actually even with our new test, to prevent cheating some kind of isolation would be necessary. However, that is only a worry for machines trying to cheat by cooperating with a human to pretend to be as intelligent as a human. For future machines far more intelligent than humans, such "cheating" would be pointless.

for that matter, "how can we devise a general purpose intelligence test?". Monkeys and other animals do exhibit some thinking, problem solving, memorizing, data collection, communication, and/or learning, but if so only on a very limited range of topics compared to humans, and usually with very much lower success rate.[9]

Now I admit that this perception is a biased one viewed through human spectacles – with a different sample space of problems, and/or a different initial base of knowledge, perceptions might change. For example, humans are probably less interested than monkeys in the question "how can I catch that termite so that I may eat it?". Humans definitely have worse memories than the American Western bird "Clark's nutcracker," which remembers thousands of sites, distributed over 100 square miles, where they hid seeds (even six months later and even if those sites are now buried under snow). Also humans probably have a smaller initial base of knowledge about what kinds of termites are more compatible with monkey digestive systems, and a smaller initial knowledge base than birds about "how to fly" or "how to build a nest," and smaller initial ability at many synthetic chemistry tasks than a humble bacterium.

However, humans under the right circumstances (and an abstract cooperative intelligent entity, certainly) *would* be willing to think about and try to solve all those problems and overcome all these limitations, if asked, and the solutions humans *eventually* dream up are comparable or better than those the monkeys or birds invent, albeit perhaps the humans may not reach comparable competence for a very long time (and in the case of the comparison with bacteria, it still has not happened). And a perfectly cooperative abstract intelligent entity (which is presumably the sort that AI researchers would want to build?) should indeed be willing to investigate *any* mental question. It is precisely that kind of adaptiveness (and persistence) that seems the hallmark of intelligence. In sum,

2. Asymptoticity: What matters when judging "intelligence" is not the initial competence, but rather the asymptotic *final* competence.

There have been certain ridiculously silly objections to "intelligent machines" raised by philosophers John Searle and Ned Block. Their objections are essentially of the form "what if the machine has got a giant preprogrammed list of all possible answers to all possible questions, and its mechanism of operation is simply lookup in the list? That obviously is not an 'intelligence' even though it could pass a Turing test." A related objection would be "what if the machine simply outputs an infinite string of random bits *but* some human or other supervisor simply discards all of the bits that do not correspond to an 'intelligent statement'?" The answer to those objections is, of course, that the lists would be far too large to fit in the universe, or the time required for the random bit generator to output the works of Shakespeare would be far too long to accomplish in the age of the universe. We can conclude from this either that simple arithmetic seems beyond the abilities of philosophers,[10] or

3. Efficiency: It is important to demand that an "intelligence" accomplish its feats without consuming ridiculously large amounts of time or memory space, e.g. that it be, essentially, a *polynomial time and space algorithm.* The space limit is more important than the time limit since we would like to demand that it easily be capable of fitting into the universe.

5 An intelligence test – initial version

Precis. In this and the next three sections we develop and discuss an informal preparatory version of our "definition of intelligence." The basic ideas are that

1. a usefully intelligent entity is one that can supply good "answers" to "questions" (both are general bitstrings),
2. we recognize that only NP questions and P-verifiable answers are needed,
3. we argue that any entity that cannot score well on certain such tests is *not* intelligent, while any entity that can score optimally well on such a test is extremely intelligent even if it can do nothing else besides score well on the test (thus "proving" that our definition is "correct")
4. somewhat sneakily implicit in that reasoning is a "lookahead" to §12 where we shall construct a "UACI," i.e. an entity which does score optimally well (asymptotically in a competitive sense) on every such intelligence test, and to §16-20 where it is conjectured that the working mechanism behind human intelligence *is* such a UACI.

Intelligence test about field of research R (consisting of 5 steps, 2 of them optional):

1. The tested entity provides the tester with a number $D \geq 0$.
2. The tester provides the entity under test with a "sample problem" from area R – meaning, a bit string. [Optionally: it is of "difficulty level D."]
3. The tested entity then responds with a "solution" – another bit string. [Optionally: there could a time limit somehow imposed for this step.]

[9] The collective strength of the human race seems to be a consequence of *both* the individual characteristic of intelligence *and* the ability of the race for inter-person communication, and (with the invention of writing) the recording and dissemination of (as opposed to forgetting) whatever impressive inventions occasionally arise. Also, the invention of money enabled useful cooperation, sometimes enormous in scale, among many humans who might otherwise never meet one another or be sworn enemies. I believe, however, that the tasks of communication, handling money, and record keeping are comparatively trivial – it is intelligence that is the hard thing for computers to duplicate.

[10] Actually, Block in his 1981 paper "Psychologism and Behaviorism" [20] claimed that (1) such limitations imposed by the size of the universe are irrelevant because merely the logical possibility of such a machine, even if unrealizable physically, suffices for his purposes, and (2) since "nothing in the laws of physics prevents infinitely divisible matter" the limitation from physics may not actually exist. I disagree with both these assertions by Block: there *are* well known limitations on physical entropy such as the "Bekenstein entropy bound" which completely destroy Block's (2), and I claim that Block needs to consider physical reality in our universe in order to be granted any sort of credit; intentional departures to other universes in order to make arguments about the meaning of intelligence seem to me unworthy of any sort of respect. If we are allowed to employ other universes in our argumentation, why don't we simply postulate a universe containing an omniscient being – terminating all arguments immediately – and leave it at that?

4. The tester than responds with a "score" – a number which is zero if the solution is judged wholy unsatisfactory, but monotonically grows larger for "better" and "more impressive" (in the opinion of the tester) solutions.
5. The tester optionally provides additional information, such as a correct solution with its score. We could also optionally allow the test-taker at any time to provide the tester with a self-generated purported problem-solution pair and ask the tester to score it, although this score will not "count" for the purpose of judging the test-taker's IQ.

This whole cycle is then repeated (with different problems each time) an unbounded number of times. Entities who get larger total scores after time T are "more intelligent in field R after time T." Note that (correctly) intelligence is seen both as multidimensional (i.e. R-dependent) and time dependent; A could be better than B at some tasks but worse at others, or A could initially be weaker than B but as time goes by could eventually become stronger (and a further reversal might happen at a later time). Also note – and this is key – that *there are absolutely no rules* about what a "problem" and what a "solution" are, aside from both being bitstrings. There is absolutely no standard language (such as English) that needs to be used. This totally avoids "language dependence." The test-taker must try to *deduce* what the "problem" means by repetitive observation of problem-solution pairs (preferably initially mostly "easy" ones).

Example 1: R could be "understanding the Rubik cube puzzle." A sample problem of difficulty D could be a bitstring representing a Rubik cube D random moves away from the start position. A bit string purporting to be a "solution" could be granted a score of 0 if it is not a sequence of M moves (in some particular format) restoring the cube to the start state; otherwise the score could be $\lceil 12^D/(1+M)\rceil$.

Example 2: R could be "recognizing the equivalence of two knots." A sample problem could be a bitstring representing two circular $(D+7)$-entry sequences of integer 3-space coordinates, and the solution is one bit which is 1 if the two polygons are equivalent knots, otherwise 0. (Score 1 for correct solution.)

Example 3: R could be "proving mathematical statements." A sample problem could be a bitstring representing some "axioms" and a "conclusion"; a "solution" could be a sequence of valid implications successfully proving (or disproving) the conclusion (score=1) or failing to do so (score=0).

Example 4: R could be "doing indefinite integration." A sample problem could be a bitstring representing some formula involving a variable x, e.g. $\cos x$; a "solution" is another such formula, e.g. $3 + \sin x$. If the symbolic derivative of the solution seems equal to the original formula (when both are evaluated at some random values of the variables) then score=1, otherwise score=0.

Example 5: R could be "recognizing printed English words."[11] A sample problem could be a (somewhat distorted) image of an English word, and a solution gets score 0 if not an ASCII representation of an English word, +1 per correct letter, and +100 if the entire word is correct.

Example 6: R could be "constructing short programs." A sample problem could be an input-output pair produced by some computer program, and a solution is a computer program C that produces that output if given that input. (One might also demand that C also work on all prior input-output pairs from previous test questions.) The score is $1/L$ where L is the length of C (if C is correct, otherwise score=0).

Example 7: R is "inferring the next value in a sequence." The kth problem is the kth integer in some infinite sequence (such as "the prime numbers") and the answer is the $(k+1)$th integer.

Example 8: R is "sorting N numbers." The kth problem is an integer $N > 0$ and N integers to sort; the score of the answer is 0 if not a permutation of the N input integers, otherwise the number of correctly-sorted integer pairs with an extra N^2 bonus for a fully-correct sorting.

In all of these examples, the tester could be, fairly easily, completely automated.

There is no requirement that the field of research R remain the same from problem to problem, nor even that its identity ever be revealed to anybody. (Of course, doing these things might be desireable to make the IQ test "easier," but it is not necessary.)

One can now propose a preliminary mathematical definition of "intelligence" which is, essentially, to do well on such IQ tests. Any entity which keeps doing better than humans would have to be judged "at least as intelligent as a human" for example. We will provide a fully formal definition later in §9, but for now let us regard this merely as a preliminary idea for a definition, which we shall now explore.

6 Discussion – in which we recognize NP

At least at first, the test seems the ultimate in simplicity and generality. Among its advantages:

1. It requires *no* prior knowledge of *anything* on the part of the test-taker. The test taker simply must *learn* by repeated examples of problem-solution-score cycles (a) what is needed to get good scores, and (b) how to provide it.
2. In particular there is no language (such as English) and no formatting requirement (such as ASCII coding of symbols) of any kind.
3. It effectively enables comparison among different intelligences.
4. It is easily automated, and **there could be an annual "intelligence contest" put on by, say, the ACM** with different people contibuting both new IQ-testees and IQ testers each year. I strongly recommend doing this if the field of "artificial intelligence" intends to be taken at all seriously.

[11] A version of this problem has already been used highly successfully as a Turing test whose goal is to *prevent* robots from signing up en masse for various internet services intended to be available only to humans.

5. Most importantly, we claim that this test really does encapsulate "intelligence." Few would dispute that being able (at least eventually) to get good scores (for some notion of "good") on this test is a *necessary* condition for intelligence. Further we soon shall argue (which is more debatable – see §7) that it also is a *sufficient* condition, i.e. that any entity capable of getting good scores on this kind of test really is "intelligent" even if that entity cannot do anything else besides score well on the test!

However, second thought reveals that our proposed test is *not* the ultimate in generality. At least if we restrict ourselves to tests that can be administered by easily constructed and computationally fast testers, our test problems are, roughly speaking, merely the "**NP problems**" [61].

It requires some thought to see this; we shall now explain that and also argue that it is *necessary* to restrict ourselves to NP problems.

First, the sets of suitable problems are those which

1. Contain an infinite (or so large as to be effectively infinite, e.g. for our "Rubik cube" example) number of members,
2. From which a random member[12] may be generated efficiently[13] together perhaps also with secret auxiliary information such that:
3. a purported solution to the problem may efficiently be scored by somebody who knows (a) the problem, (b) the purported solution, and (c) the secret auxiliary information.

These criteria correspond quite closely with the definition of "NP" but there are several worries that need to be raised and then dealt with. A minor technical worry is the allowance of randomness.[14]

A. The first major worry is that two of our examples – "proving mathematical statements (#3)" and "constructing short programs (#6)" are actually Gödel-Turing undecidable classes – i.e. far harder than NP, and a third, "knot equivalence (#2)," is not known to be in NP.[15]

However, in practice these tests would be more like "producing proofs *less than 500 pages long* of mathematical statements," or "producing short programs *which run quickly*," or "deciding knot equivalence for two knots which the tester has quickly generated in such a way that he *knows* – despite the very finite amount of time he has thought about it while operating within a very-finite-memory limit – that they are equivalent (or inequivalent)" in which case these tasks *are* in NP. The point is that with any such polynomial-length and time limits imposed, we genuinely have NP.

B. NP is quite a large class of problems, and certainly any creature capable of solving arbitrary NP problems would have to be viewed as "extremely intelligent" in "an extremely useful manner" and as "a far superior mathematician than any human," and as "a far superior theoretical physicist than any human" and as "a far superior computer programmer than any human" (and probably also as "a far superior engineer than any human") even though not all-powerful. To justify some of these claims, an NP-solving oracle[16] could quickly tell us "is there a (≤ 500)-page proof or disproof of the Riemann hypothesis?" and "is there a theory of physics stateable in (≤ 100) pages that comes with a (≤ 500)-page proof that its predictions agree with the following set of experimental observations XXXX?" and "find the most efficient runtime bound, and the (≤ 100)-page-long computer program which instantiates it, and a proof of validity for both (if one exists < 100 pages long) for an algorithm to accomplish the following formally specified task XXXX?"

C. The second major worry is that the entity under test does not actually *know* what the problem is – he must deduce that from a long sequence of problem-solution pairs. *After* he has (if ever) succeeded in understanding that, *then* it's a sequence of NP-problems from that point on. However, we claim the problem of deducing the problem format is itself an NP-problem. Why? Because the tester shall be assumed to be generating the problem-solution pairs as the output of some polynomial-time algorithm (equipped with a random bit generator). Under that assumption the problem of guessing what that algorithm is, and what its random input-bits[17] were, is an NP problem.

D. And finally we claim that it really is *necessary* for the test-problem-generator to be a polytime (perhaps randomized) computer program, because otherwise the notion of a "*correct answer*" on our IQ test would have *no meaning*! This problem is already pervasive among IQ tests that have been devised by psychometricians to apply to humans, see §17. For example for "find the next number in the sequence $1, 3, 7, 13, \ldots$" and "draw a picture of a man" – both of which are very commonly used "IQ tests" – there simply is no objectively uniquely correct answer and the "best" answer is a matter of opinion.[18]

[12] From some probability distribution chosen by the tester.

[13] Please substitute "in polynomial time with the aid of a random number generator" in place of "efficiently" to get a mathematical definition.

[14] Mainly, we shall attempt to dodge this annoying issue by taking the attitude below that, by using a cryptographically-strong random bit generator – and we shall assume such exist (since many conjecturally-good ones are available [2]) – a fully-deterministic tester effectively can garner any benefit of randomness. This dodge actually will not quite work, as we shall see below, but we shall discuss that and/or just assume the reader is capable of devising the appropriate easy alterations to restore formal correctness. Constantly carping computational complexity curmudgeons can content themselves by going to §14.

[15] I conjecture, however, that knot equivalence is in NP. Haas, Lagarias, and Pippenger [71], building on an approach dating back to Haken, proved that deciding whether a polygon is the *unknot* and deciding whether two polygons are *unlinked* are both in NP. It remains unknown whether these tasks are in co-NP. I personally suspect that any N-vertex polygon can be converted to any other topologically-equivalent N-vertex polygon by a polynomial(N)-length sequence of operations of the form (a) add one new node in the middle of an edge, (a^{-1}) remove one node from the middle of an edge (if there is no bend-angle there), (b) perform a linear-in-time homotopic motion in which all vertices move according to a linear function of t for $0 \leq t \leq 1$ and no edges cross. If this is true, then knot-equivalence is in NP.

[16] Actually, if P$\neq$NP then our UACI in §12 will not be an NP-oracle, but it will be within an asymptotically constant performance factor of any P-algorithm for trying to solve that kind of NP-problem – which is good enough to make all our points here in time polynomial in "100.".

[17] Actually, due to these random bits, it is not technically correct when we in our text say "NP"; correct insetad is the new complexity class "ME(FP)"; see §14.

[18] The answers 25, 21, 19, and 17 may all be justified for the next-number problem: 19 because the sequence is "every other prime number regarding 1 as prime," 17 because the sequence is "numbers containing the digits 1, 3, and 7 in increasing order," 21 because the sequence is

Many such tests are not really testing the ability to find the right answer so much as "the ability to think the same way as the person who created the test" – which has the excellent advantage from the test-creator's point of view that he is "the most intelligent person in the universe" but is not terribly useful for the rest of us.

But if the test problem generator is a *short computer program*, then there *is* an objectively correct answer, namely whatever that program outputs, and it can *eventually* be deduced in the sense that eventually by consideration of every possible computer program C, one can find the right C. Actually even then there could be many equivalent Cs (which would not matter), or (what does matter) inequivalent Cs which merely happen to agree on all problem-answer pairs so far, but will disagree for future problems. However, if we agree that the "best" answer is the *shortest* suitable C ("Occam's razor") then *eventually* all the other inequivalent Cs which are anywhere close to the real C in length, will become dismissible because of an accumulation of evidence against them, so that, asymptotically, a unique answer will indeed exist.

Now in order for this to be feasible (and to be compatible with the observation that the test-problem generator runs quickly) – and feasible to try to justify to anybody questioning the validity of the test! – it seems essential that C be polynomially short in length and run in polynomially bounded time, in which case we have exactly NP.

Conclusion: This discussion has now essentially proved that (aside from technical quibbles concerning permitting randomness, see §14) our class of feasibly allowable IQ-test problems exactly coincides with (and must coincide with) the "NP problems."[19]

7 Should we try to go beyond NP? No.

Human beings (and hence abstract intelligent entities) can and must (and in fact do) work on problems *not* in (or at least not obviously in) NP, all the time. For example, humans are working on "devising cosmological models to explain the birth of the universe" and "designing flood protection systems for cities" and "building software to understand human languages" and "composing a good tune" and "finding a good move in given chess positions." If somebody produces a bit string claiming to be a solution (or partial solution) to such a problem, then it is not clear at all how to judge if it really *is* a "valid solution" or how high a "score" to give it. And especially it is not clear how to do those things mechanically.

So we should **worry that by restricting our notion of intelligence to NP-problems only, we might be sacrificing something important.** The remainder of this section will analyze this and ultimately conclude this is **not a worry** or anyhow can be sidestepped.

It might, however, still be possible to include problems of *some* of these beyond-NP sorts in an IQ test, provided the tester has access to *two or more* allegedly-intelligent entities to test. But there are limitations... let us discuss this.

If it is going to be of interest to judge intelligence using tests that go beyond NP in this way, then presumably our entities are already pretty darned intelligent, since mere NP-type tests are seen as too limiting for them. So we may assume a very considerable base level of intelligence among the testees in the hypothetical discussion that follows.

Example 1: "chess tournaments": The tester could easily merely check that a chess move in a given chess position is *legal* and hence could easily conduct chess tournaments between two or more cooperative tested entities. (Use a variant of "chess" in which all games terminate within some fixed number of moves.) This would enable the tester to compare the abilities of the entities at "producing good chess moves" even if the tester himself has essentially no ability to measure how "good" a chess move is.

Solving chess (for an $N \times N$ variant of chess in which an extra rule forces all games to terminate after a polynomially large number of moves; or more simply for the artificial game "generalized geography" [55][187], or see [164][83][162]

"$n^2 - n + 1$ for $n = 1, 2, 3, \dots$," and 25 because the sequence is $2T_n - 1$ where T_n is the nth "tribonacci number" $T_n = T_{n-1} + T_{n-2} + T_{n-3}$ where $T_{-2} = T_{-1} = 0$ and $T_0 = 1$. I also have nice justifications for any of $\{5, 15, 23, 27, 29, 43, 51, 61, 73\}$. I suspect Picasso would have gotten a low score on the "draw a man" test and it is known that children from different cultures get large score differences on this test.

[19] Although we are unaware of any previous formal definition of intelligence, some people have groped a considerable way toward our definition in the past. However, the AI community's handling of the matter actually seems to have gotten *worse* with time. Two of the founding fathers of AI, Herbert Simon and Allen Newell, proposed in 1959 the "General Problem Solver" thus understanding, in theory if not in practice, the idea that the goal of AI should be about developing the ability to solve *general* problems. In a 1991 paper Lenat and Feigenbaum [104] stated in passing "Definition: Intelligence is the power to rapidly obtain adequate solutions in what appears a priori to be an immense search space." This definition is quite similar to our more formal one, except for the fact that technically, by their definition *humans* are not intelligent because they cannot crack the AES[20] cryptosystem [40]! The difficulty is that some such problems *cannot* (at least, assuming P≠NP) be solved rapidly by *any* method. So I believe that what Lenat and Feigenbaum really wanted, but lacked the machinery to express, is precisely §12's "uniformly asymptotically competitive intelligence" (UACI) which, essentially, is nearly as good at solving such problems as any polynomial-time algorithm possibly could be. But more recent (mostly) sources seem further away: Winston's AI book [225] opens with a frank admission of failure: "What is intelligence?... A definition in the usual sense seems impossible because intelligence appears to be an amalgam of so many... talents." Neither David L. Waltz in his essay "the prospects for building truly intelligent machines" [216], Ramsay's "formal methods in AI" book [158] nor Michie's AI book [126] offer any definition of "intelligence," but Michie does say on p.3-4 "if we can form a sufficiently complete and precise theory of any given aspect of intelligence, then we can convert it into a computer program... if we cannot, then although as *Homo Sapiens* we may display this or that capability, we cannot claim truly to understand, in the given respect, what it is to be human. The question... remains an open one." Haugeland's philosophy-of-AI book [72] dispenses with the issue on page 6 with "How shall we define intelligence? Doesn't everything turn on this? Surprisingly perhaps, very little seems to turn on it. For practical purposes [Turing's test] satisfies nearly everyone." End of story, and Haugeland quashes any remaining worries with a box on page 7 titled "why IQ is irrelevant." Rich & Knight's AI book [163] says rather pathetically in chapter 1: "We propose the following by no means universally accepted definition: AI is the study of how to make computers do things which, at the moment, people do better." (And essentially the same definition is in Schutzer's AI book [174].) Russell & Norvig's AI book [169], currently the most popular, does not define "intelligence" but does "define AI as the study of agents that receive percepts from the environment and perform actions." Logician Hilary Putnam opens his essay on the question [155] with "The question I want to contemplate is this: Has AI taught us anything of importance about the mind? I am inclined to think the answer is no."

for checkers, othello, and hex), is known [60] to be a PSPACE-complete problem and hence in a sense we can efficiently test *comparative* intelligence on any PSPACE problem, which is a (presumably[21]) larger class than NP.

Example 2: "peer review": The tester could ask the entities themselves to judge how good alleged cosmological theories are (in their opinions). Again, however, this "group therapy / mutual scoring / peer review" approach would only work if enough of the tested entities were competent-enough cosmologists and if the entities were isolated and hence had independent opinions. Even then this technique is very dangerous. If, for example, entities A and B were far superior to all the others, but A was far superior to B, then it might be impossible for the tester to *tell* that $A \gg B$. Another counterexample: If entities A and B were far superior to C,D,E,F,G,H but C,D,E,F,G,H all had similar approaches and similar weaknesses, then they might all vote themselves high scores and thus lead the tester into exactly the wrong conclusion.

Our conclusion (not very surprising to either experienced scientists or lawyers): **"Peer review" doesn't work.**[22] **"Formal adversarial interactive proceedings" work better.**

Fortunately, there seems to be a way out of the peer-review trap.

That is: In order that a cosmological theory (or city flood protection plan) be judged as good, it must satisfy certain *criteria* (which could be stated by the theory's author – and/or the criteria could be demanded by the poser of the problem "please seek cosmological theories," or agreed upon by a committee of the allegedly-intelligent entities) which are easily *checkable*, i.e. are solutions to NP problems, i.e. are also known as "theorems with proofs." (A cosmological theory might later be refuted by experimental data, but that possibility is not relevant to the question of comparing two such theories in the light of the data available *at that time* and pointed out by the two theory-authors.)

This makes us happy about all our nasty examples except for "building software to understand human languages" and "composing a good tune." The trouble with those two is that humans seem needed to judge what a "good tune" is or "whether you understand my language." We do not disdain these two (or other) inherently-human mental activities – we simply *forbid* them as components of any intelligence test or intelligence definition intended to be usefully and fairly applicable to non-humans even in the absence of any humans.[23]

8 Three more wrong roads

1. Since "intelligence" is about "the ability to solve problems," one might have proposed the naive idea for an IQ test that a problem-task be described (in some language) and then the testee tries to solve it. Bad idea: it is important to be able to solve problems that do not have (or do not have obvious) descriptions and definitions at all – in real life, often a large part of solving the problem first is to *find* a good problem statement. This whole paper is, in fact, an excellent example of that. In real life, one determines "what the problem is" oneself and then determines "how good the solution is" oneself also, usually with the aid of some disagreeable interaction with the external world. Furthermore, we do not want there to be any "language."

2. Since we seem to have reduced the matter of IQ testing to posing NP (or PSPACE if we have more than one entity under [comparative] test) problems, and it is well known [61] that just *one* "complete" class of NP-problems (or PSPACE problems) is reachable by efficient (polynomial time) problem-transformation from *any* NP (or PSPACE) problem – can we now conclude that we can just reduce the matter to administering test-problems of the form "solve this boolean satisfiability (SAT) problem" or "play $N \times N$ chess starting from this position"?

No – the issue is more subtle than that. The objection is that there are not just problem transformations – there are also transformations of the probability distributions of those problems. A natural probability distribution on, say, "planar graph hamiltonian path" problems might yield a very strange and unnatural distribution on 3-SAT problems. And our IQ tests are really more about problem classes *equipped with probability distributions from which we sample.*

Fortunately, several authors starting with Leonid Levin in 1986 have analysed notions of "averageP-complete" problems. To cut a long story short, it has been shown [212] that any "P-samplable" probability distribution[24] on any NP class of problems, may be solved by randomized reductions to the *uniform* distribution on certain *specific* classes of "complete" NP problems, e.g. certain graph edge-coloring problems. Also it is known that for any standard NP-complete problem, there is a P-samplable distribution that makes it average-case complete [18] and for NP search problems, P-samplable distributions do not generate harder instances than simply picking instances uniformly at random [82]; any NP search problem with a P-samplable distribution is randomized-reducible to an NP search problem with a uniform distribution.

As a consequence of all this, might it seem justified to restrict our IQ test problems to being solely graph edge-coloring

[21] This is a standard very widely believed conjecture.

[22] This, of course, is not only a problem for those of us trying to define "intelligence" but in fact is a problem faced every day by everybody trying to judge success and quality in areas lacking clear definitions of quality and success (or merely in which the judge is not very competent)!

[23] To linguists and musicians who do not like that, let me say this: Linguists: Since many people believe that language skills are an important component of intelligence and the main thing that distinguishes humans from other animals, let us note that an "intelligence" in our sense would in many ways be extremely competent at languages. For example, any NP-oracle would have no trouble with the following kind of task: "given a large number of sentences and non-sentences (pre-classified) in some language, deduce the shortest possible set of 'grammatical rules' that explain $\geq 99\%$ of that data, where the allowed 'rules' are selected from a certain fairly wide class of, e.g, quadratic-time algorithms." (Also, we note that people with brain injuries – "Broca's aphasics" [152] p.35-37, and "Brother John" [12] p.373 – exist who have lost their ability to produce, and sometimes to understand, langage, but they nevertheless clearly remain intelligent. Brother John could not even *think* linguistically but could think and function fine non-linguistically during his epileptic attacks.) Musicians: any NP-oracle that had access to some large amount of musical compositions with numerical "quality scores" on a 1-10 scale, would similarly be able to deduce a quadratic-time function attempting to best-fit the quality data... and once that function was known would be able to produce "good" musical compositions on demand.

[24] And it has been shown that "P-samplable" and "P-computable" distributions are, for this purpose, equivalent.

problems of the Venkatesan-Levin type [212] selected from a uniform distribution?

Again emphno – the issue still is more subtle. One difficulty is that by restricting ourselves to such complete problems we are providing only "hardest" problems, and discarding "easy" ones which are potentially solvable much more quickly. That discarding sacrifices the ability to discriminate between less intelligent and more intelligent entities, which might have widely differing abilities to solve some comparatively *easy* distributional classes of problems, but both of which are extremely poor on average-case *complete* problems. Another difficulty is that we really are *not* talking about a probability distribution of problems from which we sample – we instead really want a generator of an infinite *sequence* of problems, which is an incompatible notion.

So our final answer is "no"; providing only "complete" (or not complete) problems from some fixed distribution is *not* going to be adequate for an intelligence test. From this we reach the very important conclusion (at least under usual assumptions like P≠NP) that *simply repeatedly rerunning any fixed bounded-runtime randomized problem generator is inadequate for intelligence-testing purposes.* (See also §24 re the "universal intelligence test" controversy.)

3. Although allowing "chess tournaments" allows us to test comparative IQ on PSPACE problems instead of "merely" NP, I see little if any advantage to doing so at the moment. The field at present is not ready to work on that perhaps-higher kind of intelligence.

9 Formal statement of definition of intelligence

Precis. Motivated by the discussion in the last four sections, we now present a formal mathematical definition of "intelligence."

Our considerations have led to a conclusion; we shall now boil them down into a formal "definition of intelligence."[25] (Some of the ideas above will be omitted or modified for brevity and simplicity, but could be restored in a "deluxe version" aiming for improved performance.)

We employ the usual (since the days of Church and Turing) underlying computational model – a machine polynomially equivalent to a Turing machine. Then as everybody knows, an "algorithm" is a computer program that terminates no matter what the input, and it is a "polynomial time" algorithm if it does so in time bounded by a polynomial function of the input's bit-length. Let us now define something less familiar.

A **reent-algorithm** means a computer program that *never* terminates, and which keeps soliciting and accepting input bitstrings from one or several parties, outputting bitstrings in between. A reent-algorithm is "polynomial time" if the time it consumes to produce each output is bounded by a polynomial function of the total bit-length of all the inputs it has received so far. ("Reent" stands for "reentrant"; the concept is of a program that carries on an *interactive* dialogue [or several dialogues], as opposed to the old concept of "algorithm" which is a *batch* concept.) We shall also sometimes permit reent-algorithms to access a source of random bits.

Cast of characters:

PG: Problem generator
SC: Solution checker
ET: Entity under test

An **intelligence test** consists of one polynomial-time reent-algorithm PG and one polynomial time ordinary algorithm SC. The first reent-algorithm, called the "problem generator" (PG), uses random bits, and spits out an infinite sequence P_1, P_2,..., of output bitstrings called "problems;" the kth time the entity under test tells the problem generator "ready" it (beginning the next "*cycle*") spits out the next problem P_k, and it also spits out a second bitstring D_k called the "secret associated data" – but the entity under test (ET) is only allowed to see P_k and is never allowed to see any of the D_k. The second algorithm is called the "solution checker" (SC). As its input, it reads the problem P_k spit out by PG, and it *also* gets to read the secret associated data D_k. Finally, it reads as its third input, a communication from the entity under test called the "answer" A_k. It then outputs a "score" integer $S_k \geq 0$ which is a function of P_k, D_k, and A_k. It is essential that ET is never told what the underlying algorithm inside PG is, although it is probably permissible to make the algorithm inside SC public.

ET, after receiving P_k, is allowed to submit any number of trial answers A_k to SC for scoring, i.e. is allowed to invoke SC at any time to score any proposed trial answer for the current P_k. Only the *final* answer ET submits for P_k before requesting the next problem P_{k+1} from PG, corresponds to the final score S_k it gets on problem k. ET's *cumulative score* after time T, is $\sum_{k=1}^{K} S_k$ where K is the number of cycles completed before T.

[25] The *Merriam-Webster Dictionary* defined "Intelligence" as "the ability to learn or understand or to deal with new or trying situations ... also: the skilled use of reason" – and our formal definition seems happily compatible with that. But I am not sure whether it agrees with Lewis Terman, who defined intelligence as "the power to think abstractly." At a famous 1921 symposium "Intelligence and its measurement" organized by the American Psychological Association, 17 top experts were asked to define "intelligence," with the result that 3 refused and the remaining 14 provided 14 different definitions, most of them embarrassingly uncleanly and poorly worded. This was summarized despairingly by Ch.Spearman as "chaos itself can go no further... 'intelligence' has become a mere vocal sound, a word with so many meanings that it finally has none." In 1986 Sternberg & Detterman convened another symposium "What is intelligence" to answer the same question, obtaining 24 definitions from 25 experts (one was a 2-expert joint effort), thus making it clear that almost no progress had occurred during the intervening 65 years. E.g. in the opinion of A.R.Jensen ([86] p.48): "My study of these two symposia... has convinced me that psychologists are incapable of reaching a consensus on [the definition of intelligence]. It has proved to be a hopeless quest." Some of the 24 year-1986 proposals were summarized by the *Encarta Encyclopedia* as: "general adaptability to new problems in life; ability to engage in abstract thinking; adjustment to the environment; capacity for knowledge and knowledge possessed; general capacity for independence, originality, and productiveness in thinking; capacity to acquire capacity; apprehension of relevant relationships; ability to judge, to understand, and to reason; deduction of relationships; and innate, general cognitive ability." But Rex Li distilled the 24 definitions down to a "consensus" that intelligence was "thinking and learning." Meanwhile according to the consensus statement [67] organized by L.Gottfredson, intelligence involves the ability to "reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience." Both these again seem happily compatible with our formal definition. See footnote 19 for comparison with definition attempts from the AI community.

There could be many possible (PG,SC) pairs, each one generating a different intelligence test.

Entities that get higher cumulative test scores as a function of time T are "more intelligent" at least as far as that test is concerned. If some entity 1 is at least as intelligent with respect to every test (or at least every test from some set under consideration) than entity 2 (and more intelligent on some tests) in the limit $T \to \infty$, then it is simply "more intelligent."

We have allowed ET to be literally any entity. However, for the purpose of studying Artificial Intelligence it is convenient to consider "entities" which in fact are *reent-algorithms* – probably random-bit-using ones – and preferably *polynomial-time* random-bit-using reent-algorithms.

10 Short historical rant

Philosophers and psychoanalysts have argued about intelligence and consciousness for thousands of years with the non-result that, even in modern times, we still see "experts in the field" saying clearly ridiculous things and/or exhibiting their ignorance in an extremely public manner.[26] The main reason their approaches did not work was because these camps *abandoned* both the experiment-based "scientific method" and formalized rigorous reasoning. E.g. Freud and his disciples simply stated how people's minds worked, arrogantly assuming they somehow knew all, and seeing no need to conduct – and in some famous cases even actively opposing – the sort of double-blinded objective experiments that later refuted vast numbers of their claims; philosophers such as Block and Dreyfus never saw a need to learn the subject of computational complexity theory or to think about limitations set by physics on computers, and instead worked in blissful ignorance of that, often with "results" immediately seen by non-ignorant readers to be laughable.

Many of the same flaws have more recently been exhibited by **"artificial intelligence" researchers**. Their extraordinary mismatch of hype versus accomplishments has passed into legend, and also their disdain for mathematics (such as the 1980s AI conference where by majority vote it was resolved that "probability has absolutely nothing to do with artificial intelligence"; I myself once had an AI paper rejected without refereeing because it was "too mathematical"...) and their frequent use of irreproducible "experiments" involving highly dubious uses of statistics.

11 What should AIers do now?

So **we hereby propose** that AI researchers instead concentrate on

① achieving real understanding (i.e. prove *theorems*)

② and/or, when doing experimental work with computers not amenable to complete theoretical understanding, to make *reproducible* and *standardized measurements* of Progress toward the Main Goal.

We claim that both of these now are possible.

12 Theorems about intelligence

Precis. The advantage of having a formal mathematical definition of intelligence, is that we now can prove theorems about intelligence. We now do so. The most important one is theorem 5, where we construct a UACI – uniformly asymptotically competitive intelligence – which, we show, is asymptotically as intelligent (up to a constant factor, which in fact in suitable models of computation is just 1) as *any* other entity on every intelligence test simultaneously. Unfortunately this UACI consumes time exponential (2^ℓ) in the codelength ℓ of the competitor entity, but in theorem 7 that is shown (under widely believed computational complexity conjectures) to be unavoidable, i.e. best possible.

Now that we have a formal definition of intelligence, the stage is set to at least start trying to prove theorems. And to confirm that, we will now state and prove some fundamental theorems about intelligence.

1. A "godlike superintelligence" is a polynomial time reent-algorithm which achieves the maximum score achieveable by any set of test answers on any intelligence test. **Theorem:** *if P≠NP then godlike superintelligence is impossible, even if*

[26] We have already mentioned Block [20]. The philosophy professor H.L.Dreyfus achieved fame after losing a 1967 exhibition chess game to R.D.Greenblatt's early chess program MacHack after claiming that computers could never beat humans at chess. Dreyfus then, undeterred, immediately said that computers would never beat the top players. Dreyfus published two books titled "what computers can't do" [51] consisting largely of (largely justified) invective against the AI community, plus proclamations of "I told you so" concerning their failures; but at the same time it should be noted that Dreyfus also has been proved wrong in his negative forecasts for AI on various other occasions. According to Gerald Edelman ([54] p.67) "proposals that the brain and mind function like digital computers do not stand up to scrutiny" because ([54] p.225) "Turing machines have by definition a finite number of internal states, while there are no apparent limits on the number of states the human nervous system can assume... The transitions of Turing machines between states are entirely deterministic [whereas humans can incorporate randomness; here Edelman seems unaware of the standard idea of a Turing machine with access to a source of random bits]... human experience is not based on so simple an abstraction as a Turing machine; to get our 'meanings' we have to grow and communicate in a society." Such colossal ignorance about the basics of computer science and Turing machines [129][187][211] might sadly be expected from a Nobelist in Physiology & Medicine (1972), but it is somewhat more surprising to see Edelman exhibiting massive ignorance about his own area. Thus part III (the only part perhaps with any new content) of his book [54] tentatively concludes that consciousness first evolved 300 million years ago in some land-dwelling vertebrate and is possessed by "most mammals and birds" although it is "dubious" for snakes and not present in lobsters. Edelman then fails to examine (in fact completely ignores) the well known high intelligence of octopuses [144]and fishes (we discuss fish in our §21; Octupuses navigate, build dens, solve mazes, sleep, can be trained to distinguish symbols and do tricks, and learn to open twist-lid jars apparently by observing humans doing it). The great physicist Roger Penrose [151] somehow developed the idea that *quantum gravity effects on microtubules* cause subneuronal components of human brains somehow to perform super-Turing computational feats, supposedly an essential ingredient of human consciousness! (Penrose thinks that a machine relying purely on classical physics won't ever have human-level performance, and only by understanding of quantum gravity can we reach an understanding of human consciousness.) But that whole idea is immediately seen to be ludicrous by a vast number of orders of magnitude [202]. Along the way Penrose provides more fallacious arguments that human brains cannot be Turing machines, because sometimes humans solve instances of undecidable problems and can invent proofs that involve reasoning outside of some fixed logical system (unfortunately [221], Turing machines can also do those things... [123][154]). As for the Freudian psychoanalysts, suffice it to say that Neumann's book [131] must be viewed with incredulous awe: is apparently the most-reprinted and popular of all of the books the present work cites, but yet it is quite likely the single worst book that I have ever seen.

"God" knows the algorithm inside SC. **Proof:** achieving the maximum score is equivalent to being able to solve any NP optimization problem optimally – and if P≠NP that is not possible for a polynomial time algorithm to do. Q.E.D.

2. An "*approximately* godlike superintelligence" is a polynomial time reent-algorithm which always achieves within a constant factor of the maximum achieveable score on any intelligence test. **Theorem:** *if P≠NP then approximate godlike superintelligence is impossible, even if "God" is allowed to know the algorithm inside SC.* **Proof:** it is known [6] that there are optimization-problem classes in which approximation of the optimum to within a constant factor is NP-complete. Q.E.D.

3. An "*asymptotically* godlike superintelligence" is a polynomial time reent-algorithm which asymptotically on a long sequence of intelligence test cycles generated by the same *deterministic* polynomial time test-problem-and-answer generator reent-algorithm, always achieves *asymptotically* the maximum achieveable cumulative score (provided that answers that would yield unboundedly large cumulative score totals exist).

Theorem: *Asymptotically godlike superintelligence* is *possible.* **Proof:** The idea is to guess the test-problem generation algorithm[27] by successively systematically trying all possible algorithms. Each cycle a new guess is tried until one is found that agrees with all problem-solution pairs so far, then we just stay with it until a disagreement occurs, then resume guessing. Eventually (i.e. after some very large but finite number of cycles) the right guess will be found and stayed with forever after, causing optimally godlike superintelligence from then on.

There are a few tricks we need to explain in order to justify this:

1. We need to know that every polynomial time algorithm may be written in a "self proving" fashion which is a priori *known* to be a polynomial time algorithm with a *known polynomial* as its runtime bound. My favorite way to do that is to make the algorithm have a standardized straight-line-code *preface* that obviously (1) reads its input and (2) computes a polynomial P of its bitlength N; and then the *remainder* of the algorithm decrements P as a side-effect of every step it takes, self-terminating as soon as P hits zero. [Note: similar remarks instead may be made about *exponential-time* algorithms, but *not* about *all* algorithms.]
2. In that way we can generate all polynomial time algorithms in, say, lexicographic order, but without generating any super-polynomial-time programs.
3. It also is possible in numerous ways, in our intelligence test problem-answer-cycle framework, to systematically do "time sharing" among all possible such algorithms in such a way as still to keep the combined creature a polynomial-time reent-algorithm. For example, if an algorithm has worst-case time bound KN^D, then we can run it for N steps each cycle (and if not yet done, continue running it N steps the next cycle but still using the old data, and so on, until it gets done). If each cycle we add a new trial algorithm to our collection, the next effect is that N cycles get completed in $O(N^3)$ time so that we plainly have a polynomial-time reent algorithm, but with the property that every trial algorithm eventually is run on an unboundedly large number of P_k, thus assuring that one with 100% success rate eventually will be found.

Q.E.D.

Extension: Indeed, by continuing to explore all algorithms permanently with 50% of one's computer time, but using the other 50% to run the currently-best algorithm, we not only can obtain asymptotically godlike superintelligence in the above scenario, but in fact we can do so while staying within an asymptotic *factor of two* of optimizing a measure of *computational efficiency.* Even better one can run it a fraction $\max\{1/2, 1-10^{10}/\sqrt{n}\}$ of the time on the nth test-cycle, thus getting *100% efficiency* asymptotically in an appropriate computational model.

Warning: the preceding theorem and extension depended heavily on the wrong assumptions that the test-problem generator is *deterministic* and also generates the *answer.* In reality PG is randomized and a separate scoring routine SC evaluates answers (which need not be unique and which quite possibly cannot be deduced from the problems in polynomia time). In the preceding theorem, using a cryptographically strong pseudorandom number generator would have been be fine, *but* the proof breaks if the test-problem generator is allowed access to a *true* random-bit generator. Indeed,

4. Theorem: *An asymptotically godlike superintelligence is* not *possible if the polynomial time test-problem-generation-and-test reent-algorithm instead has access to a true random bit generator.*

The **proof** is trivial: on the Nth cycle, demand an N-bit answer and award score 1 if the answer matches a sequence of N freshly-generated coin tosses, otherwise award score 0.

Then the total expected score for any intelligence whatever, even cumulated over an infinitely long intelligence test, is ≤ 1, but the maximum possible score is infinite. Q.E.D.

We now state our **most important theorem**, although not in its strongest possible form.

5. *A "uniformly asymptotically competitive intelligence" (UACI)* is a randomized reent-algorithm C which, when repeatedly given any intelligence test, achieves an *expected* cumulative score at least asymptotically equal to that achieved by *any* particular other randomized polytime reent-algorithm X on the same test sequence, while consuming compute time at most a constant factor larger than those consumed by X, and memory resources growing at most linearly with time.

The UACI Theorem:[28] *A UACI* is *possible.*

Proof: The idea is to guess the competitor's algorithm by successively systematically considering all possible polynomial time reent-algorithms. Each cycle a new guess is considered. If one is found that would have yielded a larger expected cumulative score throughout past test problem-solution-pair history, then we switch to using it to generate our test answers.

Notes: we estimate cumulative expected scores as follows: each cycle, we, for all candidate-algorithms in our current

[27] For brevity, we shall often use the word "algorithm" when we mean "reent-algorithm."
[28] See also §13.

collection, try another set of random input bits on all of their past history, and update that candidate algorithm's expected-score-estimate appropriately. By the law of large numbers, ultimately the expected-score estimates will (with probability 1) approach their true values for any particular candidate algorithm up to any particular time. By switching, at some point, to exhaustive enumeration of all 2^n bitstrings with n bits rather than Monte Carlo sampling, we can in fact determine the *exact* expected-score up to the earliest point of consumption of the nth random bit, not merely an estimate (while still consuming only polynomial space). So eventually the right guess for X will be found (or something as good or better) and stayed with forever after, resulting in at least competitive performance from then on.

Note 2: to make this all work with at most polynomial slowdown, we need to use the same tricks as in the preceding proof, plus a few more. Since the candidate algorithms are eating data at different rates, we of course need to keep track of all their "cumulative scores" as well as their consumed "times" in order that we may compare apples with apples. Ultimately asymptotically all algorithms consume the same amount of "time" so the comparisons will be asymptotically fair.

Note 3: You might worry that, on some task, there might be some sequence of polynomial-time algorithms, say with runtime N^k for the kth algorithm, which achieve greater and greater scores, e.g. cumulative score proportional to k^2 after k cycles if we switch to algorithm k at cycle k. Therefore, our UACI might find these algorithms successively, with the net effect of finding an algorithm that really has superpolynomial runtime.

Avoiding that issue is in fact precisely why in §9 we defined the cumulative score to be a function of *time* T and *not* of the number of test cycles so far. If the score-producing beast SC pre-transforms its scores by some appropriate monotonic pre-transformation function it can encourage the intelligence to prefer just *one* of the algorithms in that sequence in order to get good scores without taking too much runtime to do it – while if a super-polynomial-time reent algorithm then yields superior asymptotic performance per unit time to any polynomial algorithm, it indeed will be preferred, but that is then a feature, not a bug.

Note 4: We have discussed "provable polytime algorithms" and their generation in a previous proof; we of course also reuse that trick here.

Note 5: We will never be *sure* that duplication of the best possible competitor has occurred, hence will need to continue experimenting forever, causing a slowdown by a possibly-large, even though polynomially bounded, asymptotic factor. But by devoting 50% of runtime to the best currently-known candidate algoritm and 50% of runtime to the ongoing search for improved ones, the asymptotic slowdown factor can be made to be 2, and indeed (as we explained last proof) even $1+\epsilon$. Q.E.D.

Although this existence theorem is very fine, it is not terribly useful because the proof technique – even though constructive – takes a very long time (exponential in the code-length of the competitor algorithm) before the competitor algorithm is duplicated allowing asymptopia to set in.

It is possible to address this criticism to some extent as we shall see in §22 and 15. Before doing so we also point out

6. The proofs of the preceding theorems also show that the code for an UACI can without loss of generality and without loss of performance (except for polynomially bounded factors) be required to be *short*. I.e. the "Kolmogorov complexity" (code length) of a universally asymptotically competitive intelligence is remarkably small.[29]

Indeed, it is not difficult to write down in full and complete detail, a program for an UACI, in some standard computer language such as `C` or `Scheme`.

One simple[30] way to do it in `C` is to employ a simple Turing-universal cellular automaton such as Conway's "life" [19] or Wolfram's "rule 110 = 01101110_2" (proved Turing universal by Matthew Cook [35]) or a simple universal Turing machine [129] and to systematically enumerate start-configurations. Another way is to employ "Post string-rewriting systems." In `Scheme`, programs and "treelike data structures" are essentially the same thing (actually, directed cycles, i.e. "back-pointers," are also allowed if there are recursive calls) and hence all programs may be generated systematically by tree-enumeration.

7. Although the strategy of "searching over all possible algorithms" employed in theorem 5's construction of a UACI may seem (and is) very crude and inefficient, there are good reasons to believe that it is **not possible to do better**. More precisely: We can prove under standard computational complexity conjectures that it is *not possible* to find the best algorithm (or even one merely *comparable* to the best one), even if among *quadatic-time* algorithms describable in N bits, in worst case time below exponential in the description length N of that algorithm. Indeed, it is a standard conjecture that "breaking the AES secret key cryptosystem" (i.e. given plaintext-ciphertext pairs for an N-bit long secret-key cryptosystem of the same ilk as AES [40], find the N-bit-long secret key that encodes the encryption algorithm) cannot be done in subexponential (below 2^N) time.[31]

Therefore, the only hopes for improving theorem 5's crude UACI construction are either

[29] And indeed a companion paper [190] analyses the description-length of the biological "blueprint" for *human* intelligence and concludes that either 2.4 or 32 megabits suffice, up to a factor of 3 worth of imprecision in the estimate, in two different models (specifically: it depends whether you believe "exons and introns" are important for regulation or not). This is not very large. Solo humans have written considerably larger computer programs.

[30] Maximally simple – but very inefficient!

[31] "AES-like cryptosystems" work as follows: each "stage," the plaintext is transformed by one of two reversible transformations T_0 or T_1 each with highly-bit-scrambling effects. At the kth stage one performs T_b where b is the kth bit of the secret key. The net result of composing all N of these transformations (arising from an N-bit key) is the "ciphertext." Note that the "key" here really is an N-bit-long description of an encryption (and the corresponding decryption) "algorithm" and anything capable of guessing that algorithm is capable of "breaking the cryptosystem," i.e. of rapidly producing plaintexts corresponding to given ciphertexts. So far, we have described "Feistel cryptosystems." However, as a description of the AES, it has been oversimplified; it is an important *protective modification* that the transformations T_b actually incorporate the *whole* key not just one bit, because otherwise AES would be attackable by the "fast Gray code update" and "meet in the middle" attacks that we shall describe in §15. It is a very widely believed conjecture that it is not possible to break such cryptosystems in less than 2^N steps on average.

1. To try to reduce the exponential growth rate toward whatever its minimum possible value is (which is, presumably, some constant bounded above 1) – but by considering distorted forms of the AES cryptosystem with N-bit secret key and N-bit plaintext for $N \to \infty$, we see that the best possible growth-constant presumably is 2 and hence no reduction is possible, at least in the worst case;
2. To ignore the UACI's worst case performance and try instead to improve its performance on *good* cases while still not diminishing performance too greatly in bad cases. But a limit on the ability to do that is set by the

Two-way UACI simulation theorem: *Any two UACIs A, B are equivalent in the sense that A can (and will) simulate B and B will simulate A with at most a constant additive slowdown (note: this constant may depend on A and B and may be very large) plus $\le (1+\epsilon)$ multiplicative-factor slowdown (this is valid for any $\epsilon > 0$).*

13 Related previous Universality Results

Precis. Our UACI is a natural outgrowth of previous "universality" ideas in computer science dating back to Turing.

Turing 1936 [211]: There is a "universal" Turing machine capable of emulating any other with at most polynomial slowdown (and if equipped with one extra tape, the emulation even can be done with only constant factor slowdown).

Turing further argued that "algorithms" = "the set of Turing machine programs which terminate on any input" = "the set of programs for a *universal* Turing machine which terminate on any input" = "the concept previously informally called computation." He did the latter by arguing that anything that an idealized human mathematician equipped with a pen, eraser, and an infinite amount of paper, could do (provided he could only write some bounded number of symbols per square centimeter, and provided his mind could only be in some bounded number of distinguishable states), a Turing machine could also do.

Cook 1971 [34]: NP is the set of problems whose solutions are verifiable in polynomial time. There is an NP-complete problem class (SAT[32] is one such [61]) such that any problem in the set NP can be transformed in polynomial transformation time into a SAT instance, and such that the solution of that SAT instance can be back-transformed in polynomial time to a solution of the original problem.

For transform to SAT, the transformation time indeed is only linear in the runtime of the solution-verifier.

It is conjectured that P≠NP, that is, that the set of problems whose solutions can be verified in polynomial time, is larger than the set of problems soluble in polynomial time. That conjecture remains open and whoever solves it can get a million-dollar prize.

Levin 1973 [105] (also discussed in [78]): There is a universal algorithm A which, essentially, runs program p a fraction $2^{-\ell(p)}$ of the time, where $\ell(p)$ is the binary code-length of p, then checks its output with an externally supplied checking program g, and halts with output x as soon as an output x is found that causes $g(x) = y$ where y is A's input. In particular A will solve any NP-problem (defined by a polynomial time checking program g) in time bounded by $2^{\ell(p)}T_p(x,y)K$ where $T_p(x,y)$ is the time needed to run p and then to run g on p's output x and then to test $g(x)$'s equality with y, and where we assume we are in an underlying computational model permitting emulation of arbitrary p with at most constant factor K slowdown.

Note that anybody who suspects that P=NP, does so because they suspect that there is some clever algorithm, which so far humanity has been too stupid to think of, which will solve SAT problems in polynomial time. But now we see that in fact humanity has *not* been too stupid to think of it! If P=NP, then the algorithm A will do that job in polynomial time, and indeed with the optimum possible polynomial-degree. Furthermore, regardless of whether P=NP, the algorithm A will solve SAT problems in time at most a constant factor longer than *any* polytime SAT-solving algorithm.

Historical note: Steve Cook won the Turing award for his 1971 proof that SAT (and a few other problems) were NP-complete. It was little known for a long time in the West, though, that Levin had in 1973 similarly sketched a proof that a certain 2D tiling problem (as well as a few others) was NP-complete – independently inventing the same sort of ideas Cook did – in a Russian paper so compressed it was only *two pages* long [105]! But even more incredibly, in the same ultra-short paper Levin *also* sketched the above idea for a universal algorithm to solve any problem in NP with at worst polynomial slowdown compared to the best algorithm – i.e. Levin's result was *two sided* and hence superior to Cook's one-sided result. This 2-sided nature of the situation, with NP-completeness on one side, and universal algorithms on the other, deserves to be much better known than it is.

Hutter 2002 [78]: Given any algorithm B that runs an input x, Hutter can (and does) write down a new algorithm H that (1) is provably equivalent to B, and (2) will run in time at most

$$5T_p(x) + d_p TT_p(x) + c_p \tag{1}$$

steps, where $T_p(x)$ is the runtime of p on input x and d_p and c_p are some (enormous) constants, and TT_p is the runtime needed to compute the runtime upper bound for p on input x, and p is any algorithm whatever that is provably equivalent to B.

Hutter gave the following example to illustrate how his result could be superior to Levin's (we improve and correct his discussion): the naive method for multiplying two $n \times n$ matrices over some unspecified finite non-commutative ring (where a black box is available to perform ring operations), takes order n^3 operations, but perhaps there exists some unknown method requiring only order $n^{2.01}$ operations.[33] If so, it can easily be proven that such a method exists that has both a runtime proof and a correctness proof (both unknown, but

[32] "SAT" is standard computer science lingo for the "boolean satisfiability problem," which is Cook's [34] standard NP-complete problem [61].

[33] Hutter actually had $O(n \log n)$ time in mind, but I see no reason to believe that a $O(n \log n)$-time matrix multiplication algorithm necessarily would have correctness and time-bound proofs.

they both exist). Therefore, Hutter's universal algorithm will run in $O(n^{2.01})$ time. Levin's algorithm, however, will need order n^3 time using an $O(n^3)$-time verifier.

But this example by Hutter was somewhat misleading because in fact, it is possible to verify $AB = C$ where A, B, C are $n \times n$ matrices using a *probabilistic verifier* with failure probability $\le 2^{-K}$ in $O(n^2 K)$ black box steps by multiplying both sides by a random n-vector of ring elements (and doing this K times or until an inequality is detected, whichever comes first). Levin's algorithm using this sort of verifier/denier also would run in $O(n^{2.01})$ time with arbitrarily low constant failure probability, provided the constant K was chosen large enough,[34] and quite plausibly Levin's approach would be preferable for practical reasons.

Discussion of attempt to improve on Hutter: What Hutter did not say, was the following:

1. It is possible to enumerate *only* the polynomial time (or *only* the EXPTIME) algorithms in a "self-proving" form in which it is immediately obvious by examining their code, that they are polynomial time (or EXPTIME) and what their runtime upper bound is. We have already explained how to do that in §12.
2. Proving that two algorithms for tasks in NP are semi-equivalent, is trivial if you just add a standard polynomial-time solution-verifier to the end of each of them: that forces them to be "equivalent" in the sense that they either output a solution to the NP problem, or nothing. (Similarly for NEXPTIME.)

Insert these two observations into Hutter's argument. Then Hutter's result becomes worse in the sense that we now are no longer applying it to *any* algorithm B, but only to polynomial time (or EXPTIME) B's – but it becomes better in the sense that Hutter's original result was only applicable to B's and algorithms equivalent to B for which there existed a *proof* of equivalence and *proofs* of runtime upper bounds, which (as Hutter pointed out) meant a *strict subset* of all algorithms. Our point is that we now just handle *all* polytime (or all EXPTIME) algorithms whose goal is to try to solve an NP (or NEXPTIME) problem, *without* needing to restrict ourselves to a subset.

Unfortunately, this Hutter-improvement attempt does not quite work. The first improvement idea does work, but the second fails because semi-equivalence is not the same as full equivalence (semi-equivalence can still be used, but not with Hutter's speed, only with slower-speed methods such as Levin's).

Schmidhuber's "Gödel machine" (2003-4): Jürgen Schmidhuber, in a paper presently still only available as a technical report, invented a very interesting idea he called a "Gödel machine." This is, essentially, a machine designed to run some program X that takes input from some "external environment" and produces output, thus getting a "reward" determined by the external environment, and then does this again, and so on forever. Its goal is to get the most summed reward it can. Schmidhuber's idea was that this machine's initial X would during some fraction of the time, perpetually search among all possible other programs X' and proofs P, and whenever it found a proof that X' would produce greater expected future reward than X during all future times, it would switch to executing X' instead.

Schmidhuber then speculated that perhaps "human intelligence" is really just the same thing as a Gödel machine.

That speculation, however, is false for several reasons, the most immediate being that we cannot run a Gödel machine without a utility model for our external environment, which humans certainly do not have in any form amenable to rigorous proofs. Nevertheless Schmidhuber here is thinking in a very similar direction to the present paper groping toward a definition of "intelligence."

Ray Solomonoff and Marcus Hutter 1960-2005: R.J.Solomonoff in a series of papers starting in the 1960s and continuing beyond 2000 produced some very interesting ideas for "universal learning machines" and "universal probability distributions," and invented something similar to "Kolmogorov complexity theory" before Kolmogorov did. Solomonoff's ideas come very close to our own and then Markus Hutter continued his line of thought even further. Both have ideas of the universal algorithm ilk, both recognize something of that ilk ought to be a good framework for building an AI, and both have frameworks with some resemblance to our "intelligence test" too. See §24.

14 Important intelligence-related computational complexity classes

Precis. As we previously noted in footnote 14, we adopted the oversimplification throughout §5-8 of only talking about the deterministic computational complexity classes P and NP, even though really, we should have been permitting *randomization*. We now give the appropriate replacement classes and some additional discussion.

	randomized	deterministic
1	BPP	P
2	N(BPP)	NP
3	ME(FP)	NP
4	PSPACE	PSPACE
5		EXPTIME

Figure 14.1. Important intelligence-related computational complexity classes (explained in the text). ME(FP) appears not to have been studied before and is defined here for the first time. ▲

BPP=co-BPP=BPP$^{\text{BPP}}$ (Bounded-error, Probabilistic, Polynomial time) is the class of decision problems solvable by a Turing machine eqipped with a random bit generator in polynomial time, with an error probability of at most 1/3 for all instances. (By repeated runs the error probability may be made exponentially small.)

If we take the view that the scores for answers to intelligence tests have to be *efficiently justifiable* to outside observers or

[34] Levin's method would take some constant C amount of time before "understanding" the $n^{2.01}$-time algorithm, and if $K \gg \log C$ then "false positives" become neglectible; by increasing K logarithmically as we try more algorithms any particular algorithm will be used with its own-K verifier which will get the right effect even without knowing anythign about C.

tested entities (since otherwise the legitimacy of the test can be questioned), then P and BPP are precisely the complexity classes that describe the tasks performed by the scoring device SC of §9.

N(BPP) and **NP [61]:** Our notation regards "N" as an *operator* which converts a class T of tasks into the class NT whose answers are *verifiable* by a computation in the class T. The problem faced by an intelligence whose answers are scored 1 or 0 by a P or BPP scoring device, is to get the 1 score if possible. That is an NP or N(BPP) problem. The "AI planning problem" (which is certainly a subset of what it takes to build an AI) is stated to be NP-complete in [223].

ME(FP): This class apparently has not been studied previously. our notation regards "ME" as an *operator* which converts a class T of functions (here T=FP, which is the class of polynomial-time functions f that convert bitstrings to binary integers) into the class of problems of the form "maximize the expected value of $f(x)$ by appropriately choosing its input bits x" where some known *subset* of those input bits are choosable whereas the complement subset are chosen randomly by coin tosses and are not controllable by us.

The problem faced by an intelligence whose answers are scored with a binary integer by a polynomial time scoring device that employs random bits, is to maximize its expected score. That is a ME(FP) problem; but if no random bits are used it is just NP. Of course that was assuming that SC and PG are known – but if they are not known, then the problem of guessing them given the known data (with the aim, e.g. of maximizing the correctness probability of the guess) is also an NP or ME(FP) problem.

ME(FP) completeness Theorem. *The following problem "probability-maximization SAT" is complete over ME(FP), i.e. any ME(FP) problem can be solved in polynomial time if we have access to an oracle for solving probability-maximization SAT instances.*

INSTANCE: There is a known N-bit-input, 1-bit-output polytime forward-only boolean logic circuit, which also accepts N more inputs from random coin-toss bits, for $2N$ inputs in total.

PROBLEM: to find the N-bit input which maximizes the probability the output bit is "on."

Proof: Convert poynomial time algorithms to polynomial size boolean logic circuits in the standard manner by Cook [34][61]. Add as a postprocessing step to any FP circuit with N-bit output x, a circuit comparing x to y and outputting 1 if x is greater; and then let y be N-bit random. Then the 1-bit output of the new circuit is "on" with probability p where p is a linear function of the expectation value of the old circuit. Q.E.D.

Remark. One can now, similarly to NP [61], construct a large variety of ME(FP)-complete problems. For example, it is ME(FP)-hard, given a $2N$-vertex graph, to determine the way to 3-color [200] its first N vertices in such a way that the remaining N vertices can be 3-colored in the maximum possible number of ways.

Complexity class inclusion Theorem. P⊆NP⊆PH⊆#P⊆ME(FP) *where "$A \subseteq B$" here is taken to mean*[35] *that problems in class A can be solved easily (in polynomial time) if we have an oracle that will solve problems in class B on demand.*[36]

Notation: "#" is an operator such that #T is the class of problems of *counting the number of solutions* of some problem in the class T. (Two problems known to be #P-complete are "counting SAT" and "permanent of an integer matrix.") Supserscripting A^B denotes the class of problems A *but* allowing the solver access to an *oracle* for solving problems in class B on demand.

Remark: Sipser [186] also showed P⊆BPP⊆PH.

Proof sketch. All these results are well known except for the ones directly involving ME(FP).

To prove that #P⊆ME(FP): Just set up a circuit such that if the N-input bit string is $000\ldots0$ then it evaluates some given circuit A on the N random bits, and if the inputs are $111\ldots1$ it evaluates some other given circuit B on the N random bits, otherwise it just sets the output to 0.

Then let B be some standard circuit (such as inequality test versus a constant) for which we know the output probability $p(B)$ exactly. We can decide if $p(A) > p(B)$ and hence by binary search on B we can use a ProbMaxSAT oracle to solve #P problems in polynomial time.

To prove ME(FP)⊆N($P^{\#P}$): For each choice of the choosable inputs we can use a #P oracle to find the exact probability (over the remaining coin-toss bits) that the output bit will be on. By use of a P-algorithm employing this #P-oracle (hence the notation $P^{\#P}$) we can decide if this probability exceeds some threshhold t. Then by use of the N operator we can find choosable inputs such that t is exceeded, and finally we can then do an outer binary search to maximize t. Q.E.D.

We conjecture all the $\subseteq$ in theorem 2 are strict, i.e. may be replaced with $\subset$.

PSPACE=N(PSPACE) (the equality is 'Savitch's theorem") is the class of problems soluble in polynomial space. If the intelligence tester were allowed to pose questions *depending* on the previous answers (in Hutter [79]'s terminology this would be an "active environment") then "games" would be being played between the tested entity and the tester, and it is well known that (a) some such games [55][187][164][83][162] are PSPACE-complete, while on the other hand (b) if the game-state is describable by a polynomially-long bitstring and the game ends in at most a polynomially-large number of "moves," then solving the game (and this includes games allowing "dice rolls" and with "incomplete information") is in PSPACE.

PSPACE is only of interest for higher forms of intelligence than the ones generally discussed in the present work – "higher" because the intelligence test answers' scores no longer are efficiently justifiable and hence these intelligences could no longer be efficiently measured by means of intelligence testing. However, it would still be possible to *compare two* intelligences setting up a "chess tournament" between them.

[35] This convention is convenient but slightly nonstandard.

[36] P is polynomial time. PH is the "polynomial hierarchy" of problems soluble in polynomial time by a machine that has access to an NP oracle (this is P^{NP}), in polytime by a higher-level machine with access to an oracle for *that*, in polytime by a still-higher-level machine with access to an oracle for *that*, etc. (These are the successive levels of the hierarchy.)

EXPTIME is the complexity of solving general games in which the game-state is describable by a polynomially-long bitstring.

EXPTIME would only be relevant for still-higher forms of intelligence and for them it would not even be efficiently feasible to compare two such intelligences.

15 Faster than brute force search

Precis. The naive method for trying out A algorithms, each one running for time T, takes at least AT steps. We shall now give theorems showing that, at least if we restrict ourselves to exhaustive searches over certain important subsets of algorithms, this search-over-algorithms can instead be accomplished in $O(A)$ time, i.e. $O(1)$ steps per algorithm.

A standard trick used to speed up the exhaustive tabulation of some function F of binary N-bit-words, is to use "Gray code." The naive method takes time $2^N T(N)$ where $T(N)$ is the runtime to compute F. There are many known simple algorithms (universally dubbed "Gray code"; they are surveyed in [171] and an upcoming book by D.E.Knuth) for visiting the 2^N words in an order such that each word differs from the previous at a *single bit*. In that case if there is a faster way to *update* the value of F after a single bit-change than entirely recomputing F, the algorithm speeds up. The best known Gray code is the "reversal Gray code": for N-bit words we prepend 0 to the Gray codes for $(N-1)$-bit words, then prepend 1 to the Gray codes for $(N-1)$-bit words *in reverse chronological order*.

Actually, even just naively using binary incrementing (instead of Gray code) can still be reasonably fast if the update trick is used because *on average* only a single "carry" is performed during an increment, so that the average number of updates is 2 bit-alterations per increment, i.e. naive binary incrementing is a "constant amortized time" (CAT) update method. Still, though, Gray code is to be preferred.

Also note that both the reversal-algorithm Gray code scheme, and naive binary incrementing, have the property that the average distance of the (leftmost) altered-bit from the right end of the word, is $O(1)$ on average.

A typical result achieved by such a method is

Feistel cryptosystem cracking Theorem. *The problem of "cracking" (i.e. finding all N-bit keys which would explain a given plaintext-ciphertext pair) an N-bit Feistel cryptosystem (as described in footnote 31 but* without *the protective modification) can be sped up from $2^N N^2$ to $2^N N$ time by using Gray code.*

Proof. The "Feistel cryptosystems" we have in mind encrypt the N-bit plaintext by performing N successive reversible transformations (each taking $O(N)$ time) where the kth transformation is determined by the kth bit of the N-bit key. The naive cracking algorithm is to try all 2^N keys, using each one to perform an encryption in $O(N^2)$ steps, which takes $N^2 2^N$ steps in total. To try the next key, the *Gray-code*-based cracking algorithm, which stores all N intermediate transforms of the plaintext, only needs to update the ones resulting from the changed key-bit and its successor key-bits, which is $O(1)$ updates on average, not N. Q.E.D.

Further remarks on cracking Feistel cryptosystems. If one has E different encryptions of the *same* plaintext (resulting from E different secret keys; this often arises in situations where all coded messages begin with the same header such as "Salutations from central command!") then all E of the keys can be cracked simultaneously in the same time bound $N2^N$ by using a hash-table to spot successful regenerations of a target ciphertext.

But very considerable further decryption speedup is possible by using a "meet in the middle" attack. That is, we partially-decrypt the ciphertext for s steps in all 2^s possible ways, storing the results in a precomputed hash table. We than only need to explore an $(N-s)$-bit keyspace seeking partial encryptions that match something in the hash table. For the purpose of cracking the US government's Data Encryption Standard DES) cryptosystem (which had an $N = 56$-bit key), by using a 2^{25}-entry hash table we would only need to explore 2^{31} partial keys, which could have been done using the Gray code trick on an ordinary year-2006 personal computer (no special hardware required!) in perhaps 10 minutes!

Even the later AES system (with 128-bit key) would be insecure to crackers with government scale resources, thanks to these ideas.

All this makes it clear that it is *essential*, when designing such cryptosystems, to make each elementary transformation depend inextricably on the *entire* key, not just a single bit of it. (In footnote 31 we drew attention to this modification of the basic idea.) That prevents separation of the effects of each key bit, *thwarting* both the Gray code and meet-in-the-middle attacks. And hence, as far as I know, the actual AES system remains secure.[37]

Other kinds of Gray codes. Ideas analogous to Gray code have been used in many other search contexts such as permutations, fixed-cardinality subsets, etc. The way to use this idea in *our* context – searches over *algorithms* – is as follows. Suppose we are seeking algorithms written in `Scheme` in which "program code" and "*tree* data structures" are the same thing. That is, each node in the tree is a 1-step computation whose input values are the outputs of its child-subtrees and whose output value is exported to its parent node. (Actually, in algorithms with "subexpression reuse" we would have *DAG*s – directed acyclic graphs – not *trees*, and in algorithms with "loops" the directed graphs could include cycles. But let us oversimplify by ignoring that; we thus restrict ourselves to "purely functional" N-step loop-free algorithms.)

Suppose we visit these algorithms in an order such that only one parent-to-child tree link is altered each time. Then to run the new algorithm on the same input data, we can *reuse stored subtree values* for all subtrees except for the one that is altered.

Lucas's "Gray code for binary trees" [114] makes it possible to visit all N-node binary trees exactly once by performing a single tree "edge rotation" each time [114] to move you from one tree to the next. (There is a well known "planar

[37] Although I am bothered by the fact that AES is based on operations over a finite field, when the design easily could have been altered to incorporate some non-field operations and thus presumably to make cryptanalysis more difficult.

duality" bijection between N-node binary trees and triangulations of a convex $(N+2)$-gon, in which each "edge rotation" in the tree corresponds to a "quadrilateral diagonal flip" in the triangulated polygon). Actually, we propose to use a modification resembling [115] of Lucas's original scheme. Our modified generator actually sometimes performs *more* than one edge-rotation between generated trees, but still only performs a constant number of them *on average.* (See also [96] about non-binary trees; everything we say about binary trees can be generalized to them.[38]) This modification is simple to program, and involves only $O(1)$ computational work per tree generated. Furthermore – and this remarkable fact was not stated in [115] but was realized by its author Frank Ruskey and myself in private emails[39] – a modified version of the [115] tree-generation algorithm features *average depth* (i.e. distance to the tree root) ≤ 3 to each rotated edge.

Using this tree-generator, it takes us only $O(1)$ average time per treelike-algorithm both to visit the algorithm and to rerun it on a given input dataset. The entire root-x subtree needs to be recalculated where x is the set of nodes involved in all the edge-rotations that were performed to get us to the next tree. This updating takes a number of steps linear in the node-cardinality of this little tree (which when x is a single node or path of nodes, is just the root-x path), which is $O(1)$ on average.

Actually, the above sophisticated analysis was unnecessary if we are considering algorithms with more than one possible kind of operation at treenodes.[40] In that case, we can, for each tree-topology,

1. Order the tree-nodes in order of increasing depth (distance to the tree root) with equal-depth nodes being ordered from left to right.
2. Employ the (reversal type) k-ary Gray code to consider every possible way to assign the k types of operators to the treenodes to convert the tree into an algorithm.

In that case by the previously-mentioned properties of the reversal-based Gray code (which we had only discussed for binary, i.e. the $k=2$ case, but it readily generalizes to other $k \ge 2$ and mixed-radix integers) each new algorithm output-value (i.e. tree root value) could be recomputed in $O(1)$ average time per algorithm *even if our tree-generation algorithm were pathetically poor*, e.g. even if it took N^9 steps to generate each N-node tree. We summarize with

Faster-than-brute-force Theorem. *Tree-structured loopless algorithms with N tree nodes each selected from a finite palette of possible functions of their children (and with each node having a bounded number of children), may be exhaustively generated and run on fixed input, in $O(1)$ average time per algorithm.*

Now that we have proven this $O(1)$ runtime per algorithm result in the special case of "purely functional loopless" (i.e. treelike) algorithms, let us consider expanding our empire.

Empire expansion attempt #1: allowing commutative (or more generally symmetric) functions at treenodes. In practice, symmetric functions such as $+$ and $\times$ are commonly employed at tree nodes. (Non-commutative functions such as $-$ and $/$ also are common.) The scheme above would wastefully generate *both* trees $A+B$ and $B+A$, and since many $+$s could be present in the tree, and also since we could have ternary $+$ nodes such as $A+B+C$ $(= B+C+A = C+B+A$ etc), the amount of wasteful regeneration of equivalent algorithms could be enormous.

We can avoid that waste by *only* employing symmetric functions at tree nodes whose child subtrees happen to be *sorted* in lexicographic order. Since a lexicographic sortedness check could be applied to identify all tree nodes as either "symmetric-function compatible" or "not" in polynomial(N) total time, in view of our previous remarks about inefficiency in the tree generation being OK (and assuming at least one unsymmetric function is present in our palette for each node-valency and that our palette is large enough so that an exponentially large number of labellings occur on average for each tree), we *still* get $O(1)$ average generation and running time even when we thus only explore those algorithms that are *not* trivially isomorphic as a result of a node symmetry and/or commutativity. We also can get rid of trivial isomorphisms arising from *associative* laws such as $(A+B)+C = A+(B+C)$ by demanding that $+$ nodes cannot have $+$ children (where we allow k-ary $+$ nodes), etc, and again this still takes ony $O(1)$ time per nonisomorphic algorithm if there are enough nonassociative functions in the palette.[41] So this empire expansion

[38] Indeed we remark that the *general* rooted ordered trees with N nodes can be represented as, i.e. are in 1-to-1 correspondence with, the *binary* rooted ordered trees with $N-1$ parent-to-left-child arcs by making the rightward paths in the binary tree correspond to the nodes in the general tree. This remark actually is not quite sufficient for our purposes, but it goes a long way.

[39] The proof is as follows. The [115] tree generation algorithm is based on the fact that the N-node binary trees can be got from the $(N-1)$-node trees by adding an Nth node somewhere on the path of successive right-children of the root, and then the rest of that path (below the now-inserted Nth node) needs to be made a left-child of the new node. This allows us to generate all N-node (rooted ordered) binary trees by recursively generating all $(N-1)$-node trees – with the new Nth node being irrelevant to that since it just passively hangs off the end of the right-child-path as the trees fluctuate. And then, in between fluctuations, for each $(N-1)$-node tree, we "walk" the Nth node up and then back down the right-child path by means of rotations. The crucial lemma, an early form of which was pointed out to me by Ruskey, then is that the average length of this right-child-path is ≤ 3. That is because the average number of nodes in the path from the root to the rightmost leaf node in a random N-node (rooted ordered) binary tree is precisely $3N/(N+2)$. This follows from the fact that the number $T(N,k)$ of binary trees with a k-node right-child path, $0 \le k \le N$, is $k\binom{2N-k-1}{N-k}/N$ and then

$$\frac{\sum_{k=0}^{N} k^2 \binom{2N-k-1}{N-k}/N}{\sum_{k=0}^{N} k \binom{2N-k-1}{N-k}/N} = \frac{3N}{N+2} < 3.$$

This and the easier fact that the number of N-node binary (rooted ordered) trees is $T(N) = \binom{2N}{N}/(N+1)$ both may be proven by consideration of recurrences such as $T(n) = \sum_{a,b\ge 0,\ a+b+1=n} T(a)T(b)$ and $T(n,k) = \sum_{a,b\ge 0,\ a+b+1=n} T(a)T(b,k-1)$ if $k \ge 1$ and $T(n,k)=0$ if $k > n$ with $T(n,n)=1$ for all $n \ge 0$. They also may be attacked via generating function identities concerning $F(x) = \sum_{n\ge 0} T(n)x^n$ and $G_k(x) = \sum_{n\ge 0} T(n,k)x^n$ e.g. $F(x) = 1 + xF(x)^2$ so that $F(x) = (1-\sqrt{1-4x})/(2x)$, and and $G_k(x) = T(0,k) + xF(x)G_{k-1}(x)$ so that $G_k(x) = [xF(x)]^k$.

[40] I.e. the sophisticated analysis was only necessary in the highly unrealistic case where only *one* kind of operation at treenodes is allowed.

[41] We still have (intentionally) ignored the distributive law, and the reader is warned that *non*trivial algorithm isomorphisms might be possible

attempt must be regarded as "highly successful."[42]

Empire expansion attempt #2: straight-line code. "Straight line code" algorithms consist of N "lines of code" where each line computes a value that is a function (selected from a fixed palette of allowed functions F_1, F_2,..., F_K) of some bounded number of previously computed quantities (or that line inputs a value, a possibility we regard as an extra member of our function palette) and the output is the value computed by the final line. Straight line code is best viewed not as a *tree* but rather as a *connected directed acyclic graph* (these are often called "DAGs" or "posets") with bounded fan-in and with node labels selected from a set of K possible labels.

If we had an algorithm to generate all N-node connected DAGs with bounded fan-in (without wastefully regenerating isomorphic DAGs!) that ran in polynomial(N) average time per DAG, *then* we could – by using the same Gray-code trick we just discussed to exhaustively enumerate node-labellings (*assuming* enough kinds of node labels exist so that the average number of valid labelings of a poset is exponentially large) – generate and run all straight-line code algorithms in $O(1)$ average time per algorithm.

Developing such a generator is probably possible using Brinkmann and McKay's ideas [26][124], although I make no claim that either they or I have done so. So this second empire expansion attempt should be regarded as a "plausible future success, but needs further work."

Empire expansion attempt #3: permitting loops. One could consider adding "conditional jump back" statements to straight-line code, which in the directed-graph view would be "backpointing arcs labeled with the jump condition." In this way we could allow a certain amount of "loop structure." (By using the ideas of self-proving self-forcing polynomial-time termination discussed in §12 we would not have to worry about nonterminating algorithms.) This would cause the non-isomorphic graph generation problem to become more difficult but *perhaps* still feasible to do in polynomial time per nonisomorphic graph, but *running* the algorithms corresponding to the graphs would *not* be possible by any obvious update-type method in $O(1)$ time per algorithm; we would apparently be forced to run all algorithms in their entirety. So this final empire expansion attempt must be regarded as a "failure."

16 Human intelligence vis-a-vis our definition

Let the **Human UACI Hypothesis (HUH)** denote the assertion[43] that human intelligence is built in essentially the same way as our UACI construction in §12.

The next four sections will respectively investigate 4 main lines of historical and psychological evidence that support the HUH. Also we shall sketch them below. These include somewhere between *12* and *thousands* of points of agreement in total (depending on which you count; obviously some experiments highly similar to previous ones cannot be considered "indepedent" evidence) and many of those points of agreement yield important insights about the *manner* in which the human UACI is implemented, and not merely the fact that it *exists*.

Although no one piece of this evidence is especially convincing – an alternate phrasing might be "each piece by itself is very weak" – the net bulk is enough for considerable confidence.[44] For example, if each of our pieces of evidence could have come out inconsistent with the HUH with a priori probability$\approx 1/2$, then with ≥ 12 pieces of evidence, the fact that $2^{12} = 4096$ suggests that "confidence$\gtrsim$0.999" is appropriate (at least versus the null hypothesis).

HUH – summary of points of agreement:

```
1. Spearman's positive correlation principle
   (supported by 1000s of experiments).
2. Spearman 1-dimensionality principle
   (IQ, g; supported by 10s of experiments)
   and the observed considerable universality &
   adaptability of human intelligence.
3. Piaget.
  3a culture independence
  3b "Piagetian search"
  3c some innate knowledge, behaviors, and reward-structur
  3d simplicity
  3e critical periods for certain kinds of learning
4. Forgetfulness, and hypotheses of
   utility-tracking & algorithm overwriting.
  4a usedness ==> utility
  4b memory interference (post & prior)
  4c algorithm overwrite
  4d experiments with nonhuman animals
  4e infantile amnesia
  4f rehearsal
  4g chunking
  4h memory alteration via "leading questions"
5. time consumption
  5a exponential roll out (multiplication & Karatsuba,
     Newtonian physics, chemistry, Rubik Cube, others)
  5b power law of improvement with practice (15 examples)
My count: 12-1000s.
```

The first two lines of evidence: Spearman and Piaget. "Human intelligence" seems a priori to be a multidimensional concept. In the face of that intuitive perception, there are two known empirical reasons why psychometricians' 1-dimensional concept of "IQ" can still make sense and be useful for the purpose of measuring it [88][44][45][120][86][182][183]:

that are *not* just a trivial result of re-ordering of the computations and/or commutative laws, and *this* kind of waste would *not* be eliminated by our techniques.

[42] However, it would be an even better success if we could have a more general tree generation algorithm allowing two intermingled types of nodes – with ordered or unordered children – and again preferably with the trees being generated in a "$O(1)$ average change" way with changes "average distance $O(1)$ from the tree root" to permit fast update-type computation of the algorithm outputs corresponding to each tree.

[43] Which is intentionally vaguely phrased...

[44] It is rather like seeing one square inch of the skin of an elephant and trying to prove you have an elephant. Not very convincing – but if you do this for 50 different independent square inches, then you start to get pretty convinced.

1. C.Spearman's positive-correlation and 1-dimensionality principles,
2. J.Piaget's observations [182][183] about the fixed chronological pattern of development of children's intelligence, combined with W.Stern's notion (which once was in wide use) that IQ is the quotient of mental age divided by chronological age.

Now our abstract definition of intelligence in §9 also seems a priori to be multidimensional – *but* our discovery in §12 of how to build a UACI suggests that intelligence really *can* be considered to be 1-dimensional and to some extent *predicts* both Spearman's and Piaget's principles should hold, not only for human, but in fact for any intelligence built in roughly the same way as §12's UACI construction.

So the HUH thus *explains why* both Spearman and Piaget hold, why human intelligence acts the way it does, and shows why, contrary to one's initial impression of §9's definition of "intelligence," it to a considerable extent may be measured 1-dimensionally. (Of course neither our definition nor human intelligence can *fully* be considered "one dimensional" – that is only a crude approximation.)

Sections 17 and 18 shall discuss Spearman and Piaget's principles and human intelligence generally. These two principles are generally regarded as two of the three most important topics relevant to human intelligence. The third topic is the "IQ controversy," which concerns the question of how important sex, race, and geography are for determining your IQ, why that is, and what its implications are. Fortunately, we shall be able to avoid discussing that topic because it is irrelevant to the key issues in the present work.[45]

Forgetfulness. Now consider what is probably the most blindingly obviously embarrassingly poor feature of human intelligence – the fact that we forget things.[46] (How can we possibly be designed that badly?) This is not due to inherent biological limitations that for some reason prevent permanent memories from existing – it is known that certain other "far less intelligent" animals can and do remember important information permanently, e.g. salmon which remember their birthplace stream and return there many years later at the end of their lives to spawn (after having ranged 1000 miles throughout the ocean plus 100s of miles up and down the river).[47]

We shall see in §19 that this puzzle too is explained by the UACI hypothesis.

Time-consumption behavior. Finally, §20 demonstrates that the two most important observed time-consumption behaviors of human intelligence both are compatible with the UACI hypothesis.

17 Spearman's g and human IQ tests

Precis. Starting with Spearman in 1904, psychometricians have developed a large body of theory and experiment about innate human mental abilities and IQ, often labeled "Spearman g theory" [86]. Although the underlying theory is quite simple linear algebra, I claim that no description of it in any form a mathematician would consider pleasant, was ever produced during this century-long span; and some crucial theorems underlying everything were never stated. Therefore we for the first time provide such a description and state those theorems. Everything also rests on two empirical "laws" we call "Spearman's positive-correlation principle" and "Spearman's one-dimensionality principle." It is commonly claimed that, during the hundred-year span, no convincing exception to Spearman's positive-correlation principle was ever found. Therefore we provide two. Spearman's one-dimensionality principle asserts that the high-dimensional ellipsoid describing the distribution of human mental ability is essentially "needlelike." In the hundred year span, it appears that no survey of the actual axis lengths of these ellipsoids, has never been published. Therefore, we provide one. It finds that the longest ellipsoid axis is between 1.42 and 3.92 times longer than the second-longest and average axis lengths, which, although "more one dimensional" than a sphere, is much less-dramatic "one dimensionality" than the impression one would get from much of the g literature – comparable to the "one dimensionality" of a standard construction brick ($15 \times 7 \times 5$). We also show that both Spearman's original "great" paper, and much of the subsequent g literature continuing to the present day, have involved shoddy statistical, methodological, and numerical practices. In particular, a great number of studies (continuing to the present day) devoted to the issue of higher dimensions (i.e. beyond g) of the human intellect, are shown to be almost entirely valueless because the error bars on the entries of the higher eigenvectors all are enormous. (Apparently none of the previous studies bothered to compute those error bars; and although there have been two book-length criticisms of g theory [68][90], most of *our* criticisms are new in the sense that they were not stated in those books.) Finally, we provide a survey of "biological correlates to g" (there have been others, but none simultaneously as extensive and concise as ours).

At that point, the reader will understand both what g is, and how much to trust it; we believe the main conclusions of g theory, in particular Spearman's two principles, probably are largely correct despite the shoddy nature of much of the work backing them up. We then point out that it seems obvious that both Spearman principles are *predicted* by the "Human UACI hypothesis" (HUH) that human intelligence works in essentally the same fashion as

[45] Our Spearman discussion will have independent importance because it exhibits the first examples of negative correlations between two kinds of mental performance (i.e. the first clear exceptions to Spearman's positive correlation principle) and because it re-examines and criticizes the entire Spearman-g area, pointing out some major flaws for the first time.

[46] Humans also exhibit embarrassingly poor performance compared to computers at arithmetic operations, of course. However, that perception may be misleading, because our cerebellums feature circuits that appear to be analog 3×3 matrix multipliers operating at millisecond time scales. (This is from [32] p.99-101, who cites [150] who are backed up by experimental evidence in [63].) If so, then we are only bad at *conscious symbolic* arithmetic (which is not surprising considering its unimportance throughout our evolutionary past) and our underlying hardware performance is non-embarrassing.

[47] There are also rare humans with superb memories. For example ([7] p.25) mathematician A.C.Aitken (1895-1967) once memorized 1000 digits of π. In 1937 he was tested using a passage of prose and list of 25 words. 27 years later Aitken, asked to recall the material, gave all 25 words in the right order, and also almost exactly recalled the prose passage. As a teacher, a single reading of the list of the 35 student names in a new class enabled Aitken never to need to consult it again. Aitken's inability to forget horrible memories such as the Battle of the Somme may have tormented him and he suffered a mental breakdown near the end of his life.

our mathematical construction of a UACI. This is confirmatory evidence for HUH, but we caution the reader that the phrase "seems obvious" in the preceding sentence is merely a heuristic commonsense feeling that cannot be backed up by any rigorous proof.

Spearman's first principle: positive correlations. An amazing claim about human intelligence, which dates [197] to Charles Spearman (1863-1945) in 1904, is this:

1. Define two tests A and B that attempt to measure two innate human mental abilities – it does not appear to matter much what they are.
2. Then, across a large number of humans tested on both A and B, you will find *positive* (or perhaps ≈zero, but[48] never negative) centered correlation for performances on the two tests.

The reason this claim is "surprising" is that one might have imagined that humans who perform better in a test of (say) algebra ability might do so because more of their brains are devoted to algebra, leaving less brain left over for (say) solving crossword puzzles, resulting in a negative correlation. But that rarely or never seems to happen.

Reality check – Two Numerical Examples: Table 17.1 gives two independently gathered 11×11 correlation matrices taken from the literature, with minimum and maximum correlations inside them printed in bold, and (to make the presentation more concise) the first matrix has been multiplied by 1000 and the second by 100. Observe that *all of their entries are positive*:

		1	2	3	4	5	6	7	8	9	10	11
1	Raven	1000	185	226	465	129	143	250	298	435	309	561
2	Gen.Info.	185	1000	566	247	524	566	342	479	329	346	375
3	Arithmetic	226	566	1000	220	463	562	169	513	111	247	267
4	Comprehensn	465	247	220	1000	082	256	283	**036**	372	263	361
5	Vocabulary	129	524	463	082	1000	536	236	367	217	348	240
6	Similarities	143	**566**	562	256	536	1000	128	338	043	199	261
7	Digit-Symbol	250	342	169	283	236	128	1000	233	386	281	342
8	Picture compltn	298	479	513	036	367	338	233	1000	273	367	447
9	Spatial	435	329	111	372	217	043	386	273	1000	504	507
10	Picture Arrgnmt	309	346	247	263	348	199	281	367	504	1000	501
11	Object Assembly	561	375	267	361	240	261	342	447	507	501	1000
	1	100	36	**72**	55	59	59	52	50	45	32	26
	2	36	100	46	47	36	40	23	31	32	**14**	27
	3	72	46	100	48	70	67	49	51	45	32	32
	4	55	47	48	100	47	43	30	41	44	33	28
	5	59	36	70	47	100	58	46	42	39	29	30
	6	59	40	67	43	58	100	52	53	46	40	33
	7	52	23	49	30	46	52	100	48	45	41	26
	8	50	31	51	41	42	53	48	100	43	36	28
	9	45	32	45	44	39	46	45	43	100	58	36
	10	32	14	32	33	29	40	41	36	58	100	25
	11	26	27	32	28	30	33	26	28	36	25	100

Figure 17.1. Two mental-ability correlation matrices. Note that all their entries are *positive*. ▲

The first matrix is on page 422 of Kranzler & Jensen [97] and arose from testing 101 University of California students (52 female, 49 male) on the respectively named subtests. (For a better description of these 11 tests, see the original paper). The second matrix is from page 7 of [45] (data collected by Crawford) and arises from a representative sample of 365 Scots taking the 11 subtests of the (English language) Wechsler Adult Intelligence Scale-Revised [218]:[49]

1. Answer general knowledge & information questions,
2. Digit span (remember & repeat digit sequences forwards & backwards),
3. Vocabulary (describe the meanings of words),
4. Mental Arithmetic,
5. Comprehension (explain facts, actions, and concepts),
6. Similarities (asks about ways different seeming objects or abstract things are similar),
7. Picture completion (notice the missing parts of a number of line drawings),
8. Picture arrangement (organize a series of line drawings to make them form a story),
9. Block design (construct set patterns from cubes with different-color faces),
10. Object assembly (jigsaw-like puzzles), and
11. Digit-symbol association task (enter codes below numbers according to a coding system printed at the top of the page).

[48] In real-world statistical data, *exact* zero correlations do not exist.

[49] For two representative Dutch WAIS samples in 1967 (1100 people) and 1998 (77 people) see p.516 of [220]; as you then will see the Dutch correlation matrices show considerable differences both from our Scottish matrix and from each other. Typically, the Dutch-1967 correlations differ from the corresponding Scottish ones by 0.05 on average (0.065 if RMS difference) but still contain only positive entries in $[0.05, 0.76]$. This gives you some estimate of the "one σ additive error bars" appropriate for each correlation matrix entry, namely about ± 6.5 for our second matrix (as scaled by 100) and about ± 100 for our first matrix (as scaled by 1000).

The reader who wishes to examine more and larger correlation matrices may find some on page 190 of [70] (22×22 matrix from 981 Swedish 6th graders; min and max correlations 0.08 and 0.8), page 268 of [1] (13×13 matrix from 899 people; min and max correlations 0.09 and 0.57; an additional table on their page 266 gives two additional rows which one can adjoin to the matrix to get additional correlations to auditory pitch discrimination and color discrimination tests; these additional values range from +0.03 to +0.49), page 164 of [157] (7×7 matrix from 284 people; min and max correlations 0.134 and 0.569), pages 110-112 of [205] (57×57 matrix from 240 people; min and max correlations −0.22 and +0.8; note this matrix *does* contain some negative correlations, although they are rare – that will be discussed below), [206], [203], [214] and see also [108][28].

Spearman's second principle: One-dimensionality & g. Suppose we subject many humans each to N different mental tests for some fixed value of N (such as $N = 11$ in the examples above), and each test result is a real number. Suppose the net distribution of each real behaves (thanks to a *standardization* to make its mean be 0 and its variance 1) like a Gaussian "standard normal distribution"[50] [51] and indeed suppose the set of N-dimensional score *vectors* is distributed approximately according to an N-dimensional Gaussian density centered at $\vec{0}$:

$$\rho(\vec{x}) \;=\; \pi^{-N/2}\sqrt{|\det M|}\,\exp\left(-\vec{x}^T M \vec{x}\right) \tag{2}$$

for some *positive-definite* symmetric matrix M describing the shape of that Gaussian. (Specifically, M is half the *inverse* of the correlation matrices given above, so the preceding claim was that every entry of M^{-1} is nonnegative.)

Now Spearman's *further* claim is that the characteristic N-dimensional ellipsoid "shape" of the Gaussian is *needlelike*, i.e. approximately *one-dimensional*. That is, M^{-1} is very nearly a *rank-1* matrix, i.e. has a unique maximally-positive eigenvalue corresponding to its Perron-Frobenius unique all-positive eigenvector, and with all the other eigenvalues *much smaller* in norm.

A closer look of course reveals that the distribution is not exactly 1-dimensional, and the eigenvectors sorted in order of decreasing eigenvalue-norm give orthogonal bases which span increasingly-dimensional subspaces giving increasingly better descriptions of the distribution of human intellect. But Spearman's point is that, to a good first approximation, this distribution is described by a *single* dimension, later dubbed "Spearman's g," saying how far you are along a *single* direction in N-space, namely the Perron-Frobenius eigenvector.

Numerical reality check #2: Consider the two 11×11 matrices above. The 11 eigenvalues of the first matrix are

$$4.27, 1.77, 0.98, 0.86, 0.71, 0.52, 0.48, 0.46, 0.34, 0.31, 0.30 \tag{3}$$

and the 11 eigenvalues of the second matrix (the 1σ error bars on these appear to be about ± 0.09) are

$$5.28, 1.08, 0.91, 0.75, 0.60, 0.55, 0.48, 0.40, 0.38, 0.33, 0.22. \tag{4}$$

(The reader may check that these 11 numbers have sum=11 as they should.) The principal axis lengths of the ellipsoids are proportional to the *square roots* of the eigenvalues. Thus for the first matrix, the longest axis is $\sqrt{4.27/1.77} = 1.55$ times longer than the second-longest axis and $\sqrt{4.27/1} = 2.07$ times longer than the root-mean-square axis. For the second matrix, the longest axis is $\sqrt{5.28/1.08} = 2.21$ times longer than the second-longest axis and $\sqrt{5.28/1} = 2.30$ times longer than the RMS axis.

[50] Incidentally, it is false – no matter how many psychologists claim it is true – that mental tests necessarily yield a Gaussian-like score distribution. That might be true for composite tests with a large number of different kinds of problems, and definitely is true for tests "validated" to obey it, but it is easy to construct tests in which the result distribution is bimodal because test takers either understand how to do that kind of problem, or are helpless; and it also has been found that the raw-score distributions on certain kinds of tests can be highly skewed. For example the annual American Math'l Society competitive "Putnam exam" usually has a median score of *zero* out of a possible 120. Also, the tails of the distribution of IQs do not fall as fast as the tails of a true normal distribution with mean 100 and standard deviation 15 fall, otherwise IQs above 200 would not exist, but the purveyors of IQ tests claim they exist. For example Cox & Terman [37] in 1926 claimed (based on accomplishments up to age 17) that poet and playwright J.W. von Goethe had IQ 210, followed by religious writer Emanuel Swedenborg with IQ 205, psychometric-test-pioneer Francis Galton, Catholic Cardinal Thomas Wolsey and economist and political philosopher John Stuart Mill each with IQ 200. Those topped the comparatively piddling IQs of mere scientists such as Isaac Newton (190), Galileo (185), Michael Faraday (170), and Charles Darwin (165) and the American President Abraham Lincoln (150). Physics Nobelist Richard Feynman was tested as a schoolboy and found to have IQ 124 [64], despite the fact that Feynman later as an undergraduate won the extremely difficult nationwide Putnam (competitive math) exam and still later co-won the Nobel prize as probably one of the top 5 physicists ever. DNA-structure co-discoverer and Nobelist James Watson stated in a 2005 public lecture that his IQ was 120. More recently, it has been claimed that an almost-unknown recluse and child prodigy named William James Sidis (1898-1944) had an IQ between 250 and 300, while the contemporary advice-column writer Marilyn Vos Savant (1946-) was listed for 5 years in the Guinness Book of World Records under "highest IQ" for both childhood and adult scores with 228. Sidis is discussed next footnote.

[51] Supposedly Sidis read the *New York Times* at 18 months, read books in Latin [self-taught] at age 4, calculated mentally the day of any date in history at age 6, was fluent in 8 languages by age 8, enrolled at Harvard at age 11 (passed entrance exam at age 9 but not allowed in until 11 – youngest ever admitted?) where he also at age 11 delivered a much-publicized two-hour lecture on four-dimensional bodies which included Euler's face-count formula $F_0 + F_2 = F_1 + F_3$ and the existence of the 120-cell and 600-cell regular 4-polytopes. (None of these 3 results were new, but Sidis quite plausibly did not know that.) His A.B. was awarded cum laude. Sidis then got a job teaching math at Rice (Houston) at age 17, but only lasted 1 year, He remembered and quoted facts from books including the page number (superb memory), enjoyed doing crossword puzzles entirely in his head, claimed to be able to learn a language in one day (knew about 40)... considerable although not entirely accurate information about Sidis is in the book [215]. Sidis's greatest work was his book *The Animate and the Inanimate* written in 1920, published in 1926, and now available electronically. Although this book is enjoyable to read and in terms of choice of topics was quite ahead of its time, its fundamental proposals are simply wrong: the main false idea was that life is an example of entropy decrease contradicting the second law of thermodynamics; and Sidis falsely conjectured the same about nebulae and stars – perhaps all three contained some magical substance that made that possible – and thus that in *net* entropy in the universe was not increasing and thus that the second law of thermodynamics was not in contradiction with the reversibility of fundamental physical laws. It has been claimed by people I would call "Sidis cultists" that in this book Sidis proposed the reality of black holes and of "dark matter" well ahead of their acceptance by everybody else, but I deny both those claims. The book is almost entirely free of mathematics and Sidis exhibits in it no evidence of understanding of either General Relativity or Quantum Mechanics (which is not surprising since both were in their infancy in 1920). Sidis got a succession of menial jobs for the rest of his life, writing, after hours, a succession of both published and never-published books about socio-economic, historical, and hobby topics – often under pseudonyms such as "Barry Mulligan." He appears never to have done anything of lasting intellectual value and seems to have been tormented by his childhood's repercussions.

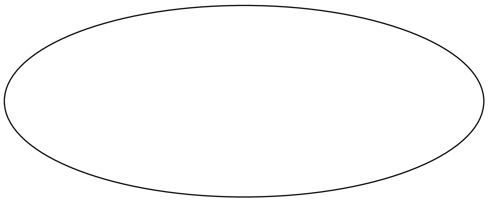

Figure 17.2. A 2.5×1 ellipse. ▲

The top few eigenvalues in a compendium of other real-world correlation matrices are tabulated by Linn [108], and for some more see [28][157][214]; I have retabulated these in table 17.3.

ultimate source	#people	#tests	$\sqrt{\lambda_1/\lambda_2}$	$\sqrt{\lambda_1}$
Harman 1960	145	24	1.97	2.85
LL & TG Thurstone	437	21	1.76	2.71
in 1941	710	60	1.74	3.92
Wiggins	250	26	2.29	2.91
Aircrew classifn tests 1944	5158	21	1.42	2.30
Guilford 1955	364	52	1.56	3.13
Schutz 1958 (corrected)		9	1.93	2.11
Rabbitt 1988	284	7	1.78	1.83

Figure 17.3. Principal axis ratios (square roots of ratio of the top two eigenvalues of correlation matrix, and of top eigenvalue divided by RMS eigenvalue) for various sets of mental tests on various sets of people, from [108][28][157]. (Our two 11-test numerical examples are external to this table.) ▲

Verdict: The common acceptance, among psychologists investigating intelligence, of Spearman's 1-dimensionality principle is not justified because – as anybody with sewing experience would agree – an axis ratio of 1.55 to 2.30 (from our two numerical examples; or 1.42 to 3.92 in table 17.3) does not a "needlelike" ellipsoid make. However it *is* justified in the sense that these axis ratios all indeed are *much* larger than the ratios that usually arise from randomly generated artificial IQ test score uncorrelated "datasets" which, of course, usually feature both positive and negative correlations (typically zero) and which lead to comparatively *spherical* Gaussians, with the largest two axes probably within a factor of $1 \pm O(N^{-1/2})$ in length, and the largest axis $\approx 4/\pi \approx 1.27$ times longer than the average axis length and $\approx \sqrt{3/2} \approx 1.22$ times longer than the RMS axis length.

Six well known theorems from linear algebra fundamental to our subject:

1. A real symmetric matrix has all eigenvalues and eigenvectors real, and all its eigenvectors (wlog) orthogonal.

2. The $n \times n$ matrix of the inner products of n linearly independent vectors (e.g. any correlation matrix) is automatically real-symmetric (with unit diagonal if the vectors had unit norms) and *positive definite*, e.g. all eigenvalues are positive.

3. The sum of the eigenvalues of a matrix equals its trace, and their product equals its determinant.

4. The Perron-Frobenius theorem states that a square matrix M with all entries positive real, has a unique (non-multiple) eigenvector consisting entirely of positive reals, and it automatically corresponds to M's largest-norm eigenvalue, which automatically is a positive real.

5. The best rank-k approximation (in the Frobenius norm, i.e. sum of squared entries) to a symmetric $n \times n$ matrix M is got by using its top k largest-norm eigenvalues ($1 \le k \le n$) and their corresponding eigenvectors *only*, discarding (i.e. zeroing) the rest.

6. A symmetric random $N \times N$ matrix $A = B^T B$ where B is $N \times N$ with i.i.d. standard random normal variates as entries, has square-roots-of-eigenvalues (i.e. "axis lengths" for the ellipsoid; also these are the singular values of B) distributed according to a density function whose plot is exactly the upper-right quadrant of an origin-centered circle chosen so that the mean-square eigenvalue is N. (This all is from [178] theorem 2 and he cites [219]). Consequently the ratio of the largest two axis lengths will generically be $1 + O(N^{-1/2})$ and the ratio of the largest to the mean axis length will approach $4/\pi$ and the ratio of the largest to the RMS axis length will approach $\sqrt{3/2}$ all with probability$\to 1$ as $N \to \infty$.

What is g as a mental test? By simply writing down the Perron-Frobenius eigenvector, we get, effectively, a definition of g as a certain positively-weighted linear combination of the N subtest scores. This weighting in fact defines a supertest whose purpose is precisely to measure g.

It also yields a mostly-automatic procedure to devise "a good IQ test":

1. Create a very large number N of test problems,
2. try them out on a large diverse set of humans,
3. compute the Perron-Frobenius eigenvector,
4. discard the test problems with small variances and/or small resulting weights,
5. and then redo the trials and eigenvector computation to recompute the weights of the undiscarded problems.

The result will be a not-too-long IQ test (problem set with Perron-Frobenius scoring weights) which measures g excellently.

Numerical Examples. In the two 11×11 matrices we have been considering, the top eigenvector ($\lambda = 4.27$) of the first matrix is approximately

$$(28, 36, 31, 24, 30, 29, 25, 32, 29, 31, 35) \tag{5}$$

which would give the optimum weightings of the 11 respective parts of Kranzler & Jensen's test for the purpose of measuring g (assuming all 11 parts had been pre-scaled to have equal raw score variances). In other words the most g-discriminating segment of their test was the second subtest ("general information") and the least g-discriminating part was the fourth subtest ("comprehension"). The top eigenvector ($\lambda = 5.28$) of the second matrix is (with multiplicative error bars on each entry which appear to be about $100 \pm 10\%$)

$$(35, 24, 36, 30, 33, 35, 30, 30, 30, 25, 21), \tag{6}$$

giving *its* optimum g-weightings, indicating that on the WAIS-R test, the most g-discriminatory subtest is the third (vocabulary) with the first and sixth subtests (general knowledge & info; similarities) nearly as high; and the least g-discriminatory subtest is the last (digit-symbol association task).

Note that these two vectors are not in very good agreement on their fine details – for example the Kranzler-Jensen data says Object Assembly is second-top in g-discriminatory power, whereas Crawford's WAIS-R data ranks it third-lowest (and there is a similarly big swing for "similarities")! This moderate-to-poor level of agreement is typical of what you get when trying to deduce "what g is" from different datasets.

The *second-top* eigenvectors are in even poorer agreement! They are respectively

$$(-18, -41, -25, -18, -24, -05, +23, +08, +41, +63, +14) \tag{7}$$

$$(+36, -26, -36, +29, -33, -41, +19, -16, +39, +17, +26) \tag{8}$$

for the Kranzler-Jensen ($\lambda_2 = 1.08$) and WAIS-R ($\lambda_2 = 1.77$) data. The great lack of agreement here is not surprising considering that the error bars on the entries of the latter, second eigenvector (which I estimated by the Monte Carlo approach of adding 1σ random-normal perturbations to the correlation matrix entries and re-evaluating the eigenvector) are *huge* – of order comparable or larger than the entries themselves – indicating the extreme difficulty of attempts to model/deduce the dimensions beyond the first (g) of the distribution of human intellect.[52] Obviously, vastly larger (at least 100×, and quite possibly 1000 or $10^4\times$ larger) quantities of data would be required to get rid of this immense "noise." Since such data is very expensive to collect and there are many mysterious confounding effects (see critical discussion below) and in fact no study I am aware of has either collected more than *one* order of magnitude more data or even *considered* some of the more obvious confounding effects, I suggest that all efforts to do so are doomed and a pointless waste of time.[53]

What is g biologically?

Struck by this one-dimensionality, Spearman later advanced the notion that g must be "something of the nature of an 'energy' or 'power' which serves in common the whole cortex (or possibly even the whole central nervous system)." Later investigators attempted to find biophysical or biochemical quantities that correlate with g, see table 17.4.

Figure 17.4. Alleged genetic, biophysical, and biochemical correlates to g that have been claimed in the literature during 1940-2005, sorted in roughly decreasing order of importance. I do not necessarily endorse any of these results,[54] many of which I collected from secondary sources [44][86][120][135][141][213], although others were collected from primary sources, mainly papers in the years> 1990 of the Elsevier journal *Intelligence* and to a lesser extent *Personality and Individual Differences* and the Minnesota Study of Twins Reared Apart (MISTRA) [23][98]. Nor am I necessarily making any implied claim about which of two correlated effects causes the other (or whether they are both consequences or artifacts of some other cause).[55] Some of these correlates are heavily supported by tremendous amounts of data (the positive correlation of head size measurements to IQ has at least "ten nines" of confidence) while others are supported by only small amounts of data (PTC tasting ability: study of only 122 undergraduates [62]). The claim myopia is positively correlated to IQ contradicts Spearman's claim[56] that all sensory and mental abilities are positively correlated, but if Spearman's myopics all were tested in corrective eyewear then this is not necessarily a contradiction. If PTC-tasting-ability really is *negatively* correlated to IQ that refutes Spearman's sensory↔IQ universal positive correlation hypothesis. Throughout the present work we shall therefore regard "Spearman's positive correlation hypothesis" as only pertaining to *mental* and *not* (as Spearman originally proposed it) also to sensory tasks – although Spearman's original conception perhaps still is valid if altered to embrace only some particular large subset of sensory tasks. The self-assessed skin color refers to a study by Lynn & Rowe [117] correlating self-assessed skin color on a 5-point scale with a 10-word vocabulary test, and had $p < 0.01$ significance level. One balanced survey of racial correlates[57] to IQ is ch.5 of [120]; see also [118][168][86][88][68]. Numerous studies have found conflicting α-rhythm frequency↔IQ correlations ranging from 0 to about +0.60. There are many ways that "brain wave" electrical signals may or may not be correlated to IQ, and the area is difficult. One popular idea is AEP (averaged evoked potential) – the subject is stimulated by e.g, audible clicks, and EEGs are collected after each click and averaged over a large number of clicks (which averages out noise). The result is a multipeaked curve. High IQ seems to correlate with more peaks, shorter latencies,

[52] I am pointing this put because I unfortunately have failed to find similar error-bar estimates in the literature, confirming yet again my impression of its overall shoddiness. Although it is a well known consequence of Gershgorin's circle theorem that finding the eigen*values* of a symmetric matrix is a numerically well-conditioned task [65], finding eigen*vectors* corresponding to nearby eigenvalues can be extremely ill-conditioned. For example, for the identity matrix, this condition number is infinite. Consequently it can be extremely difficult to determine (accurately and with confidence) the eigendecomposition of correlation matrices arising from statistical data, although eigenvectors arising from eigenvalues such as g that are *well separated* from the others, are much easier to deduce.

[53] Vast numbers of studies – both books and papers – continue to be made up to the present day claiming to draw conclusions about this, despite not employing more data than in our numerical examples above. I believe none of them, and find it no surprise whatever that these studies usually continue to find contradicting conclusions.

[54] Many early "results" in this area are known to be bunk; see [68].

[55] Let me repeat some standard warnings. If true, these correlations are mere statistical correlations, valid over samples of 1000s of people, but which can easily fail in any individual cases. For example famous satarist Anatole France (1844-1924; Nobel Prize for Literature 1921) had an exceptionally small 1017g brain, while the almost-as-famous Russian writer Ivan Turgenev (1818-1883) had an exceptionally heavy one at 2020g, regardless of the claimed positive correlation of brain weight with IQ. (Typical brain weights are 400g at birth, 850g at 11 months, 1.1kg at age 3, and 1450g at adulthood [95].) Also, although one might imagine from our table that the ultimate intellect would stand 3 meters tall and have a 30kg brain with pH 14 and zero (or negative!) glucose uptake rate, in fact, any human meeting those specifications would have IQ=0 since he'd be dead. Finally, the function $F(\vec{x}) = 2\prod_{n=1}^{20} x_n - \sum_{n=1}^{20} x_n$, despite having negative correlation with every x_n on the boolean cube $\vec{x} \in \{0,1\}^{20}$, has unique global maximum at the point $\vec{x} = (1, 1, 1, \ldots, 1)$. Such nonlinearities would in fact appear necessary to explain the PTC-tasting gene's empirical negative correlation with *both* IQ and cretinism [62][179].

[56] Spearman [197]: "the common and essential element in the Intelligences wholly coincides with the common and essential element in the Sensory Functions... there really exists a something that we may provisionally term 'General Sensory Discrimination' and similarly a 'General Intelligence,' and further that the functional correspondence between these two is not appreciably less than absolute."

[57] Incidentally, let me recommend – as an excellent laboratory for studying racial-genetic versus environmental effects on average IQ – studying the IQ of South versus North Koreans.

smaller amplitude, and greater habituation (that is, decrease of the effect over time in a long sequence of clicks) according to Jensen [86] p.153-5. Despite conflicting studies, Deary & Caryl in [213] summarize that "a variety of measures of EEG and averaged evoked responses correlate with IQ in adults as well as children and in samples of above-average IQ as well as those including retarded subjects." But on the other hand Mackintosh ([120] ch.7) in a later survey concludes that that entire area is a "mess" containing numerous highly contradictory "results" of various studies, leading to the conclusion that most claimed correlations between EEG signals or attempted measurements of nerve conduction velocity (NCV) to IQ, cannot be trusted. For a still-more-recent claim that brain NCV is positively correlated to IQ, see [161], but the correlation appears to be low, $\approx +0.08$. Until recently attempts to devise IQ tests for children below 4 years of age showed little correlations with IQ later in life, but recently 7-month-old's "habituation rates" and "novelty preferences" have been shown to be have correlation about +0.4 with both IQ, achievement test scores in reading and math, and language proficiency scores 4-18 years later [33][165][184]. The "Flynn effect" is the observation [136], mainly documented by James R. Flynn in numerous analyses of IQ tests over time in many countries, that average IQ scores have historically risen by about 10 IQ points per 30 years for the last 100 years as is revealed by the need to keep "renormalizing" the tests [136]. The Flynn effect is very hard for proponents of racial and gene-based IQ to explain. It is claimed that males [46] and left-handers ([36] p.175-177; [17][142]) exhibit greater IQ *variance* than females and right-handed people (and also there are 10-35% more left-handed males than left-handed females, but this sex link appears not to be enough to explain either greater variance) [125]. Great future progress is expected by finding correlates of IQ inside human DNA. ▲

quantity allegedly correlated to your IQ	centered correl. coeff.
IQ of your identical twin (reared together)	+0.87
IQ of your identical twin (reared apart)	+0.7
IQ of your fraternal twin (correl. decreases with age)	+0.6
Your total years of education	+0.55
brain intracellular pH in boys	+0.52
Your IQ measured at age 11 versus age 77	+0.5
Your school grades (correls of 0.4 to 0.73 depending on subject)	+0.5
glucose uptake rates in 32 brain regions (PET measurement after radioactive glucose injections)	−0.12 to −0.92
IQ of your sibling (reared together)	+0.47
parental estimate of child IQ	+0.44
IQ of your spouse	above +0.40
IQ of your biological parent (who rears you)	+0.42
Body symmetry assessed by 10 measurements	+0.39
MRI direct measurements of brain volume in vivo	+0.34
"Inspection time" required to decide which of two vertical line segments is shorter	−0.30
Self-estimated IQ (11 studies)	+0.19 to +0.49
IQ of your biological parent (but reared apart)	+0.22
α rhythm EEG frequency (7.5 to 12.5 Hz in adults) in 8-year-olds	+0.50
α rhythm EEG frequency in adults	no correl found
Parental socio-economic status	+0.33
Mother consumed more than 1.5oz alcohol daily during pregnancy	−0.32
IQ of unrelated children (reared together)	+0.3
lung capacity (2 studies)	+0.23 to +0.29
High blood lead levels	−0.25
IQ of your sibling (reared apart)	+0.24
external head size measurements (4 sources representing 21 studies of humans)	+0.12 to +0.62, typ= +0.2
Visual color discrimination ability [1]	+0.31
You took the same IQ test twice over last 2 days (depends on test type)	+0.03 to +0.5
8-choice reaction-time task (push the lit-up one of 8 buttons as fast as can)	−0.23
Parental income	+0.22
myopia (in 3 large and many small studies)	+0.20 to +0.25
Auditory pitch discrimination ability [1]	+0.21
facial features – observers of your photo judging "how intelligent you look"	+0.20
PTC (phenylthiocarbamide) tasting ability (genetically determined) [62]	−0.2
age of menarche	−0.20
Your adult height (in 7 studies)	+0.12 to +0.29
IQ of your adoptive (non-genetic) parent (decreases with age)	+0.19
Parental income (for adoptive parents)	+0.18
self-assessed skin color ("very dark" increasing to "very light") among 442 US blacks	+0.17
head size measurements corrected for age and sex (8 studies)	+0.08 to +0.22

quantity allegedly correlated to your IQ	centered correl. coeff.
Deaf	−0.16
IQ of your cousin	+0.15
serum urate level in blood	+0.10
birth weight	+0.05 to +0.17
basal metabolic rate: conflict – some studies +0.6 to +0.7, others no correl	?
brain nerve conduction velocity	+0.08 ?
Boys weight (in given age group)	+0.051
Girls weight (in given age group)	+0.035
asthma and other allergies	+
"general health"	+
likelihood of dying in automobile accident	−
wealth and income	+
inbreeding (have genetically related parents)	−
Participation in Maharishi International University curriculum & Transcendental Meditation [38]	+
mean high winter temperature where you live	−
open-ended smell identification ability [43]	−
longetivity	+
Schizophrenia and depression risks	−
Obsessive-compulsive disorder risk	+
Incarcerated in prison	−
amount of religious belief	−
log(GDP/capita) in your country [50]	+
blood groups	no correl found
vitamin supplementation [207]	no correl found
Individual variations in how much you REM-sleep [180]	no correl found

Criticism of "Spearman's great discoveries" and his field generally. My above sketch of g-factor theory has, unfortunately, been far more concise and clear than the usual literature sources, and the historical development of the area was not nearly as nice as that. It is depressingly easy to criticize that field from top to bottom.

Spearman's 93-page 1904 paper [197] (which Jensen calls "one of the 3 or 4 most important papers in the history of mental testing") in fact was quite shoddy both mathematically, statistically, and methodologically, and to add injury to insult, R.B.Fancher [56] redid Spearman's calculations and found about 50 erroneous results apparently due to wrong arithmetic and with the signs of the errors showing an amazing fortunate tendency to "improve" the validity of his conclusions (Spearman's uncorrected numbers were then reprinted by Jensen [86] p.24). Specifically, Spearman seemed unaware of much of linear algebra, for example *never* mentioning eigenvalues and eigenvectors in his paper. Instead, he noted (essentially) that any 2×2 subdeterminant of a rank-1 matrix is zero, and proposed to use this as a test for rank-1. The fact that all the 2×2 subdeterminants *were* small for his matrix proved the existence of Spearman's g. However

1. This is a stupidly inefficient method since computing all eigenvalues and eigenvectors by standard numerical methods takes $O(N^3)$ steps for an $N \times N$ matrix while giving you much more useful information than computing all the 2×2 subdeterminants, (of which there are order N^4) so it takes longer to get less useful stuff;
2. Spearman and subsequent writers have kept calling these not "2×2 subdeterminants" but rather "tetrad differences" presumably due either to ignorance of linear algebra or desire to keep the reader confused, both pervasive throughout the psychometric field;
3. Even working by hand via Jacobi's method, finding all the eigenvalues and a few eigenvectors would have been easy for Spearman compared to the work he had to do anyway, but he never did it;
4. Spearman's "great" discovery of universally positive correlations between various mental and sensory test scores was actually "obviously predictable" in many cases such as, for example: students with sensory deficiencies in eyesight or hearing might be *expected* to do worse in school subjects such as Classics or French *because* of their poor senses (and *not* because students due to having a lot of some mysterious magical "energy" g had both good senses and good innate mental talents). Incredibly, this simple possibility was not even *considered* by Spearman!
5. For another example, it is commonly claimed (e.g. in manuals advising chessplayers) that your mental performance is decreased soon after eating, perhaps because of increased blood flow to the digestive system and lessened flow to the brain. If that is so, then in any IQ testing study where some test takers ate before testing and others did not, we would generate spurious "positive correlations" which would have vanished if everybody ate at the same time. To my knowledge, in the 102 years since Spearman, *not one* study has ever even considered this effect![58]

[58] I do not know to what extent the menstrual or eating effects are responsible for Spearman's positive correlation principle. (E.g. if they are 100% responsible for it – which I doubt – then that principle may be considered dead.) We can tell immediately from the claimed size 3-9 IQ points of the menstrual effect, and the postulation that the eating effect is at least as large, that these effects *are* large enough to be of the same order of magnitude, i.e. they cannot be ignored as all prevous workers did.

6. It has been claimed that tests reproducibly show women have increased mental functioning, equivalent to a gain of 3-9 IQ points, during the "luteal" phase of their menstrual cycles [93]. (Also, men's scores on certain kinds of mental tests seem to vary with their testosterone levels, which in turn vary with both the time of day and season of the year.)[59] In any tests in which some women took the whole battery at one phase and others at another, this would result in spurious positive correlations for the same reason (which could in principle be eliminated by randomizing subtest order and scheduling subtests spaced out over 28 days). Again, to my knowledge, in the 102 years since Spearman, *not one* study has ever even considered this effect!
7. For another example, students might be expected to do well in *both* or *neither* of French and Algebra in school not because of correlated mental innate ability levels in the two subjects, but simply because of a good or poor school attendance record (which would be expected to be *correlated* over all subjects at that school since a student who did or did not attend that day would attend or miss *both* classes). This "attendance hypothesis" might have been testable by examining attendence records – but was never even *considered* by Spearman!
8. Also, Spearman obtained student grades in all subjects from the *same* schoolmaster, allowing devil's advocates to hypothesize all grades were correlated simply because the schoolmaster *liked* or disliked that student, so that *really* Spearman's discovery was not about the innate nature of human intelligence, but rather about the innate nature of that schoolmaster.
9. This kind of accusation of bias can be avoided to a considerable degree by use of "blinding" techniques, but neither Spearman, nor most workers in his field over the next 100 years, used those techniques, despite *tremendous* motivation to do so.
10. Hence it really remained entirely conceiveable that all of Spearman's "great discoveries" in this study were in fact merely "an illusion."
11. Also, suppose students across the USA went to either "good" or "bad" schools (where good/bad schools tend to have good/bad teachers in all subjects). In that case we would expect, in any study of nationwide test scores, to find a positive correlation among test scores in different subjects even *entirely in the absence* of any such positive correlation in innate mental abilities inside each student's mind! In that case any such study would not be revealing the "innate positively-correlated nature of human intelligence" at all, but rather "the innately positively correlated nature of school district disfunction."
12. Since all this reveals that it can be quite difficult to root out spurious correlations, and since most everybody in the literature uniformly claims that Spearman's paper was "great" without pointing out these huge flaws, one may be forgiven for a certain amount of skepticism that the self-admitted less-great researchers in this area really can be trusted to know what they are doing.
13. And indeed it is clear [68][90][177] that many of the most important and dominant early researchers in the IQ field were frauds who simply invented their data (Cyril Burt [74]), or racists and/or incompetents (Lewis Terman, Robert Yerkes).[60] This does not prove they were wrong (frauds, racists, and incompetents can be right!), but does suffice to generate considerable mistrust.

[59] Unfortunately, the study that "showed" the menstrual effect [93] was based on testing a total of 17 males and 23 females at one particular time for each testee. In other words, their sample was so small that most or all "conclusions" of this study cannot be taken seriously and it was completely ludicrous for it to ever have been published. However, since it was published by the top journal in the area, and it is all the data on the matter that we have to go on, let us for the moment, for the purpose of playing devil's advocate, accept it.

[60] Here is one of the questions added by Terman to Binet's IQ test to create the "Stanford-Binet test" ([68] p.176)

> An Indian who had come to town for the first time in his life saw a white man riding along the street. As the white man rode by, the Indian said – "The white man is lazy; he walks sitting down." What was the white man riding on? [Correct response: "bicycle." Incorrect: tricycle, unicycle, horse, wheelchair, rickshaw, automobile, motorcycle.]

Obviously, this falls rather short of the claimed goal of the test to measure innate intelligence. And then Terman recommended all sorts of uses for his test well beyond where Binet had been willing to go, such as to exclude low-IQ people from various professions (IQ 75 is "an unsafe risk in a motorman or conductor" – how did Terman know?). The US Supreme Court in the 1971 *Duke Power Co.* case, unanimously ruled unlawful the company's requirement of a high IQ score for promotion of an employee to "coal handler" and demanded that any test given relate directly to the skills needed for the job. However, this decision has had near zero impact on such practices US-wide: in one amusing more recent case. Robert Jordan, a 49-year-old college graduate, in 1996 took an exam to join the New London (Connecticut) police. He scored 33 points, equivalent to IQ=125. He was rejected as having *too high* IQ – New London police only accepted candidates who scored 20 (corresponding to IQ=100) to 27, on the theory that too-high scorers would get bored with police work and leave soon after their costly training. Jordan launched a Federal lawsuit but lost because "the same standards were applied to everyone who took the test." ([208]; the decision was handed down by the second US circuit court of Appeals on 23 August 2000). In 1922 an IQ test was used to justify removing children from the care of their mother because she had a "mind equivalent to a thirteen year old." An IQ test devised by a committee headed by Robert M. Yerkes (head of the American Psychological Association) and also including Terman as a contributing member, became the first given to vast segments of the population – all World War I US army recruits, by 1919 nearly 2 million of them. It included such "unbiased" multiple choice questions as "Crisco is a: (1) patent medicine, (2) disinfectant, (3) toothpaste, (4) food product?" and "Christy Mathewson is famous as a: (1) writer, (2) artist, (3) baseball player, (4) comedian?" which quite likely both would have stumped Albert Einstein (and definitely Isaac Newton). This test was used to conclude that various US immigrant racial groups suffered from low IQ, and that subsequently was used as the basis for excluding them via quota in the 1924 immigration restriction act. Yerkes also found that the "average mental age" of the army draftees taking this test was 13 years (where IQ is mental age divided by chronological age) leading to the amazing conclusion that the average is much less than the average. (A later study of two variants of the Yerkes test found even lower average adult mental ages of 12.2 and 10.8 years.) This by itself should have been enough to realize the test was extremely poorly designed and/or administered, but Yerkes instead concluded "We have never heretofore supposed that [13 years] was the average of the country or anywhere near it... feeble-mindedness... is of much greater frequency of occurrence than had been originally supposed." Personally, my reaction would have been to advise Yerkes to look in the mirror, but regardless of Yerkes' own personal feeblemindedness, this experience illustrates the *immense* discrepancies between different "nationwide IQ scores" yielded by different IQ tests – engendering very great doubt about all conclusions of that nature. This as usual is yet another massive indictment of the entire psychometrician community. Yerkes' 13-year-old=average finding also by transitivity leads to the amazing conclusion that the average parent should have their children forcibly removed from their custody. I would also argue that the "great" founders of psychiatry such as Sigmund Freud were also incompetent frauds, e.g. see [172], but that is another battle.

14. The Educational Testing Service – designers of the SAT, [61] the USA's most important nationwide standardized test – *intentionally discard all* test questions not shown to exhibit both high correlation with the total score on the rest of the SAT, and high variance. (They call this "validation" and they have used this design procedure ever since the SAT's inception [146].) This *forces* Spearman's g to appear to exist on the SAT, regardless of whether it actually does or not!
15. As Gould [68] pointed out, after the appearance of the Stanford-Binet IQ test, all or most other highly used IQ tests and subtests were first "validated" by their designers by studies intended to prove their positive correlation with Stanford-Binet, and with revision until good validation occurred. This *forced* Spearman's g to appear to exist from then on, regardless of whether it actually did exist! In particular, in both our two "confirmatory numerical examples" above, I believe that *all* of each's 11 subtests were known before beginning the experiment to be highly correlated with g, in both cases causing the "verification" of both Spearman principles to be a foregone conclusion!

The entire field then continued to exhibit rather poor apparent understanding of linear algebra up to the present day (the very fact they keep using the name "factor analysis" as opposed to "eigendecomposition" proves that; and the fact that in the entire IQ literature I have *never* seen the Perron-Frobenius theorem, or the rank-k best approximation theorem – both of which are fundamental to their area – even *mentioned*). I urge anybody in this field to learn linear algebra from a decent book [65][77]; I also urge either the abandonment of, or much-better justification of, all forms of "factor analysis" that are different than merely computing the eigendecomposition.

So in view of all this, it would appear that there is still room for a skeptic perhaps to dispute the existence of (and certainly to dispute the magnitude and importance of) Spearman's g, *despite* the commonly heard claim that this is the oldest, and most important, and most well confirmed, finding in all of psychometrics!

And indeed, psychometricians who do believe that intelligence is multidimensional continue to exist. The one who seemed to go the furthest in that direction was J.P.Guilford [69] who proposed a 120- or 150- or perhaps even 180-dimensional model of human intellect ($150 = 5 \times 6 \times 5$ arose from a scheme Guilford had involving a 3-way classification of mental abilities into 5, 6, and 5 categories each way).[62] Guilford spent many years systematically trying to find tests of his "abilities" which were orthogonal to Spearman's g and/or to each other, and claimed to have succeeded in great profusion.

So how did Jensen, the top bulldog for Spearman's g, respond [86]? He simply dismissed Guilford with the claim that he had been refuted by Alliger [3], giving a quick oversimplified sketch of what Alliger did. But when we actually examine Alliger's paper, what do we find? Nothing convincing. Alliger just says that Guilford foolishly tested mostly white male college students with a military connection, a restricted sample. That restriction artificially reduced a lot of Guilford's correlations to effectively zero instead of positive, conjectures Alliger. Good point. But then Alliger says that by "correcting" Guilford's data by multiplying it by a "correcting factor" to enlarge every correlation, he gets a lot of negative and positive correlations – but mostly positive. My response is: that is garbage. Thanks to Alliger, I'm willing to concede many of Guilford's "zero correlation" findings may have been nearly meaningless, but if so, then Alliger's findings got by "correction" definitely are meaningless and in no way refute the Guilford's hypothesis that there are a great many important intellectual quantities orthogonal to g and each other – if Guilford were 100% correct that the disputed correlations were zero, then Alliger's "corrections" would still have come out basically the same as they did![63] Furthermore, even if Alliger's "corrected" figures were 100% justified, then the fact he found a lot of *negative* correlations still would serve to refute the whole Spearmanesque claim that negative correlations do not exist.

Further, let us return to Gould's criticism that due to "validation" of most other highly used IQ tests and subtests, Spearman's g was *forced* to appear to exist from then on, regardless of whether it actually did. In view of this (and in view of the fact that our two 11×11 numerical examples above appear to 100% confirm both Spearman's principles *and* Gould's refuting argument!) probably the only place we can look in the literature to find data *truly* capable of refuting or confirming Spearman, is data from the early days, before too much "validation" occurred, but not so early on that Spearman's atrocious experimental and statistical practices were repeated.

The prime candidate I have been able to find for such data is Thurstone's [205] "primary mental abilities" dataset published in 1938.

[61] SAT used to stand for "Scholastic Aptitude Test" although its makers now tell us that it simply stands for SAT with no meaning other than that. This has nothing to do with the "satisfiability problem" SAT from boolean logic which is a fundamental NP-complete problem [61].

[62] Many others also have multidimensional proposals, but usually far fewer dimensions than Guilford! The "mainstream" view, dating to L.L.Thurstone after corrective interaction with Spearman, seems to be that Spearman g exists – one big eigenvalue of M^{-1} – and the other eigenvalues are substantially smaller *but* not zero and thus lead to higher-dimensional notions of intelligence, but with the nature of the next few dimensions being considerably less clear and less agreed upon than g. For example Thurstone once claimed there were "seven primary mental abilities" which he labeled Verbal Comprehension, Word Fluency, Number Facility, Spatial Visualization, Associative Memory, Perceptual Speed, and Reasoning. Another model was that there was "fluid" and "crystallized" intelligence, with some adding "visual" intelligence to the mix [70]. Another idea was there was "rote" intelligence typified by performance on tasks such as repeating a sequence of numbers, as compared with "level 2" intelligence typified by repeating a sequence of numbers *backwards*. Yet another idea was to try to identify kinds of intelligence with certain regions of the brain, because it is known that brain injuries in different regions affect different kinds of mental abilities [160][125]; and another was "practical intelligence" which concerns mental manipulations thought to have more impact on everyday life than more abstract "academic intelligence." The validity of the latter idea was totally disputed by L.S.Gotfredson and her battle with R.Sternberg may be read in *Intelligence* 31,4 (2003) 343-397.

[63] Although they would have come out about 50% positive and 50% negative, whereas Alliger got majority positivity, so in that sense there is *some* justification to claiming Alliger refuted Guilford, but 70% positivity would not be a "refutation" of Guilford but rather merely a refutation of 40% of Guilford's "zero" correlations with the remaining 60% remaining unrefuted.

Thurstone tabulates the 1596 correlations among 57 kinds of mental tests as taken by 240 people on pages 110-112 of [205]. Lo and behold, only 15 of them were negative (in the range -0.22 to -0.05), while 52 were in $[-0.05, +0.05]$, and the remaining 1529 were positive in $[+0.05, +0.85]$. That mostly supports "the Spearman hypothesis" that all mental tests are positively correlated – but not wholy. A Spearman advocate would speculate the 15 exceptions were just "noise" in which case this would be 100% support.[64] Upon examining the 3 most negative correlations, I find a remarkable thing – see table 17.5.[65]

This seems too much to be a coincidence! Hence, I conclude that this is **a counterexample to Spearman's first principle** in the sense that a genuine negative correlation between two kinds of mental ability appears to exist.

Skeptical? I also re-examined the independently collected 1941 data of Thurstone & Thurstone 1941 [206] and found a *second* such example[66] (with $p \approx 0.00001$). Since it is inconceivable that *both* these counterexamples are delusory, we can now take it as settled that clearly **a counterexample to Spearman's positive-correlation principle exists.**

correl	the two tests
-0.22	100-word vocabulary test // Recognize pictures of hand as Right or Left
-0.16	Find lots of synonyms of a given word // Decide whether 2 pictures of a national flag are relatively mirrored or not
-0.12	Describe somebody in writing: score=#words used // figure recognition test: decide which members in a list of drawings of abstract figures are ones you saw in a previously shown list

Figure 17.5. The three most negative inter-test correlations found by Thurstone among the $1596 = 57 \times 56/2$ correlations tabulated for 240 people taking 57 mental tests on pages 110-112 of [205]. There is a stunning resemblance between these three test-pairs, don't you think? Also, of the next six most-negative pair-correlations (all of which had value -0.10) five of those pairs involved at least one of the 6 mental tests tabulated here. Considering there were 57 mental tests in all, the chance that at least 8 among the 9 most-negative correlation-pairs would, in a random ordering of the correlations, all happen each to involve the 6 particular among those 57 mental tests that constituted the 3 most-negative pairs, was $6[(57 \times 56 - 51 \times 50)/(57 \times 56)]^5 \approx 0.002$, and if we multiply this by the a priori probability ($\lesssim 0.04$) that the particular 3 most-negative pairs would seem this amazingly conceptually similar, we get a very low probability $p \lesssim 0.00008$ that this all is merely a fluke. ▲

Nevertheless (defenders of Spearman would riposte) only 15 negative correlations out of the 1596 (in the Thurstone 1938 study) is not a lot (only 1%), whereas the 1529 positive correlations is a lot (96%) so Spearman was still *mostly* correct. However, attackers of Spearman would riposte that many of Thurstone's mental tests were known a priori to be positively correlated with g (Gould's criticism again) so the 1-versus-96 comparison is misleading. Assuming half the Thurstone tests met this criterion (actually, I have no idea how many did) would reduce us to only about 390 correlations about which we were a priori ignorant, not 1529, of which 15 are negative – which would not be 1% negativity, but rather 4% negativity.

Another bothersome problem – IQ tests lacking right answers:

It is quite an amazing thing to me as a mathematician, but the psychologists and educationalists who devise IQ tests seem to care remarkably little about whether the "correct" answers on their tests actually *are* correct.

For example, the criteria for the "right answer" to Raven's matrices are pseudo-logical or aesthetic, but there is *never* any *proof* any answer is right or wrong.[67] So this test – despite the plentitude of praise heaped upon it by Jensen at every opportunity – is *purely* a "popularity test" or "conformity test" testing whether your aesthetic preferences (or preferences when employing pseudo-logical argumentation), happen to *agree* with more or less of society (specifically, the normalization subsample of society) or with the test-creator, and is *not* a test of how good you are at finding the "right answer."

Many SAT test questions also have historically lacked objectively correct answers (such as "verbal analogy" and "find the

[64] One could try to assess that if we had Thurstone's data so that we could do cross-validation, but he did not publish his data or error bars on his correlations, so we cannot. This is yet another example of the shoddy statistical practice with which the field is so rife.

[65] I am mentioning this because in all previous literature, I have failed to find any mention of a clear case of a negative correlation between two mental abilities. This appears to be one.

[66] We thank Wendy Johnson for emailing us this 60×60 correlation matrix based on 710 Chicago schoolchildren taking 60 mental tests (the tests are re-described concisely in [89]). T&T in 1941 used a new set of tests only partly overlapping their old set, and in particular *eliminated* the problematic tests that had been in Thurstone's earlier 1938 dataset, preventing those negative correlations from reappearing in 1941. ("Validation" in action?) Of the $60 \times 59/2 = 1770$ centered correlations in T&T 1941 matrix, I find that the three most negative ones, with values -0.161, -0.152, and -0.138 respectively, are the pairwise correlations of the performance on the "scattered Xs" test (circle the Xs in a random scattering of letters) with these three tests: (a) Sentence completion (Choose the word appropriate to complete a given sentence), (b) Reading comprehension II (Read paragraphs and answer questions about them), and (c) Reading comprehension I (same description as b). Again, it is difficult to believe this also is a coincidence! The probability that (if this all were just a fluke due to "noise") the three most-negative correlations from the 1770 all amazingly would arise from the *same* row of the matrix was 1/3600, and the probability all three columns would arise from tests this amazingly *conceptually similar* also was small ($p \approx 0.05$?) yielding net likelihood $p \approx 0.00001$ if it were all a mere statistical fluke. So we presume this is a second example of two genuinely negative correlated mental abilities.

[67] Each Raven problem consists of a 3×3 array containing eight related abstract black and white pictures plus one blank spot. A 2×2 illustrative example is given page 37 of [86] and the Raven tests are also discussed in ch.14 of [85]. The goal is to determine, from a set of 4 more pictures, which one would "best complete the pattern." Typically one's reasoning is something like this: "Picture$_{11}$ has three "arms" sticking out and picture$_{12}$ has four; picture$_{11}$ has two little white circular "holes" and picture$_{21}$ has three; therefore the missing picture$_{22}$ should have four arms and three holes – is there such a picture among the 4 candidate answers? Yes!" Well this is all very fine, but, objectively, it is possible to make up a story to justify any answer and there is no clear notion of the "correct" answer. (E.g. I could also easily justify copying picture$_{11}$ into the blank spot.) Raven's matrices are discussed in [27] where in table 1 p.408 five "rules" are stated that apparently underlie Raven's design of the test. Any test taker aware in advance of these 5 rules would have a huge advantage over the test takers going in ignorant.

next term in the numerical sequence" problems) and their answers instead are "justified" because they agree with the most people who do well on the whole SAT test; similarly many other classic IQ subtests ("draw a man," "identify the 'pretty' and 'ugly' girl in two pictures") are non-objective. In many IQ tests the grader often is face-to-face with the test taker, hence is free to exhibit (e.g.) his racial biases when grading subjective tests such as "comprehension" and "vocabulary." While this may not matter much in some settings [85], it was the entire purpose of the "IQ tests" used to deny suffrage to blacks in the US South during the "Jim Crow" era ($\approx$ 1900; one typical "literacy test" was to demand the testee recite the entire Constitution from memory, but such tests were not required in cases where the testee was "obviously" literate[68]).

I must say that I have always found tests without objectively correct answers very disturbing – and, in the case of the SAT (used to decide college admissions) and IQ tests used to make hiring decisions, outrageous. Indeed there have been several well publicized examples in which the SAT's answer was in fact *objectively incorrect* despite the fact that the SAT test-creators and a plurality vote of the normalization sample of society all agreed on it. In those cases the test takers who gave the objectively correct answer, thus demonstrating their superior intelligence to both the SAT test-makers and the bulk of high-SAT-scorers, were downgraded but (after a lengthy protest procedure consisting mainly of the ETS mechanically sending out pre-written form letters explaining why the protestor was a total idiot) in some cases were ultimately upgraded but *still* the SAT-makers always refused to downgrade the bulk of society who gave the old-right but now-wrong answer. (This course of events has only occurred on the math SATs. On the verbal SATs the ETS has, to my knowledge, *never* admitted it was wrong [146], although I am quite confident they were.)

With conformity-testing rather than correctness-testing it seems to me we are incapable of measuring IQs of hypothetical entities *more* intelligent (or differently intelligent) than humans. That is perhaps acceptable if the goal is to measure "Spearman g" which is proclaimed to *be* a conformity-measure and only of interest when testing *humans* (that would appear to be Jensen's attitude when he praises the Raven matrices as essentially "a pure measure of g") but not fine if the goal is to measure "intelligence" as an abstract notion defined independently of the existence of humans.

For example, the test questions "does God exist" and "is Darwin's theory of evolution wrong" would for a conformity test feature (in both cases, according to large polls of contemporary USA natives) the answer "yes," but I rather suspect that an intelligence greatly superior to the present day human average would give a different response.

So – does Spearman's g really exist? Where do I stand?

In spite of the above considerable criticisms (and I am not happy about the poor standard of work in this area) I feel that the *preponderance* of the evidence indicates that

1. Spearman's first principle (positive correlations between mental test results) was mostly right, in the sense that exceptions to it are rare (between 0.1% and 5%).
2. Spearman's second principle (the substantially 1-dimensional nature of human intellect) requires "1-dimensional" to be interpreted quite generously. My interpretation of the data is that the human intellect is extremely complicated and that its higher dimensions (eigenvectors beyond g) are very difficult to deduce reproducibly from the data, *but* g is significantly more important than the other dimensions in the sense that the length of the g-axis of the characteristic ellipsoid is 1.42 to 1.97 times longer than the second-longest axis and 2.07 to 3.92 times longer than the average axis.

The confounding effects we have mentioned probably are not enough to destroy these conclusions but are large enough so that it was unacceptable for previous workers to ignore them.

It seems to me that the proponents of such important hypotheses as Spearman g, heriditary- or racial-linked IQ, and the theory that large racial and geographical IQ differences exist and have determined and will determine world history and economics [118], and those proposing "Eugenics" [116], must provide a high standard of proof to justify their hypotheses! The *critics* of those hypotheses – who are inherently less ambitious since they are not proposing a theory of their own, merely critiquing an attempted theory – in my view are not required to provide nearly as high a standard of proof for their negative case. Thus Alliger's critique of Guilford is fine as a negative statement, but it was nonsense for Jensen to pretend that Alliger's work met the high standard required to provide a *positive* proof of Spearman's g-factor hypothesis. The nature of Spearman's two hypotheses are such that they will require eternal verification and could be refuted at any time. Guilford's proposed tests should be considered as among the prime candidates for further inviligation of Spearman's hypotheses, and they should be investigated by a better investigation than Guilford made – not simply dismissed and shoved under the rug!

The connection to the present work – our theory predicts both positive correlations and Spearman g!

I apologize for the rather long lead-up. My ultimate point is simple.

Hypothesis that human intelligence works like a UACI: Let us *hypothesize* that human intelligence works rather like (precisely how much like, is hard to know or even define) theorem 5 of §12's construction of a UACI – *universal* intelligence (albeit running on rather bizarre ultraparallelized wetware...).

I consider this hypothesis very plausible and indeed I feel that the UACI theorem has largely demystified intelligence by showing that at bottom, there need not be much to it – it is simply a matter of setting up a search-over-all-algorithms and then trying to optimize its design.

Consequences: If this hypothesis is true, then it is not at all surprising that every[69] kind of human mental ability would

[68] Over 130,000 blacks were registered to vote in Louisiana in 1896, but then Jim Crow measures were implemented, with the result that there were only 1342 on the rolls in 1904.

[69] The rare exceptional negative pair correlations – we hypothesize are because one of the particular mental abilities in that pair is substantially governed by something else other than the general purpose UACI. For example, ability to distinguish left- from right-handed images might be an

be positively correlated with every other, since it is at bottom just one universal algorithm that works trying to emulate every possible algorithm for every task. So indeed, positive performance correlations are exactly what one would expect.

Warning: Unfortunately it is not quite that simple. One could imagine a parameterized search algorithm inside the UACI in which adjusting one parameter emphasizes certain specific parts of the search over algorithms while de-emphasizing others, resulting in negative correlations. So we need to hypothesize that such things are rare enough or unimportant enough or undetectable enough or just nonexistent (e.g. because the algorithm is fixed not parameterized or because the parameters act in a low level manner buried under so much complexity that they do not have clearly different effects on different high level mental skills) that they are far outweighed by just one factor (g) which we identify with the overall performance of the UACI algorithm.

Now one can further argue by using quantitative forms of the Perron-Frobenius theorem (such as the one we shall now state), that, *if* the pairwise correlations all are bounded below by a sufficiently positive number, *then* the existence of Spearman g (i.e. high relative dominance of the top eigenvalue) too would be *forced*:

One strengthened form of Perron-Frobenius theorem [77][128][176]: Let U be a square matrix with all entries positive. Then U has a unique (i.e. non-multiple) positive real eigenvalue r, such that $r > |\lambda|$ for all eigenvalues λ of U with $\lambda \neq r$. This eigenvalue corresponds to the unique eigenvector of U with all entries positive real. And indeed

$$\frac{|\lambda|}{r} \leq \frac{1-\mu}{1+\mu} \quad \text{where} \quad \mu = \frac{\min_{ij} U_{ij}}{\max_{st} U_{st}}. \tag{9}$$

A simple bound on r is $\min_j \sum_h U_{jh} \leq r \leq \max_s \sum_t U_{st}$; many stronger bounds are available in [77][128]. This Perron eigenvalue r corresponds to a unique (up to scaling) eigenvector $\vec{x}$ of U (obeying $U\vec{x} = r\vec{x}$). This $\vec{x}$ consists entirely of positive real numbers, and indeed obeys

$$\frac{\max_j \sum_h U_{jh}}{\min_s \sum_t U_{st}} \leq \frac{\max_j x_j}{\min_h x_h} \leq \max_j \frac{\max_{s\neq j} U_{sj}}{\min_{t\neq j} U_{tj}} \tag{10}$$

so that $\min_h x_h$ is not only positive, it furthermore cannot be too small.

18 Piaget's observations – lessons learned from children

Precis. We attempt to summarize the state of "Piagetian theory" concerning the development of intelligence in children. We then see that it is compatible with HUH, and thus the observations of Piaget and his followers may both be regarded as confirmatory evidence for it, and as casting a useful illuminating light on the nature of the human UACI's "search over algorithms."

Jean Piaget (1896-1980) began his career by studying development, growth, and adaption to changed circumstances in mollusks. He then turned his attention to the development of intelligence in small children, arguing that "It is with children that we have the best chance of studying the development of logical knowledge, mathematical knowledge, physical knowledge, and so forth." Piaget was responsible for the notion that children's intelligence (as well as many facets of it) develops in a sequence of stages. The sequence is invariant in different cultures and no stage is skipped. At each stage the child progressively gives up erroneous ideas for more correct ones, or more precisely transforms initial inadequate ideas and strategies into more adequate and sophisticated ones.

Piaget described his findings in many books. These books are annoying to read because they consist in large part of extremely detailed anecdotes about interactions with children. The stages are denoted by Piaget with names like "stage IIIb." Because Piaget mainly worked with small samples in a rather unscientific manner (he was a careful observer, but he did not do large or controlled experiments and did not employ statistics) some but not all of Piaget's claims have held up under later scrutiny. Mostly, they have proved reproducible with children from several cultures (Chinese, American, Australian Aborigine) tested during the 1960s and 1970s. The main cases where Piaget's original ideas have been shown to be incomplete are in cases where later investigators have had much better equipment than Piaget, such as videotape, electronic nipples which babies can suck to cause different reactions, and devices which examine a child's eyeballs to get a record of what they look at and for how long. These allow informative experiments even with very young babies.

E.g, at the age of *2 days* babies prefer to listen to speakers of their mother's language over speakers of foreign languages; this has been taken as evidence that they were listening and processing statistical data while in utero.

Example: objects. Babies have a small amount of innate capabilities and innate understanding of objects.[70] At the age of *1 day* babies already can visually distinguish faraway from nearby objects. Two month olds try to learn about objects by grasping and sucking them (which seems an innate preprogrammed knowledge-acquisitional behavior). According to Piaget and some others, children initially act as though objects only exist when they are in their visual field. They will not try to find objects they witness being hidden behind an opaque screen (and appear to be when surprised that the object still is there when the screen is removed) – but if the screen is transparent the babies will try. This is not because babies have no memory – it can be demonstrated that babies this age remember other events for days or weeks. But at 8-9 months, the child develops the notion that objects permanently exist (e.g. they remember them, and may even try to find things they can no longer see).

However, by using experimental techniques unavailable to Piaget,[71] some other experimenters [9] *have* detected *some*

innate ability, *not* an algorithm created by the UACI. Because of the light they cast on the UACI-in-humans hypothesis, all putative negative correlations deserve deep experimental investigation to try to determine how innate they are and/or how they develop in infants.

[70] Other newborn animals have considerably more innate competence than human babies, for example horses can walk, see, and process visual information almost immediately after birth, and newborn turtles are fully capable of independent existence.

[71] On such experiment is as follows. A board swings back and forth, pendulum fashion, along an axis. The baby witnesses an object being placed behind the board and hence out of its line of sight, in a position where it will intersect the swinging board causing it to "bump." However (thanks

amount of understanding of the following four concepts even in 3 month old babies:

1. unsupported objects fall (although the babies do not understand *how much* support is required),
2. objects move along continuous paths,
3. solid objects cannot pass though each other,
4. objects still exist even when hidden from view.

This is perhaps because these ideas are innate, or perhaps because their development happens on a chronological continuum which, contrary to Piaget, *cannot* necessarily be fully neatly compartmentalized into different stages.

8-to-12 month old babies, although now realizing that objects are permanent and they will search for objects hidden behind opaque barriers, sometimes make an interesting mistake. If they see an object hidden twice under cup A, they retrieve it successfully each time – but if the object is now hidden under cup B, they sometimes will still look under cup A for it. Between 12 and 18 months of age this error ceases and they now look directly under cup B (or wherever they last saw the object hidden). Between 18 and 24 months, babies even understand complex recursive sequences of hiding, such as hiding an object under a cover, then both together are hidden under a pillow, then the cover is removed so the toy remains behind the pillow.

At 9 months babies get the idea that objects can also be investigated by shaking, hitting, and throwing them, and at ages 12-18 months they perform more kinds of experiments with objects such as dropping them systematically from different heights, singing to dolls, and putting them in bed. 1-year old babies observe how people react to objects and will imitate them (sometimes remembering the new behavior and first imitating it a week later) as a means of "plagiarizing" new experimental methods, and 1-year-olds also point at objects in an apparent effort to elicit parental reactions to the object. As a consequence of all this experimenting, babies learn more and more properties of objects. They do not initially understand most of them. For example, a baby whose leg is tied to a mobile will move its leg in order to cause the mobile to shake and rattle, but will keep kicking even when the rope is disconnected; it does not understand the function of the rope. Babies will pull a cloth with a toy on top of it toward them; but if the toy is placed to the side of the cloth, this no longer works as a method of acquiring toys, but the baby will pull on the cloth anyway and seems surprised to see nothing happens. By the time babies are 18 months old these two conundrums are understood and babies will demonstrate primitive uses of tools such as pulling out-of-reach toys toward them with a rake. At 5-7 years they learn to classify objects into groups and hierarchies ("animals," "blocks of different shapes") and mentally, rather than merely physically, manipulate objects. (DeVries [49] found that many 3-year-olds thought that a cat, after the mask of a fierce dog had been put on its face, was now a dog, whereas 6-year-olds refused to buy that notion. Piaget found that children below age 7 or 8 will say there are more cats than animals, even though cats are inherently less numerous.) Careful logical reasoning about them happens considerably later.

Example: area, volume, and weight. At age 4, children have only a one-dimensional length-based notion of quantity. This notion, while incorrect, still is an adequate approximation for many purposes. For example, children think there are "more" checkers if the checkers are spread out further in a longer row, and most children below age 7 do not understand that the volume of a liquid does not depend on the shape of the container – they usually will claim that a taller and narrower container holds "more" liquid even if they witnessed the liquid being poured into it from a shorter and wider container. (Also, if asked to draw a tilted half-full bottle, they do not initially understand that the water in the bottle will have a horizontal surface even when the bottle is tilted.) Similarly children do not initially understand that area is invariant under rearrangement of subshapes, and that the weight of a clay ball is invariant under changes of shape. During age 4-12 years, according to Piaget, they realize the invariance of area first, weight second, and volume third.

Between ages 2 and 4, children have trouble even with their 1D quantity notion – they usually experience great difficulty if asked to sort seven sticks in order of increasing length. They are capable, however, of distinguishing *one* object from *two* even in their first half-year, although they cannot reliably distinguish *four* from *5 or 6* objects until age 3-4.

Additional examples concern the child's acquisition of language, and his increase in ability at logically planning and acting. All these may be broken down into developmental stages which seem to happen in a fixed order according to an approximately predictable schedule in a culture-independent manner, and in which there is a preprogrammed desire for, and mental "reward" structure for, knowledge-acquisition.

In some cases abilities are *lost* rather than gained with time. For example, 1-to-7-month old babies seem capable of distinguishing every sound used in all human languages. But while the English language employs both r and ℓ sounds, the Japanese language does not. Japanese babies permanently lose the ability to distinguish r from ℓ sounds at about 10 months of age, whereas English babies retain and strengthen this distinguishing ability. (This is environmental, not genetic; Japanese babies brought up English learn r versus ℓ fine, and there are other language- and sound-pairs where the same phenomenon happens.)

There is evidence that some mental feats are *only* possible *during certain age ranges.* A 13.5-year old California girl named "Genie" [170] was rescued after spending most of her previous life tied to a potty in a small room and beaten whenever she made noise. Her father growled at her like a dog. After rescue, she seemed *unable* to learn language. (She initally knew a few words and eventually several hundred, but could not link them into valid sentences longer than 2 words: "Mike paint." "Genie cry ride." "Applesauce buy store." "Neal come happy; Neal not come sad." Her language development appeared permanently stalled at about the same level as a 2-year-old's although progress perhaps was continuing at an extremely slow rate. Tests suggested she had high nonverbal intelligence.) "Chelsea," who was born deaf and wrongly diagnosed as mentally retarded, regained her hearing with

to some optical tricks) the board does *not* cease to swing in full oscillation, accomplishing the impossible. The baby is apparently surprised by this in the sense that eyeball-tracking equipment shows it looks longer at the situation than at analogous non-paradoxical situations.

electronic aid at age 31 but was also unable to learn language ("Combing hair the boy," "The woman is bus the going," "Banana the eat." [152] p.296-299). But a similar life story was experienced by a Ohio girl named "Isabelle" who was locked in an attic with her deaf mother until age 6.5 by her grandparents [122]. Isabelle was unable to speak or understand language and acted like an animal when rescued, but during the ensuing 2 years under special care, she learned to speak and understand language and developed other human skills at approximately three times the normal rate, so that she apparently became a normal 8-year old girl.[72] Children who learn second languages at ages 3-7 perform like native speakers on various tests but their performance declines with age-of-learning from age 8 through puberty, then becomes flat (no age correlation) for languages learned after that [139]. Kittens cannot learn to see out of their left eye if that eye is covered up to an age of 80 days. Male white crowned sparrows learn to sing during a 30-day window between age 20-50 days, but if they do not hear the right song during this window they cannot ever sing normally and consequently cannot breed.

Conclusion. Now all this is entirely compatible with the idea that the child is *running a fixed super-algorithm whose purpose is to build an intelligent entity, and which proceeds by successive exploration of algorithms*[73] *with more refined modified algorithm versions being tried once a good one has been found.* And that is exactly what is predicted from the hypothesis that human intelligence works similarly to §12's construction of a UACI, *but* with the important changes that

1. The "exploration of all algorithms" proceeds, not simply in lexicographic order, but instead via exploring refinements of initial skeleton-algorithms, then if a refinement constitutes an improvement, further refinements are considered and
2. Some "master scheduler" or "homunculus" permits certain kinds of explorations to happen (or provides appropriate kinds of "score rewards") only during certain time-windows, and
3. the homunculus has preprogrammed innate structures, [74] innate knowledge-acquisitional behavior and preprogrammed ways to evaluate what learning is "successful" and should be "rewarded," and
4. There is preference for "simpler" algorithms.[75]

Call this kind of algorithm-exploration strategy "**Piagetian search**." It has some obvious advantages over plain lexicographic exploration, namely one can "rapidly get a crude system up and running" and (hopefully?) there is much greater search-efficiency this way. (It might be thought to have the disadvantage that some algorithms may be unreachable by Piagetian search, but some simple "safety backup" procedures such as conducting a true-exhaustive search 5% of the time, could overcome that problem.)

19 Forgetfulness

Precis. Both the existence of and the nature of human forgetfulness are shown to be compatible with HUH. This againd may both be regarded as confirmatory evidence for it, and as casting a useful illuminating light on the nature of the human UACI's "search over algorithms."

Speaking as a human, it is utterly embarrassing to consider how poor our memory is. Although computer programmers cannot currently duplicate human intelligence, they would have to *intentionally try* to build an *incredibly bad* memory to be as bad as human memory. And computer engineers via error-correcting codes *still* know how to make computer memory essentially perfect even with unreliable underlying hardware.

① It appears that human skills and memories stay sharp only if they are continually *used*.

One quite elegant demonstration of this fact was a self-experiment by Marigold Linton. Every day for 5 years she woud note in a diary two events that happened that day. According to a pre-designed schedule she would randomly select diarized events and judge whether she could recall that event. Because of the random sampling, some events were queried more than once. The results ([7] p.106; [109]) were: there was 65-100% forgetting after 4-5 years for events queried once and only once; but less and less forgetting for events queried more times, with events tested 4 or more times being only 40% forgotten even after 6 years.

But – and this is our point – this is entirely compatible with the UACI hypothesis.

Specifically, hypothesize that "human skills" are algorithms which are continually being modified in a partly-randomized Piagetian manner to try to improve performance. Now, in our IQ test, imagine our UACI is asked to keep solving 3-SAT problems. Eventually, suppose it learns to become quite good at that. Then one day, the problem generator instead starts posing "prove this mathematical statement from these axioms"' problems. The UACI then will desperately modify its algorithms in an effort to get good performance on the new task. After a long time, let us suppose it eventually gets good at that. If the problem generator now switches back to 3-SAT problems, will our UACI immediately fall back on its old 3-SAT solver and be able to get good performance immediately? *No!* The modifications will quite probably have destroyed the old 3-SAT solver (although enough of it will probably remain that relearning how to solve 3-SAT problems would then get accomplished more quickly than the first time. (② And psychological experiments confirm that relearning something is typically faster than learning it [137].)

So what is stopping the UACI from simply *remembering* its old 3-SAT solver (you ask)? Well, it is not so easy. Our UACI construction does not "know" that one kind of problem is "3-SAT" and another kind is "axioms→proof." All it gets is bitstrings. If it "realized" that there were two distinct kinds of

[72] But this data is necessarily from a very small sample. We cannot be certain that the key quantity here was age, or whether some other difference in the experiences of Genie, Chelsea, and Isabelle were the key.

[73] And of all "theories of the world," which however can just be regarded as "algorithms."

[74] See footnote 92.

[75] This explains the Japanese babies losing their unnecessary initial ability to distinguish ℓ from r – it was "simpler" for them to drop that from their repertoire.

problems and which was which, it could copy its 3-SAT solver and then work on developing a separate theorem-prover. However, by the time it "figures that out" it may already be too late. (The "new" problems might just be a harder kind of 3-SAT problem with a slight format change, for all the UACI knows.)

This problem could largely be solved by continually making "backup copies" of the UACI's "brain" enabling "old versions" to be "recovered" [52]; but I conjecture that biological limitations prevent humans from "copying their brain" easily. (Really, you would want to copy only the "correct parts" of your brain – constituting the 3-SAT solver – but it is not necessarily easy to identify what those parts are, especially in a biological and multitasking setting.) Further, keeping a record of all old-versions, or even a small fraction of old versions, might be expensive enough to render this solution not worth the cost from the viewpoint of Darwinian evolution.

Enquiring more deeply into the nature of human (and animal) remembering and forgetting leads to more evidence that humans and animals work like our UACI, and more understanding of the details of that. (We shall continue numbering facts with circled digits to make it easy to keep track of them.)

③ Cockroaches can be trained to remember not to enter attractive-looking areas, because when they do they get an electric shock. Only 25% of *immobilized* cockroaches forget that training 24 hours later, but of cockroaches allowed to wander around during the intervening 24 hours, thus (presumably) keeping their minds more busy, 70% forget ([7] p.108). In other words, memory interference effects happen in cockroaches as well as humans, presumably for the same reason: the UACI hypothesis.

④ Humans trying to learn things suffer from "interference" ([7] p.111, [189]). E.g, if they are given a passage of prose to remember facts about, their recall is better if they were *not* later exposed to other prose passages about related subjects. And there have been a very large number of other experiments (e.g. [84][99]) demonstrating interference effects of various kinds on human memory.

According to A.D.Baddeley: "Interference" is the assumption that "forgetting reflects the disruption of the memory trace by other traces, with the degree of interference depending on the similarity of the two mutually interfering memory traces." A deeper idea [196] is the theory that human memories can be regarded as *associations* between two things "A-B" (e.g. a name with a face, a word with a definition, a country with its capitol) and then interference arises when you, either *before* or *after* remembering the A-B association, remember or consider an A-C association. Both experimentally reduce recall of A-B association, but the former kind more, and the effect is more serious the longer the tested memories are to be retained.

⑤ A related effect is that most people's memories can be both qualitatively changed and quantitatively distorted by asking them "leading questions" and by asking questions incorporating new spurious information. For example, people shown film of a car crash could be made to alter their speed estimates and be made to recall totally spurious events and objects by asking them such questions. Were these people merely trying to please the experimenter by telling "little white lies" or were their memories genuinely altered? Loftus tried to distinguish by offering several different schemes of monetary rewards to encourage getting more right answers – but these had no effect, which she regarded as evidence that the memories were genuinely altered. ([111][113], [7] p.183-185, [110] p.118-120).

⑥ Almost all mature humans exhibit "infantile amnesia," i.e. they can remember essentially nothing that happened before they were 3 years old. This is not because infants have no memory; even 3-month-old infants can be demonstrated clearly to remember things for several days ([7] p.215-221). This is compatible with the hypothesis that the human UACI is "tuned" by an external "scheduler" in such a way that heavy rewriting and algorithm overturn happens during ages 0-3 but it is less heavy at later ages. This *scheduled tuning hypothesis* is also supported by findings ([7] p.223-224; [29][59]) that (a) children's memories are more easily distorted by misleading questioning the younger they are; (b) forgetting is faster among 6-year-olds than it is among 9-year-olds than it is among adults.

⑦ Human memorization is tremendously improved with overt or internal-mental "rehearsal"; if people are prevented from rehearsing by being required to count backwards by threes from some moderately large number, their recall rates fall by a factor of ≈ 10. ([228] p.31 discussing 1958-1961 work of Brown, Peterson, and Murdock).

⑧ Memory is improved by "chunking" tricks. For example, a famous experiment by Herb Simon which you can try yourself is to remember these 10 words:

> Lincoln way criminal differential milky address lawyer calculus galaxy Gettysburg.

Simon and most people are unable to recall all 10 words after a short exposure. *But* Simon and many people have no trouble recalling them all if they are reordered as follows:

> Lincoln's Gettysburg address
> milky way galaxy criminal lawyer
> differential calculus.

Another trick is to use "mnemonics" (silly rhymes and so forth) or the "method of loci."[76]

Points ①-⑧ all seem entirely compatible with the UACI hypothesis *with these additional modifications*:

1. "Human skills" are algorithms which are continually being modified in a partly-randomized Piagetian manner to try to improve performance.
2. The algorithms being altered in a semi-randomized manner are preferably the ones that are *activated*.
3. Memories and (what is the same thing) stored algorithms keep track of their "utility" (which is some increasing function of how often they used and how successful those uses are) and low-utility algorithms are

[76] To memorize a sequence of items, you imagine taking a sequential trip along some familar route, associating each item with the mental image of each location along your route. Then you recall them by mentally re-walking along the route retrieving the stored items. This technique supposedly often is highly effective.

the ones that are preferentially overwritten. Successful retrieval and use of a memory causes its utility to rise. Memories from infancy have lower utility. Other humans talking about a memory cause its utility estimate to rise. Memories with more "links" to others[77] ("chunking," "method of loci," and "mnemonics" create such links, in the latter two cases artificially) get higher utility estimates and memories one tries to remember more times have higher utilities (hence the usefulness of artificial "rehearsal").

4. Much of this is also true in animals very far from humans in the evolutionary tree.

20 Time-consumption behavior

Precis. First, we argue that UACIs and humans both exhibit "exponential roll out" behavior. Second, we consider the "power law of improvement with practice" exhibited by humans on a wide variety of tasks. We prove a theorem that our UACI construction (for certain kinds of tasks) also exhibits power law learning, and argue (under standard computational complexity assumptions) that on those tasks it is not possible to learn faster than a power law. Both of course are wholy consistent with the HUH.

"Exponential roll out." An important typical behavior of our UACI construction in §12 is that the UACI solves a long sequence of problems very *poorly* for an *exponentially long* time (more precisely, exponentially long in the *code length* of some algorithm that would rapidly solve that class of problem well[78]) until it finally "catches on," after which it solves problems of that ilk in great profusion, rapidly, and well. We now claim that is *also* true of human intelligence.

1. Consider the problem (which we discuss as example 7 in §22) of multiplying two numbers. It took humankind about 150,000 years to devise the first polynomial-time algorithm for multiplying two numbers, and then it took an additional 5000 years to invent the first subquadratic-time algorithm. In the meantime perhaps 10^9 man-years were wasted by humans using inefficient methods to multiply numbers! However, modern humans who know those algorithms multiply numbers in minutes, and with mechanical aids in seconds. The problem of inventing multiplication is deemed so difficult and so worth solving that all schoolchildren spend years learning these algorithms. Our point is: the *reason* that is deemed to be worth this huge human time-cost is evidently because even that huge cost is deemed smaller than what rediscovery would cost.

2. Now consider Newtonian physics. Obviously, it is central to engineering everything from sailing ships to catapults, hence central to understanding and using our environment, and essential for the well-being of modern humanity. Once known, it is easy to use. But it took 150,000 years for humanity to invent it, during which time an enormous number of problems were solved more poorly and/or more slowly than they could have been solved with Newtonian physics.

Interestingly, the development of Newtonian physics can be considered analogous to Piaget's "stages of development of intellect in children" by successive transformations of less adequate into more adequate and sophisticated conceptions. I.e. when an Ancient Roman engineer designed a catapult or ship, he was of course immensely handicapped by ignorance of Newton's laws. But evidently he was not entirely helpless. He could, for example, design and build a (non-working) catapult, then decide what went wrong ("arm too short, rope not strong enough") and by a procedure similar to binary search a hopefully better design could be built, and so on, until eventually a good design was found, then the Romans could stay with it. And the Roman engineer could use arithmetic to calculate, e.g. that a ship would need X timbers which would weigh Y and require Z worker-years to produce. These kinds of procedures, while not as good as understanding Newton's laws, were eventually adequate to build both ships and catapults.

3. Although chemistry, especially of bio-related chemicals, is central to our lives, humanity *still*, for the most part, has not managed to equal the chemical expertise of bacteria, despite 150,000 years of investigative effort. Again, the history of chemistry has exhibited Piaget-style transformations of less-adequate into more-adequate and sophisticated conceptions.

4. Finally, consider "Rubik's cube" puzzle [11][185] discussed in example 2 of §22. Surely the reader would agree that typical humans take a very long time to determine how to solve the puzzle, but once they finally have developed solving-algorithms, they can unscramble Rubik cubes at a (comparatively speaking) extremely fast rate from then on?[79] F B

The reader can doubtless think of other examples (besides the Rubik cube) from his own experience of where a task became easy only after a long preliminary investigation (e.g. learning to walk).

So it should be clear from these examples that human intel-

[77] Including links to pre-existing memories. For example, Chase & Simon [30] and de Groot [47] found that chess experts did not exhibit better memory of randomly-constructed chess positions than average people, but on chess positions genuinely arising from *games*, the chess experts exhibited clearly superior memory, presumed to be caused by them recognizing and remembering entire *patterns* of chessmen instead of one-by-one, or caused by establishing mental links to patterns pre-existing in their memories.

[78] And 7 of §12 argued that this exponentially long delay was, in fact, *unavoidable*.

[79] In competitions, the world's fastest solvers unscramble cubes in under 20 seconds average time, but (by their own accounts) their algorithms took years to develop. It is commonly *conjectured* that the "superflip"

$$S = S^{-1} = D^{-1}R^2F^{-1}D^2F^2\ U^2L^{-1}RD^{-1}R^2\ BFR^{-1}U^2L^{-1}\ F^2R^{-1}U^2R^{-1}U^{-1}$$

(which flips all 12 edges) is a configuration furthest in the face-turn-move-count metric (20 face turns) from the start position. [The superflip commutes with every element of the Rubik group (and is the unique nonidentity element which does so) i.e. is the group's "center." It is known that 20 moves are necessary to accomplish it and that S locally maximizes the distance-to-start function; also *every* configuration of the 8 corner cubies with the edges held in the superflip or centrally-reflected superflip position has been optimally solved by Tomas Rokicki with the result that over 1000 distance-20 configurations were found, but none with distance$\geq$ 21. Systematic investigations of every cube configuration with enough symmetry have also been done by H.Kociemba & Silviu Radu, with the same results.] $Q = L^2FBR^2U^{-1}B^2D^2B^{-1}R^2UDL^2B^{-1}U^2F^2U^{-1}$ Tens of thousands of cube-configurations are currently known at distance 20 but none at distance 21. However, this conjecture remains unproven even *32 years* after Rubik invented the cube in 1974, despite the fact that *millions* of people have played with these cubes since commercialization in 1977.

ligence *does* exhibit the same kind of "exponential roll out" phenomena that our UACI constructions do – which could be regarded as yet more confirmatory evidence for the Human UACI Hypothesis.

"Power law of improvement with practice." An empirical psychological law is the "power law of improvement with practice." It is highly reproducible and robust over a wide range of different experimental variations [134]. It states that if people repeat some mental and physical task (but with significant mental component, but not so complicated that people are not immediately capable of doing it) N times, they get better at it in such a way that the time to do the Nth task is proportional to a negative power of N. For example this has been demonstrated for this task: "a subset of 10 lights suddenly light up. Your job is to push the corresponding subset of 10 buttons as rapidly as possible (simultaneous pushes allowed)." It also has been demonstrated for this task: "a 4-digit number is suddenly presented to you. There is a certain fixed set of rules for successively converting the first two digits to a single digit, and your job is to repeatedly apply the rules until only one digit remains, then to report it." The first task [175] seems to exhibit exponent≈ -0.32; the second [226] has exponent about -0.4; and the 15 exponents tabulated in [134] p.4 range from -0.06 to -0.81.

But what does this tell us about (a) how humans work, or about (b) how to build an intelligence? One's initial response to (b) is that quite possibly it tells us nothing because quite possibly an intelligence superior to humans might exhibit entirely different and better behavior. Indeed even *humans* do not exhibit this behavior on mental tasks difficult enough that they need to make "aha" breakthroughs, leading to sudden, often-enormous increases in performance, resulting in stairstep-like rather than power law behavior.

Newell & Rosenbloom [134] in 1981 devised a model of human mental operations they called "chunking" which they argued explained the power law. And indeed, upon programming a computerized chunking-involving learner [166], they discovered that indeed, it empirically did exhibit power-law learning, thus "confirming" the theory that humans work via chunking. However, quite possibly other learning and improvement mechanisms having nothing particularly to do with chunking[80] could also exhibit power law behavior.

Indeed (just to name one random example – there may be more) Baum & Durdanovic [13][14] proposed as a model of human mental operations, or as a way to build an AI, the construction of an "artificial capitalist economy" in which "monetary" rewards were externally granted for successful completion of mental tasks, and "agents" could "bid" in an "auction" for the right to perform next mental step and then garner as their share of the profits, whatever the difference between their winning bid and the next auction's winning bid, was. Agents were randomly generated all the time R2 U- B2 D2 B-R2 with unsuccessful agents going "bankrupt" and being eliminated while successful ones stayed around and could spawn variants via random perturbations. Baum's idea was that the world's actual capitalist economy all the time allows impressive cooperative feats to be accomplished (such as building a large set of deep sea oil-drilling rigs and a network of distribution pipelines) by enormous numbers of workers who are "idiots" in the sense that they understand only a tiny fraction of the whole task (such as "how to cast alloys used in corrosion-resistant valves"), and all without there necessarily being any centralized controller. Baum's motivating thought was that the construction of an intelligent entity from comparatively idiotic components should also exhibit such properties, and he also pointed out that far "more perfect" capitalism could be achieved in such an artificial construct than is achieveable (due to moral and other constraints) in the real world.

This whole artificial-capitalistic-economy idea seems very interesting and inspiring, and could also lead to more understanding of economics. It is one of the key ideas explored in Baum's popular-science book "What is thought?" [12]. For our purposes, what matters is that Baum & Durdanovic built several computerized economic systems (called "Hayek$_n$" for $n = 1, 2, 3, 4$) to experiment with all this,[81] and some of their empirically observed Hayek learning curves (although they of course call them "accumulated profit curves") printed in their papers [13][14] *also* appear to exhibit power-law learning – as well as stairstep "aha" effects and random-looking noise resembling "stock market price graphs."[82]

The moral is that we could equally well conclude that the psychological experiments "confirm" the theory that the human brain is not a chunking system, but instead is an artificial capitalistic economic system!

And indeed this confirmation is even better than for chunking because humans also exhibit random fluctuations and "aha" insights. Indeed, I suspect that one could also use the same data to "confirm" other vague theories of how the human mind works – I very much doubt that chunking and artificial capitalism are the only ideas compatible with power-law learning.

Let us now return to the present paper and the UACI of §12. If the UACI lives inside a computational model in which chunking or a Baum-Durdanovic artificial economy is the best way (up to a constant factor) to learn to solve some particular class of tasks, then the UACI (by the competitive asymptotic optimality theorem) will eventually also exhibit the same "learning curve" as these systems.

There might be tasks, however, on which the UACI will exhibit a *better*-than-power-law learning curve.

Power law learning Theorem. *There exists (and we shall construct) a class of randomized tasks on which (a) power-law improvement with practice is achievable (and will be achieved by a UACI), (b) it is not possible for any polytime algorithmic entity to do better than that power law (under the standard conjecture that AES-like cryptosystems cannot be broken faster than by exhaustive search).*

[80] Although the issue is complicated because "chunking" is vaguely defined, and other learning paradigms also often are vaguely defined and two such paradigms might well be said – or not – to overlap, the question being subjective...

[81] Incidentally, the setup of Hayek$_n$ was in its broad outlines quite similar to our proposed "intelligence test." Similar remarks could be made about the setup of "neural net learners"and of "statistical data-fitting systems."

[82] *Warning*: B&D did not point out and did not consider this power law, and hence any conclusion that it really exists is much more dubious than if they had made a careful and specific examination of this issue. I am working extremely crudely by just examining some of their printed curves by eye.

Proof. Fix a real constant $s > 1$. The task (IQ test) is as follows. PG selects a random integer $n \geq 1$ with probability $n^{-s}/\zeta(s)$. This integer (in binary format) is the "problem" P_k. The corresponding "answer" A_k is a single bit determined as follows: Encrypt the binary integer n, padded with 0-bits, using an AES-like cryptosystem with some secret key, and the first bit of the result is the answer. SC awards correct answers score 1, and gives incorrect ones score 0.

We assume the AES-like cryptosystem has word length and key length large enough to (a) U D be effectively unbreakable (so that A_k is effectively a random boolean function of $P_k = n$) and (b) so that the event that n is too large to encrypt, effectively never happens.

We assume the ET soon recognizes (or is initially informed) that the answers are always 1-bit long and that they are a deterministic function of n.

In this case the best that any ET can do (under the standard conjectures AES is unbreakable in polynomial time, and that it is impossible even to get any statistical prediction advantage about AES bits) is simply to *remember* the answers to previously-seen problems and regurgitate them whenever the same problem rearises.[83]

The following behavior will then happen. If we run ET for $c > 1$ times longer, then it will experience roughly $c^{1/s}$-times-larger values of n and hence will build a table roughly $c^{1/s}$ times larger. Consequently, the probability that it will not know the answer to the next problem, will be $c^{1-1/s}$ times smaller. This (for any fixed $s > 1$) is precisely a power law diminution in error rate with time (and the exponent is adjustable by adjusting s), and it is not possible to do better. Q.E.D.[84]

So we have demonstrated that both humans and UACIs exhibit power-law learning under the right circumstances, consistent with the Human UACI Hypothesis.

21 Consciousness – still a mystery?

Precis. Although psychologists had experienced immense difficulty trying to devise a consensus definition of "intelligence" (a quest which hopefully has now ended), "consciousness" is an even *more* murky and elusive concept. Because of that murk, we cannot confidently provide a consensus-inspiring definition. Nevertheless we try by providing a "tentative proposal" of a definition. If it is accepted, then consciousness is trivialized because any "intelligent" entity automatically is conscious. We then provide a negative discussion that refutes all the most commonly-heard alternative notions about consciousness.

The most dramatic feature of human consciousness is the fact that it is turned off (sleep) one-third of the time. We provide a long discussion of sleep. What is our rationale for including that discussion? We claim that sleep is a logically-crucial issue for the following reason. We have in the last three sections provided a large amount of confirmatory evidence for the Human UACI Hypothesis (HUH) extracted from the experimental psychology literature. A critic might now carp that perhaps that evidence was "cherrypicked," i.e. that I looked through the psychology literature seeking confirmatory evidence but ignoring evidence that mitigated against the HUH. That is not the case – my search simply did not uncover any countervailing evidence. However, the closest thing I know to countervailing evidence, is sleep! That is because the HUH does *not* predict sleep. HUH also does not forbid sleep, but conceivably some other hypothesis about how human intelligence works, *would* predict sleep – and if it also predicted all the same phenomena that HUH predicts, then the experimental confirmation of that other hypothesis would have to be judged superior. Furthermore, we are capable of exhibiting (and we do) two kinds of computer programs whose performance is inherently increased by "sleep," causing us to worry that there indeed *might* be such an alternative hypothesis lurking. That is our rationale for examining known facts about sleep in considerable detail. We find (in agreement with previous workers) that the most obvious guesses about sleep all are known to be false. The only proposed sleep explanations that currently appear still to be standing are the hypotheses that sleep is either an evolutionary accident or merely intended to keep animals "quiet and out of trouble," and for many animals – to a large extent including for humans – sleep is in no way necessary nor even helpful for any known facet of our intelligence. Indeed, it appears likely that quite-intelligent animals exist that never sleep. In view of this (pending any increases in our understanding of sleep) we conclude that sleep *cannot* constitute a challenge to the HUH.

What is "consciousness"? Although we have proposed a mathematical definition of "intelligence" we have not proposed one for "consciousness."

The reason is because I did not understand what consciousness is. In fact, for a long time, to a good approximation, the closest I was able to come was the (rather pathetic) "a consciousness is an intelligent entity which asserts it is conscious"!

However, since we feel the need to do *something* about this issue, we propose, as a preliminary stab at it, the following (which, if correct, largely trivializes the question):

Tentative Proposal: A consciousness is "an intelligent entity which interacts with some law-obeying randomized *external environment* in an effort to increase some kind of numerical *reward*."

Note: Any entity "intelligent" according to the present work's definition then would automatically also be "conscious"!

Unfortunately, our Tentative Proposal perhaps would classify the human *un*conscious mind (discussed below under the name "the puppeteer") also as a "conscious" entity – just a rather inaccessible and not-directly communicative one. (The human conscious mind would be one part of the puppeteer's "external environment," and the reverse is also true.) Fortunately, I doubt that, because the puppeteer probably could not be made to take or pass an intelligence test. That speculation, however, is an experimental question that will eternally

[83] Of course, once ET is run long enough to actually break the cryptosystem, then the power law will end and there will be 100% correct answers from then on. Similarly humans cannot power-law improve forever since there are physical limits on their performance. However, in practice the power law in both cases can persist for a very long time – in the cryptosystem case for a time exponentially long in the key length.

[84] The scenario in this proof is not as bizarre as it sounds; really *any* task whose answers are hard to predict but easy to memorize and whose questions may be regarded as integers arising from some power-law sort of distribution will do.

be subject to possible refutation when and if better means of communication with the unconscious mind are devised.[85]

We shall now discuss consciousness and come to many negative conclusions saying what a consciousness *isn't*:

1. It need not be (and for humans is not) the top (root) level in the tree of subroutine calls.
2. It also is not the bottom (leaf) level.
3. It also can be (and for humans is) very unaware of many of its own workings – the commonly-heard claim that an important characteristic of conscious entities is that they are "self-aware" seems complete nonsense.

The involvement (or not) of top level conscious thought is an important technique used by humans. Many tasks we learn, such as learning to walk, learning to play tennis, or learning the right suffixes for Spanish verbs, require conscious thought (such as trying to model what will happen when one swings the racquet in a certain way), but later, the able walker, tennis player, or talker accomplishes the same or greater feats more ably *without* conscious thought.

Some problems (such as words one cannot think of, but suddenly occur to you a day later) evidently are being attacked "in the background" by some semi-consciously created (but therafter autonomously and *un*consciously operating) mental agent.

Other activities humans engage in *never* appear to be fully consciously controllable, although they plainly are controlled by and sensed by your brain. Examples include sexual arousal and fertility, sensations of "hunger," "pain," "fear," and "pleasure," hormone levels,[86] heart rate, full muscle control,[87] and control of and understanding of what most of your internal organs are doing most of the time.

So evidently there is, in your brain, some other thinking, sensing, and controlling entity or entities besides your consciousness. One might call it the "unconscious" (which Freudians have without evidence divided into two parts, the "id & superego"), the "dark silent overlord," the "homunculus," or the "puppeteer." This other entity knows what your internal organs are doing and controls them. It decides whether and when to tell your conscousness to experience "fear" or "pleasure" or "pain." It controls and guides the development of your conscious intelligence, perhaps administers "rewards" or "punishment" to it, perhaps "eavesdrops" on it, and decides when to activate it and (to a large extent) when to shut it off. This puppeteer evidently controls and uses a vast amount of mental resources, perhaps more than your consciousness uses. For example, when the puppeteer for reasons of its own decides to alter a man's testosterone levels based on whether he won or lost a tennis match, making that determination required a considerable degree of perception and knowledge about tennis. That knowledge could not have been innate and hence must have been obtained by "eavesdropping" on the consciousness somehow. As a more impressive example, suppose a lion suddenly appears a few meters in front of you and roars. Your puppeteer instantly sends a FEAR signal to your consciousness and instructs it to stop doing whatever it was doing and concentrate on survival. This all is not what most people would normally describe as a "conscious decision." Now consider the mental processing that was required in order to accomplish this. Images and audio signals had to be processed so that it could be recognized that this was a lion and life-endangering situation. This is a difficult computational feat, but it happens nearly instantly. So evidently the puppeteer has a lot of processing power and it is the one ultimately in control, not your consciousness.

Computer engineers have found themselves forced to devise similar-sounding mechanisms. Although computers are capable of rewriting any memory location or bit stored on disk, most of today's computers have, built into their hardware design, a notion of "operating system" versus "user program" capabilities. *Only* the operating system has full rewriting capabilities, and it delegates *restricted* capabilities to user programs, which can only rewrite *their designated subsets* of memory locations. (If they try to circumvent the restrictions, those attempts will fail.) This is nowadays deemed essential to make computers reliable – granting full power to all user programs would have been too dangerous. Moral: intentional *restriction* of access to computer power is an essential *safety mechanism* in modern computers.

Probably if heart rate and the operations of one's internal organs – including the mental operations inside much of one's brain – had been under full conscious control, that similarly would simply have been too dangerous. Similarly it also would have been too dangerous and causative of disfunctionalty to allow people to have *two or more* consciousnesses. This all is despite the fact that both greater body control and multiconsciousness capabilities plainly would have been advantageous under the right circumstances.

Some speculation about what consciousness is. So I believe that in humans and other mammals, consciousness is a *subroutine*, designed to have *highly restricted computational and control power*, that is launched by, and heavily influenced by, an overarching unconscious "operating and maintenance system" ("homunculus," "puppeteer") which is in substantial part "pre-programmed" at birth. In contrast, it is easily seen by contrasting the information content in your DNA versus the (larger) unpredictable information content you learn over life and the (larger) information content required to describe the unpredictable connection pattern of your neurons, that your intelligence itself must be, at least in considerable part, "self-generated" rather than pre-programmed.

Consciousness is apparently judged to be so dangerous by

[85] Some psychiatrists believe that the unconscious mind reigns totally during sleep (since the conscious mind is not operating) and that in "hypnosis" the conscious mind somehow also becomes downgraded in relative importance. However, some people simply do not become hypnotized and it is hard to know exactly when a person is "hypnotized" or "not." There is no clear definition. Consequently some psychiatrists have argued that anybody deeply involved in reading a book has entered a self-administered hypnotized state, while some skeptics have argued that hypnosis does not even exist.

[86] Famously, endocrinologists claim they can tell which of two men won a tennis match purely by looking at the graphs of their saliva or blood testosterone concentrations versus time [21]. The same is true for players of chess matches, and for the testosterone concentrations in male *fans* of televised soccer matches [41].

[87] If most people attempt to move or not to move their muscles in certain ways that risk damage, they will fail because of built in "protection mechanisms" and "reflexes."

Darwinian evolution that it is not even turned on 24 hours a day; it is periodically shut off with the consciousness itself having only limited input into the timing of that. Dolphins are known to sleep with half of their brains while the other half remains conscious, with the halves switching roles for the net effect of conscious external appearance 100% of the time. This is essential so that the dolphin can continue to breathe. The consciousness of dolphins, being continuous and uninterupted, has thus reached a superior level in its battle for control versus the puppeteer, compared to the level that human consciousness has attained.[88]

The purpose of sleep remains a mystery. I believe that there is no inherent need for it in the sense that intelligent creatures could easily have evolved which did not sleep. Let us refute the usual hypotheses trying to find a purpose for sleep:

1. I do not believe that the purpose of unconsciousness during sleep is simply to reduce energy consumption, because (a) it is still possible to rest while remaining conscious and (b) the power reduction during sleep versus awake resting is small both for the body ($\approx 10\%$) and the brain.[89] The REM-sleeping brain actually appears to consume *greater* power than the awake brain. Non-REM sleep does lessen brain power consumption by 10-30%, but this seems an insignificant energy savings when one considers that sleep is only 1/3 of one's life.

2. I also do not believe that some sort of "toxins" build up in the awake brain which need to be "cleaned up" while sleeping. Nobody has identified any differences in the chemicals produced by the waking and sleeping brain aside from chemicals specifically intended to turn sleep an and off, and many of your neurons (such as those in charge of heartbeat) keep working continually day and night without suffering from toxins.

3. Humans who cannot (due either to brain lesions or to taking certain drugs) REM-sleep do not lose their sanity or their intelligence. Non-REM sleep seems more important, since without it human functionality decreases due to buildup of desire to sleep, but if the "desire to sleep" brain circuitry is dramatically turned down, then this does not appear to result in bad consequences. In 1973 British researchers reported a 70-year-old woman who claimed she slept only an hour a night without daytime naps. In one 72-hour test, during which she was under constant watch, the woman stayed awake 56 hours, then slept only 1.5 hours. Yet she remained alert and in good spirits.

4. Finally, there are fairly intelligent creatures which appear not to sleep. Most fish do seem to sleep (some in cocoons or nests of their own devising and in a sufficiently deep state of unconsciousness that divers can pick them up without response) – but because most fish do not have eyelids and do not exhibit profound brain wave changes during sleep (unlike mammals, birds, and reptiles), and because some fish spend most of their time motionless even when "awake," this is not entirely well defined. Some fish, such as the streamlined sharks, must swim perpetually to keep their gills oxygenated. They do not appear to sleep, at least from the outside. Some other fish such as rockfish and grouper, don't appear to sleep at all. They rest against rocks, bracing themselves with their fins.

The point of examining deep-sea fishes or blind cave-dwellers is that they experience the same environment at all times of day and hence (a) have no environmentally preferred time to sleep and (b) presumably therefore would be expected *not* to sleep unless there were some necessity for sleep or other advantage to be gained from it.

And fish seem quite intelligent; they have been trained to swim through hoops, "play fetch" (retrieving balls for human owners), recognize printed symbols, jump out of the water, and "kick soccer balls into goals"; they observe other fish and imitate them in order to find food sources and recognize predators; they recognize other individual fish and remember past interactions with them; they cooperate in various ways and can act to protect each other; and blind cave dwelling fish have been shown to use sonar-like pressure senses to make mental maps of their surroundings which they use to avoid bumping into walls and other fish; and they react to human-imposed changes in that map by changing their behavior adaptively.

In view of these four refutations, why *do* we spend 1/3 of our lives sleeping? It seems that the only theories still standing are simply that (a) it is an evolutionary accident: the common ancestor of all land vertebrates was some fish that did sleep, and so its descendants continued to sleep; (b) sleeping has the advantage that it keeps animals "out of trouble" by preventing them from wandering around at a time of day or night for which they are comparatively ill-adapted and vulnerable.

It is possible, though, that although sleep is not *necessary*, animals have evolved for 10^8 years of time during which most of them *did* sleep, and hence we should expect that their evolutionary development from that point on took whatever

[88] Perhaps that has something to do with their high intelligence?

[89] "During non-REM sleep, cells in different brain regions do very different things. Most neurons in the brain stem, immediately above the spinal cord, reduce or stop firing, whereas most neurons in the cerebral cortex and adjacent forebrain regions reduce their activity by only a small amount. What changes most dramatically is their overall pattern of activity. During the awake state, a neuron more or less goes about its own individual business. During non-REM sleep, in contrast, adjacent cortical neurons fire synchronously, with a relatively low frequency rhythm [0.5-4 Hz "delta" and 4-7Hz "theta" waves]. (Seemingly paradoxically, this synchronous electrical activity generates higher-voltage brain waves than waking does. Yet just as in an idling automobile, less energy is consumed when the brain idles in this way.) Breathing and heart rate tend to be quite regular during non-REM sleep, and reports of vivid dreams during this state are rare.

A very small group of brain cells (perhaps totaling just 100,000 in humans) at the base of the forebrain is maximally active only during non-REM sleep. These cells have been called sleep-on neurons and appear to be responsible for inducing sleep. The precise signals that activate the sleep-on neurons are not yet completely understood, but increased body heat while an individual is awake clearly activates some of these cells, which may explain the drowsiness that so often accompanies a hot bath or a summer day at the beach.

On the other hand, brain activity during REM sleep resembles that during waking. Brain waves remain at low voltage because neurons are behaving individually. And most brain cells in both the forebrain and brain stem regions are quite active signalling other nerve cells at rates as high as – or higher than – rates seen in the waking state. The brain's overall consumption of energy during REM sleep is also as high as while awake... Specialized cells located in the brain stem, called REM sleep-on cells, become especially active during REM sleep and, in fact, appear to be responsible for generating this state... heart rate and breathing are irregular.

Animals made to go without REM sleep undergo more than the usual amount when finally given the opportunity." [180]

advantage it could of sleep – in some cases perhaps becoming dependent on it. Total sleep deprivation in rats leads to death within 10-20 days (which is less time than they take to die from *food* deprivation) but nothing comparable seems to happen to humans.

Let me give **two examples** from *computer* practice in which something analogous to sleep definitely does yield advantages. The first is taken from experience in computer chess. The "opening books" of some computer chess (or other game) players are capable of "learning from experience." One of the first methods for accomplishing that was implemented by David Slate in a chess program he called "mouse." Imagine that mouse's normal chess playing method is to search all possible lines of play 10 ply deep, computing a heuristic estimate ("evaluation function") of how "good" the chess position is at the end of each such line – and then to make the minimax-optimal move. This strategy is not optimal because the evaluation function is not perfect in most situations (although it is perfect in checkmate or stalemate positions). Now, suppose mouse plays a chess game and (say) loses. After the game is over, it goes through all the positions in the game in *backwards* order, for each performing a search to 12 ply depth (i.e, two ply deeper than normal, which requires about 25 times the normal amount of compute time) and *remembering* the evaluation of every position its search reaches, and *not* re-evaluating any position that it previously evaluated. The result is that mouse, during this postmortem analysis, effectively gets, not 10 ply depth, but in fact 24 ply search depth, along the lines that actually occurred in the game – with 12 ply depth along nongame lines. Now mouse stores a permanent record of its conclusions (i.e. the evaluations of all the deeply analysed positions that arose during this postmortem) which it uses in all future games instead of its usual evaluation routine on those chess positions which fortunately happen to be present inside this permanent store. The result is a "learning" chess program. As Slate demonstrated, this program is capable of learning through experience to avoid falling into opening traps, even ones too deep for its normal search to see (although sometimes several near-repetitions of the trap-experience are necessary before it becomes "convinced").

As Michael Buro later demonstrated with his othello program "logistello," game playing programs equipped with such learning opening books have a tremendous advantage over ones with a static opening book – in any long sequence of games between the two programs, the learning program eventually will *always* win (even if it inherently is so weak that it would normally lose 95% of the time)! Consequently, every top othello program nowadays is equipped with a mechanism of this ilk.

Now regard these gameplaying programs as "conscious" when they are playing a tournament game, but "sleeping" during the postmortem analyses during which they "learn." As we have seen,

1. The "sleep" period greatly improves the performance during the conscious period.
2. If the "sleep" tasks were instead run simultaneously with the "conscious" tasks (which would in principle be possible) then logistello would lose far more tournament games than it does, because (a) opponents could keep on playing the same opening trap against it for more tournament games before that trap "stopped working" and (b) the conscious component would operate at diminished speed because some of its "mental resources" would be "stolen" by the "sleep" component.

As our second example, consider "code optimization." There are many ways, some of which have been automated, to examine the code of a computer program and then "optimize" it by replacing some of that code by different code fragments which yield the same functionality but at higher speed. Sometimes it is not clear which of two code versions will be faster and the only way to know is to run both through a long series of empirical tests. However, doing such optimizations *while that code is running* could be very dangerous and would risk system failure (with the risk being especially high if the modifications are to "hot" parts of the code, which are exactly the parts one most *wants* to optimize), even though doing them *between* runs of the program is harmless, indeed beneficial. Again, in this case dividing everything into "sleeping" (code optimization) and "conscious" (running the code and collecting performance data) phases has a clear beneficial effect.[90]

Important open topic about human intelligence: Can the consciousness affect and alter the puppeteer (and in what ways)?

22 How can we build an intelligence?

Precis. The crucial ingredient of §12's construction of a UACI was a "search over algorithms." We now examine how to try to improve that search to get a "real world" UACI. We both provide many improvement ideas and examine how improved (and unimproved "brute force") UACIs would fare on several example-tasks. This analysis will make it clear both that a brute-force UACI using lexicographic search will be far too slow to be capable of real-world success; and that it is at least *plausible* that improved non-brute UACIs could achieve at least some success.

Despite that fact that in principle, *mathematically* speaking, we now know how to build a simple UACI, developing any *practically useful* artificial intelligence will be a very hard task since our mathematical construction corresponds to a horribly slow algorithm. We now list several approaches to improving performance and then discuss how they might work in practice.

[90] It might be possible to redesign computers to eliminate this obstacle, but (a) we are speaking of them as they currently work, and (b) this redesign might be harder than it looks. Some computers actually *forbid* "self modifying code" or more precisely warn programmers that attempts to use such code will result in unpredictable results because of mysterious interactions with the cache-memory subsystem; other computers allow the OS to totally forbid self-modifying code as a "security feature." (Many "buffer overflow" and "intentional rewrite" tricks designed to breach system security have been based, in essence, on self-modifying code and this all is not possible under an OS that forbids that. The interesting compromise idea of Tsukamoto [209] of allowing self-modifying code *only within a specified subinterval* of the memory controlled by a user-program, has unfortunately not been tried.)

[91] We have already mentioned rigorous speedups of brute force search in §15, but we shall soon see that those by themselves are inadequate to achieve decent performance. Now, however, we shall discuss less-formal and less-rigorous – but probably more practically effective – ideas. Both

Ideas to improve UACI performance.[91]

1. Instead of making the universal UACI algorithm search over *all* polynomial-time algorithms, only search over "non-stupid" ones for which it is not immediately obvious that one can improve them by some standard code optimization trick. Although this would represent an enormous reduction in the size of the search, it still would remain enormous.

It is well known that, often, exponential time searches can be sped up to involve an exponential growth constant smaller than the obvious ones. E.g. Schöning [173] showed how to solve arbitrary N-bit 3-SAT problems in $(4/3)^N \text{poly}(N)$ expected steps instead of the naive 2^N. His algorithm essentially works by choosing a large set of random "initial guess" bit strings, then performing randomized searches over their "local improvements." Similarly, heuristic "local optimization" algorithms that perform a non-exhaustive search over travelling salesman tours empirically find near-optimal tours very quickly. So the analogous idea for us would be to search over "local" changes to an algorithm, which are kept only if they result in improved performance, and then we again try for an improvement. We initially start either with random algorithms, or algorithms from some large initial set of pre-programmed ones considered likely to be promising starting points.

2. Another kind of "local optimization" is *numerical optimization* of undetermined coefficients in some algorithm to optimize some performance measure. (For any algorithm that runs in T real-arithmetic-operation steps to evaluate some performace measure M, it is well known [15] how to evaluate *all* partial derivatives of M with respect to all coefficients inside it (or inputs to it) simultaneously in $O(T)$ time. Standard numerical optimization approaches [48][153] then can use this function and gradient information to seek local optima.)

3. Today's algorithms for rigorously finding the optimum solution to traveling salesman problems empirically exhibit extremely small exponential growth factors (since they have solved instances with over 20,000 cities [101]). They work by "branch and bound" exhaustive search with the aid of extremely good bounds to prune off subsearches which (the bounds prove) cannot contain the optimum tour. The analogous idea for us would be to try to entirely avoid considering algorithms which (something proves) cannot perform as well as the best algorithm found so far.

4. Instead of searching over *all* algorithms, perform non-exhaustive search over just the subset of algorithms considered *promising*. (This may or may not sacrifice provable correctness.) Similarly, one could pre-program a lot of useful algorithm *components* – or have a search for promising-looking components – and then only search over ways to interconnect the components. The point is that most human-written software makes heavy use of various general-purpose algorithm components ("differential equation solver," "linear system solver," "list-manipulation package," etc.) Some other ways to regard this are that (a) we are searching for algorithms but in a "good" programming language, (b) we human programmers aim to "help" the dumb algorithms-searcher by allowing it to take advantage of a certain amount of human knowledge in the form of a database of pre-programmed routines (instead of having to rediscover that for itself) (c) we also may "tell" the UACI something about the nature of the IQ-test problems instead of requiring it to deduce everything about them, and (d) the UACI can build its own "useful algorithm components" tool set and keep track of statistical estimates of "how useful" those components are.

5. Algorithm design manuals often concentrate, not on algorithms per se, but rather on certain promising algorithm-design techniques such as "memorization of a table of answers," "inductive solution of bigger and bigger subproblems," "divide and conquer," "optimization of parameters," "work by analogy," "dynamic programming," "greed," "use of data structures," "linear programming," "backtrack search," "reduction to simpler problems," "recognition as a special case of some more general problem," "pre-sorting," etc. It seems entirely possible to design an algorithms searcher that already knows about these algorithmic structures in the form of pre-programmed "algorithm stencils" and then searches only for the subalgorithms to put inside each "box" of the stencil.[92]

6. The UACI could use the "experimental" technique of seeking correlations between certain bits or numbers available to it, and correctness. If it found them, then it could try to use those bits or quantities preferentially when seeking algorithm improvements. It could also try to seek certain algebraic combinations of quantities avalable to it (optimizing over undetermined coefficients) with the goal of obtaining quantities with higher amounts of such correlation. (Precisely this kind of idea is used in automated "decision-tree learners" [156].)

Thought experiments analysing the UACI in action on real world problems:

Example 1: Inventing the quadratic formula. Suppose we want to find a root x of $Ax^2 + Bx + C$. One answer is

$$x = \frac{-2C}{B + \sqrt{B^2 - 4AC}}. \tag{11}$$

This formula may be written in unary-binary tree form as

```
          *
        / \
  unary-  /
   /     / \
  2     C   +
           / \
          B  sqrt
              |
              -
```

classes of ideas can be used simultaneously.

[92] The whole Chomsky "principles and parameters" framework which is heavily supported by evidence and which dominates modern linguistics is precisely the notion that somehow, the brain has a lot of preprogrammed algorithms or algorithm stencils whose discrete *parameters* are what babies learn in order to understand grammar and syntax – and that therefore only certain kinds of grammars are possible/allowed in human languages. In 1992 Myrna Gopnik [66] deduced the existence of a single dominant "grammar gene," defects in which disrupt grammar by preventing unconscious understanding of making verbs past tense or pluralizing nouns. This deduction was via a study of a single family containing both affected and unaffected individuals. Although the story later was seen not to be nearly as simple as Gopnik & Crago had thought, they were correct that a single gene mutation was responsible for the defect, because the gene was actually located on chromosome 7 by Lai et al in 2001 and named FOXP2 [121].

```
           / \
    square   *
      |     / \
      B    *   C
          / \
         4   A
```

Is this tree small enough that the quadratic formula could be found by brute force search over all formulas of this tree-size or smaller? Not with current computer speeds: the number of unary-binary trees with 7 leafs (labeled A, B, C, 1, 2, 3, or 4), 3 internal unary nodes (labeled sqrt, unary−, or square), and 6 internal binary nodes (labeled *, /, +, or −), is about 10^{17}, i.e. beyond reach in a CPU-year. The cubic and quartic formulas would be far more vastly out of reach.

On the other hand, given a large number of (A, B, C, x) numerical 4-tuples, a computer could easily recognize that all the 4-tuples satisfy $Ax^2 + Bx + C = 0$ to high accuracy (thus mechanically "understanding what the problem is") by **searching for linear relations** among the cubic monomials of (A, B, C, x). That can be accomplished by Gaussian elimination. Then it could call upon some commercial symbolic manipulation program to spit out a symbolic solution of this equation (and the same for cubics and quartics). This is an excellent example of where brute force search fails but a cleverer non-exhaustive search, searching only among "promising" algorithms, quickly succeeds.

Example 2: Learning to solve "Rubik's cube" puzzle (4.3×10^{19} configurations). To make our job easier, we shall assume that it is *known* what the goal is, that "move" primitives to perform the 6 fundamental cube face-turning operations (customarily denoted U,D,R,L,F, and B [11][185]) and their inverses (U^{-1}, D^{-1} etc.) and squares (U^2, D^2 etc.) are already available, and finally that "inspect" primitives for determining any desired among the 54 cubie-face colors in the current cube-configuration also are already available. (This approximates the situation faced by a human novice solver.)

Then here is one approach which will learn to solve the Rubik cube reasonably quickly.

Build catalog of useful algorithm components: Search for short sequences S (1-4 moves long), and T (1-3 moves), such that either S^nT (S iterated n times followed by T) for some value of n with $1 \le n \le 12$, or $S^{-n}T^{-1}S^nT$ or $S^{-n}TS^n$ or $T^{-1}S^nT$, if applied to a pre-solved cube, will alter 1 to 4 of the 26 surface cubies. Catalog the resulting low-alteration sequences. (This step requires exploring about 10^7 sequences, although 48 times fewer if the symmetries of the cube are taken into account. The catalog will contain about 10^4 sequences.)

Solve a given scrambled cube: While the cube remains unsolved, find an operation of form QCQ^{-1} or QM that reduces the number of wrong cubie faces, where C is a catalog operation, M is a single move, and Q is a move sequence 0, 1, or 2 moves long. (E.g. choose such an operation at random with Q of minimal length, with a bias favoring shorter sequences.) If one exists, do it; otherwise do random moves to rescramble the cube.

The catalog-building step is the "learning" step and may be done only once. Thenceforth the "solving" step will run reasonably quickly on randomized cubes.[93]

I believe the above approach approximates the one used by most of the human novices who succeed in solving the cube – i.e. it is basically the same idea, but has rather more "brute force" and less "intentional design" character than what humans do.

A second approach to solving Rubik's cube. Although the preceding approach is similar to what most human solvers do, it will not get very close to "God's algorithm" (the minimimum possible number of face turns to descramble the cube). And, of course, we want to build an AI superior to humans. The best cube-solving methods that humanity so far have been able to create are *not* based on catalogues of clever move-sequences; they instead all are based on Thistlethwaite's "nested subgroups" approach.

generators	cardinality	ratio	depth	how to recognize
L, R, F, B, U, D	$8!3^7 12!2^{11}$	2048	7	all reachable configurations
L, R, F, B, U^2, D^2	$8!3^7 12!/2$	1082565	10	all 12 edge-orientations correct
L, R, F^2, B^2, U^2, D^2	$8!^2 4!/2$	29400	13	...& all 8 corner-orientations correct & LR-midslice membership correct
$L^2, R^2, F^2, B^2, U^2, D^2$	$4!^4 2$	663552	15	...& all 3 midslices have correct memberships & edge-cubie permutation's parity is even & only 96 corner-cubie permutations possible
1	1	1	0	fully-solved position

Figure 22.1. Thistlethwaite's 4-stage nested subgroup algorithm for solving Rubik's cube. ▲

In this approach, we devise a sequence of nested subgroups of the full Rubik cube group (restricting to a subset of the full 18-element set of faceturn moves generates only a subgroup). At each cube descrambling stage, we search for a minimal-length sequence of moves that will get us into the next smaller subgroup. Thistlethwaite's original algorithm (figure 22.1) involved 4 stages. The stages respectively required at most 7, 10, 13 and 15 face turns (these bounds were proved by exhaustive searches of the respective coset spaces, which was feasible to do with computer aid since the largest such space had only 1082565 elements) thus proving the cube could be solved in at most 45 face turns.

[93] Although it will require more moves than "God's algorithm," it usually is within a factor of 10. Solvers requiring fewer moves and/or less thinking may be built later by e.g. constructing a list of (precondition, cube-move-sequence) pairs, where the solve algorithm is to apply the first move sequence whose precondition is satisfied, and then tuning the list.

Thistlethwaite's algorithm has been implemented independently by several people. Each stage is done by a search method such as "IDA*" [169].[94] The resulting codes (on 2006-era computer hardware) solve randomly scrambled cubes in typically about 100 milliseconds and about 30 face turns. Thistlethwaite's particular 4-subgroup sequence is an especially good choice because all the ratios are small and because all his subgroups have large generator sets, enabling small numbers of moves using those generators (large subgroups with only a small number of generators, such as $\langle U, R\rangle$, would have been unwise).

Later Kociemba got rid of stages 1 and 3 of Thistlethwaite's algorithm (i.e. combined stage 1&2 and 3&4), resulting in a 2-stage algorithm, which (it was proved) requires at most $30 = 12 + 18$ face turns. Kociemba made his algorithm not rest once it found an optimal solution for the first phase, instead also investigating other first-phase solutions (including suboptimal ones). Kociemba's resulting program solves a random cube in under 20 face turns in about 1/3 second. This is nearing God's algorithm since cube configurations that require 20 face turns are known.

Now in order to invent nested-subgroup-type algorithms *automatically*, we need

1. A preprogrammed searcher such as IDA*
2. Fast ways to recognize when we have entered a given subgroup.

We now explain how it is possible to invent subgroups, and to invent fast ways to recognize when a cube configuration is in a given subgroup, totally mechanically. To explore a subgroup generated by some generator subset S (and there are only 2^{18} subsets of the 18 Rubik generators) simply make random S-moves. Now, using linear algebra, search for linear combinations of (color, location)-features of the resulting cube configurations which assume *constant* values (but which assume nonconstant values in the full Rubik group). Those linear combinations are the red flag that indicates subgroup membership. As you can see from figure 22.1, all Thistlethwaite's subgroups can be recognized from such features as "sum, over all edge-cubies, of orientation-correctness, equals 12." So by simply searching for constant-valued linear combinations of cubic monomials (just like in example 1 – linear-relation finding evidently is a very useful technique) most necessary "red flags" would be found – and in particular, all of the ones needed for Kociemba's algorithm. Unfortunately this would be a very large linear algebra problem.

Example 3: Inventing matrix algorithms using some "stencils" and/or "prebuilt algorithm components." A remarkable variety of useful matrix algorithms can be built using a small toolkit called the BLAS of "basic linear algebra subprograms," embedded inside a small variety of prebuilt algorithm "stencils" defining various acceptable/interesting loop structures, such as "for $i = 1$ to N, for $j = i$ to N, for $k = j$ to N."

Suppose we already know how to find the singular value decomposition (SVD) of a 2×2 matrix, i.e. this small prebuilt algorithm component is available. Suppose we now want to do various interesting things with arbitrary $a \times b$ matrices M, such as find their SVD, eigenvalues, determinants, pseudo-inverses, rank, or solve linear systems. Well, all you need to do is to repeatedly run SVD on a random 2×2 submatrix (rows j and k and columns j and k with j, k distinct random) of M, converting M to $U^{-1}MV$ where U and V operate as 2×2 orthogonal matrices. The result (after doing a large number of these) is that M becomes transformed to its diagonal matrix D of singular values, while the product of the Us gives M's left singular vectors and the product of the Vs gives M's right singular vectors. With the SVD known it is trivial to compute the determinant (product of the singular values), rank (number of nonzero singular values), pseudo-inverse in SVD-form (reciprocate all nonzero singular values to get D^{-}). It also is easy to solve a linear system $M\vec{x} = \vec{y}$ by computing $\vec{x} = V^T D^{-} U \vec{y}$ where $M = U^{-1}DV$ and $V^{-1} = V^T$ since V is orthogonal – for this all we need is a matrix-vector multiplication routine, available inside BLAS.

Similarly, if we already had a prebuilt routine to find the Schur decomposition a 2×2 matrix M (that is, to find a factorization $M = Q^{-1}RQ$ where R is upper triangular and Q is unitary and all letters here are 2×2 matrices) then we would be able to find the eigenvalues of an $n \times n$ matrix by repeatedly applying the Q-conjugacy transformations got from the 2×2-size Qs arising from Schuring random 2×2 submatrices of M (i.e. rows j and k and columns j and k with j, k distinct random).

The resulting routines would only be a constant or log factor slower than LINPACK's superbly designed matrix algorithms for the same tasks to reach 10-decimal accuracy, but would be simple enough that they arguably could have been invented by quite dumb brute force searches.

I further would argue that most of the algorithms in LINPACK could basically be invented by (1) *a dumb search* that searches over all matrix algorithms that "look like" the algorithms in [65], i.e. which use the same kinds of loop structure and in which we "fill in the empty boxes" in the stencil with standard generalized forms of typical formulas and with undetermined coefficients inside them and then later (2) *use a general purpose numerical optimizer* to solve for the coefficient-values needed to make each trial algorithm "work best" on a large set of examples.

This kind of approach may seem silly and does not approximate the design approaches originally used by the human inventors of most matrix algorithms. But it would have succeeded in finding some algorithms which humans were unable to find for decades. For example, Gram-Schmidt matrix orthogonalization has a numerically far-superior, but much less obvious, form called "modified Gram Schmidt" which was only discovered far later by Björck [65]. Gram and Schmidt themselves did not notice this improved form of their algorithm, but it would be just as easy for the computer to find as the unmodified Gram-Schmidt approach.

Example 4: Neural net classifiers. Neural nets with "backprop learning" and "forward prop classification" [167][75] are known to work quite well for the purpose of recognizing pixel images of handwritten digits, but this was only accomplished with considerable care, also known as "black art," to design a good kind of neural net. Assuming a general purpose

[94] IDA* stands for "Iterative Deepening A* search."

neural net backprop and forward prop code is assumed already pre-programmed as part of our "toolkit," nothing stops a mechanical brute force searcher from investigating numerous kinds of neural net designs to see which works best – and I believe this approach could generate comparably good or better neural nets than those produced by human designers.

Example 5: Inventing efficient sorting algorithms. When I was a child first learning about computers, it seemed "obvious" to me that the best way to sort N numbers was to repeatedly remove the minimum number, leading to an obvious $O(N^2)$-step ("step" meaning comparison or movement) algorithm. It therefore made a considerable impression on me when I first heard about $O(N \log N)$-step sorting algorithms.

How could such an algorithm be invented automatically by a "dumb" searcher? Assume we already know the goal in the sense that we only consider algorithms which *permute* the N numbers in such a way that some externally supplied "sortedness checker" agrees the numbers are now sorted. (To give the searcher less foreknowledge, we could also externally supply a checker that some list of numbers really is a *permutation* of some other list.) Suppose the searcher already has in its bag of preprogrammed tricks, the "divide and conquer" algorithm paradigm. So we agree only to search over algorithms of the form "divide the N numbers into two bags of $\lfloor N/2 \rfloor$ and $\lceil N/2 \rceil$ numbers, sort them recursively, and then somehow merge the results." Given that this is the case, we really only need to search over the algorithm *components* needed to "fill in the empty boxes" in this stencil, namely the base-problem of sorting a list containing exactly one or zero numbers, and the merge-problem of merging two lists. Such algorithm-components ought to be fairly easy to learn provided we know that we can use the external checker to check each individually, and especially if our search for merge-algorithms also knows about fundamental list-maintenance operations and about the "inductive solution of bigger and bigger subproblems" algorithm-design technique and is willing to apply the external checkers on each merging subproblem (i.e. merging only the first k items in each list for $k = 1, 2, 3, \dots$).

So it seems possible to make a UACI capable of inventing an $O(N \log N)$-step merge sort algorithm for itself. Then via local improvement experiments it even seems possible for it eventually to convert the result into fairly *efficient* code, especially if our UACI also knows about "standard code-optimzation tricks of the kind inside many present-day optimizing compilers."

Example 6: Inventing a multiprecision adder. Assume our UACI knows there are *two* N-bit input numbers A, B and one $(N+1)$-bit output number C and that a checker that $C = A+B$ is available (externally supplied). A UACI searching divide-and-conquer algorithms would divide one or both of the input binary numbers into two halves (the most significant bits and the least significant bits). It might find out by experiments with "easy" problems that $AB = BA$ and hence postulate that it was a good idea to search only over algorithms that treated the inputs symmetrically. It would then try to divide and conquer by dividing both numbers in half and somehow combining the results of the four recursively solved subproblems (A_j, B_k) for $j, k \in \{0, 1\}$. The UACI's task would then merely be to invent the "base" (adding a 1-bit-long number by another) and the "merge" algorithm components.

Base: If the task is to input two 1-bit numbers and output their 2-bit sum, the UACI could quickly discover a valid algorithm by brute force (simply remember every answer to every problem). If the task is to add a 1-bit number to an N-bit number (which arises in the divide & conquer framework in which only one of the inputs is being subdivided) then the UACI if it knew about the "inductive solution of bigger and bigger subproblems" algorithm-design technique and is willing to apply the external checkers on each base subproblem (i.e. adding only the last k bits of the N-bit number to the 1-bit number for $k = 1, 2, 3, \dots$) ought to be able to succeed.

Merge: For adding with the both-subdivided divide & conquer paradigm, the "merge" consists merely of outputting $A_0 + B_0$, then outputting $A_1 + B_1$. This seems very feasible to discover, because both of these quantities are already pre-available to the merger. This merge will succeed 50% of the time and ought to be learned quickly. But the other 50% of the time, we need to use a more complicated merge: if $A_0 + B_0$'s most significant bit is 1, then the second output instead is all bits but the least-significant of $A_1 + B_1 + 1$. Assuming we already have an "add 1" subalgorithm (invented as part of the "base") then this improvement to reach 100% accuracy too seems feasible to discover, and the fact that the original method's answer always is wrong if $A_0 + B_0$'s most significant bit is 1 but that all other available bits seem much less correlated, could have been discovered and used to motivate a fix of the form "if(bit=1) then ...".

Inventing an adder a different way – with a different preprogrammed algorithm stencil. If the following algorithm stencil were available (where $a[1..N]$ and $b[1..N]$ were the bits of the two N-bit input numbers)

```
for j=1 to N {
   x = F(a[j], b[j], x)
   output x
}
```

then an adding algorithm could be invented quickly. There are only 256 possible functions F mapping 3 bits to 1 bit, so all could be explored, with the result that a binary addition algorithm would be discovered except for computing the final bit of the output – which, it would then quickly be discovered, was always equal to x. This same algorithm stencil also will do *subtraction* and *comparison* of two N-bit numbers; it also will work for addition, subtraction, and comparison of numbers in nonbinary radices, and it will do various bitwise combinations (such as ANDing and XORing) of two binary N-bit words, and, e.g, vector addition. Hence it would appear fully justified to put this algorithm stencil in an AI's toolbox.

Example 7: Inventing quadratic- and subquadratic-time multiplication algorithms. Also as a child I wrote a program to multiply N-digit numbers, using a "schoolboy" method, in $O(N^2)$ steps, and consequently it again made an impression on me when I learned there are subquadratic algorithms.

Assume our UACI knows there are *two* N-bit input numbers A, B and one $2N$-bit output number C and that a checker that $C = AB$ is available (externally supplied). A UACI searching divide-and-conquer algorithms would divide the input binary numbers into two halves (the most significant bits and the least significant bits). It might try without success to get a subquadratic algorithm that only divided up one of the numbers, but might succeed in getting a quadratic-time algorithm of this kind (by similar methods to what we shall discusss below). It might find out by experiments with "easy" problems that $AB = BA$ and hence postulate that it was a good idea to search only over algorithms that treated the inputs symmetrically. It would then try to divide and conquer by dividing both numbers in half and somehow combining the results of the four recursively solved subproblems (A_j, B_k) for $j, k \in \{0, 1\}$. The UACI's task would then merely be to invent the "base-solve" (multiplying a 1-bit-long number by another; this seems feasible to do!) and the "merge" algorithm components.

Now *assuming* a N-bit adder was already available as a preprogrammed algorithm component, the UACI might try to combine pair-products by adding them in all possible combinations, and if so would soon discover a combination that yielded a merge algorithm that worked more than 10% of the time. The failures would again be seen to be highly correlated to certain bits which would motivate fixes which hopefully would discover the need for "carry" bits and if an "add a 1-bit number" preprogrammed component were available (see the previous example) then it might then discover the fixes necessary to improve to 100% correctness. It would then have discovered an $O(N^2)$-step N-digit multiplication algorithm.

If an adder/subtractor were available, then the UACI might investigate all short ways to employ sums and differences of the A_j and B_k as inputs to the recursive half-length multiplier, with the specific goal in mind of doing only *three*, not four, calls to it and thus obtaining an $O(N^{\lg 3})$-step *sub*quadratic algorithm. Such an algorithm exists – "Karatsuba's algorithm"

$$(A_1 2^N + A_0)(B_1 2^N + B_0) = \tag{12}$$

$$A_1 B_1 4^N + [(A_0 + A_1)(B_0 + B_1) - A_0 B_0 - A_1 B_1] 2^N + A_0 B_0$$

Although it would be very difficult for a UACI to discover the Karatsuba merge formula in toto, it is plausible to imagine that it could discover simplified forms of it valid only when no carries occur – and these merges immediately would be useable as a speedup to the $O(N^2)$-time algorithm that could be called only if certain carry bits were 0, and which would reduce its *average* runtime to subquadratic: The UACI could then try to find altered forms valid in various carrying events, ultimately hopefully discovering Karatsuba's algorithm in full.

It would help if a substantial fraction of the sample problems were "easy" ones with $A_1 = B_1 = 0$ or $A_0 = B_0 = 0$.

Example 8: Triangulation of a polygon. Chazelle [31] discovered a linear-time algorithm to find a triangulation of a given simple N-gon in the plane by means of $N - 2$ diagonals. It would seem extremely difficult to make an AI duplicate that discovery!

Verdict. In view of the above examples, it is clear that the simplest, "completely brute force" sort of UACI is far too inefficient to be useful, *but* it also does not seem ludicrous that a UACI of our general sort might be possible that, thanks to a lot of pre-supplied algorithm design stencils, components, and tools, could reach a fair amount of competence in practice. It also seems entirely possible that such a UACI could continually be improved for an almost arbitrarily long number of years. However, the development time to reach high competence might be very large indeed.

Eric Baum made a similar point in his book [12]. He argued that what matters is not duplicating the hardware power of the human brain – as we saw in §2 that is pretty much already accomplished – but rather, duplicating the *enormous* amount of Darwinian evolutionary "computation" that went into *developing the algorithm* that our brains run. Specifically, we estimate[95] that the total number of lifeforms born during Earth history is 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. Even assuming (probably extremely optimistically?) that each's contribution to evolution could be effectively simulated by performing an average of 10^6 computer instructions, the total amount of computation needed to duplicate evolution would be of order 10^{48}-10^{53} instructions, which on a 1 GHz sequential machine

[95] I got this estimate as follows. Two genera of plankton, both often erroneously called "blue-green algae" but which in fact are cyanobacteria or close relatives, > 95%-dominate [149][159] the count of cells per liter in seawater: *synecococcus* and *prochlorococcus*. The former, which are somewhat larger and equipped with flagellae for locomotion, are ubiquitous in all marine bodies and occur in concentrations of (0.01 to 1.3) $\times$ 10^9cells/cm^3. The latter, which is the smallest known (0.6μm) and most abundant photosynthesizing lifeform, mainly only occurs at latitudes more tropical than 40° but is by far the most abundant there, occurring at concentrations of (1 to 4) $\times$ 10^9cells/cm^3 throughout the mid-ocean (once erroneously thought to be a biological desert). Plankton occur at depths 0-200 meters and exhibit a concentration maximum at depths of 50-75 meters (where they have 3-10 times the surface concentration) perhaps due to a nitrite maximum there. Prochlorococci have been estimated to provide the world with 30-80% of oceanic O_2 production, while NASA has estimated 80-90% of worldwide O_2 production is oceanic. Soil bacteria are comparably concentrated ($\approx 10^9$cells/gram) in fertile soils but 1000 times less concentrated in poorer soils [227], but soils are of much thinner depth than the oceans and land is only 29% of Earth surface area – so freeliving cell counts on land are comparatively negligible. Hence just counting these plankton alone should get us within a factor of 2 of the count of all lifeforms. Taking account of the surface area of the Earth, we estimate the total count of these plankton to be 7×10^{31} *cells to within a factor of 2* and the total count of of all lifeforms to be within a factor of 2 on the low side and 4 on the high side of this. These plankton are observed to replicate typically once a day (usually late in the day) in Pacific atoll lagoons, but probably multiply more slowly in the open ocean (less nutrients). Observations published by Corlett in 1953 in the Northern Atlantic (60°N) showed seasonal plankton population blooms by a factor of 2000 between mid-April and mid-June, so at least 11 doublings occur in 60 days at this time. It seems plausible that multiplication rates at least half that occur all year in the tropical latitudes favored by *prochlorococcus*, hence we estimate that there is on average *one replication every 1-10 days.* Bacteria are believed to be 3.5Gyear old and the oxygenated atmosphere cyanobacteria produced caused a massive extinction about 2.2Gyr ago and the appearance of most of today's iron ore deposits. Assuming these rates and populations have persisted for 2.2-3.5Gyr, multiplication yields an estimate that *the total number of lifeforms during Earth history has been between* 2.8×10^{42} *and* 3.6×10^{44}. If we also count bacteriophages infecting these plankton as "life," then this count could significantly increase. Measurements indicate [222] that 6-12% of cyanobacteria die by phage lysis, and in plankton blooms this increases to 34-52%. So assume that 25% of all these plankton historically died by lysis caused by phage infection. (In other words, 50% of all deaths were caused by phage and 50% by other causes.) Further assume that each lysis produces 60 to 1000 new phage particles. In that case including phages would increase our count by a factor of 16 to 251. (Baum [12] instead estimated "10^{30} to 10^{40} including viruses" on his page 445 and described his derivation, but I think Baum's estimate is of inferior quality.)

would take over 10^{31} years – and even devoting all computers currently available on the planet to the task, we are still taking well over 10^{20} years – dwarfing the age of the universe at a mere 10^{10} years. This is truly a vast computation.[96]

However, Darwinian evolution is (probably) a poor way to develop an intelligence, and humans are (certainly) a poor kind of intelligence. Hence AI researchers can try to console themselves with the thought that 10^{40} years this is (probably) merely a quite weak *upper bound* on the computing that would be required by some better search technique to build some much superior brand of intelligence.

And even this estimate fully suffices to make our point that UACIs based on search techniques better than brute force lexicographic search, should be *far* better than brute force ones. Specifically, I estimate[97] that a brute force UACI first will exceed human capabilities at essentially everything, only after an initial runtime *delay* of somewhere between 10^{999} and $10^{99999999}$ computer-years.

23 Hold contests! (and why that will work)

Precis. We explain how the AI field can and now should adopt a modus operandi based on standardized annual intelligence contests. This would ensure measurable progress each year toward a practically useful artificial intelligence. The historical parallel to computer chess is cited as clear evidence that such a modus operandi will be both necessary for success, and (perhaps) also sufficient.

Reread §11 to remind yourself of what the main subgoals of AI should be. We can accomplish goal ② by holding annual "intelligence contests" with the result that measurable AI progress – increased measured intelligence – *will occur every year.*

The idea of holding annual computer chess tournaments has already been tried, and it *did* ratchet annual progress, i.e. chess "ratings" approximately monotonically increased. This was probably *essential* for the eventual triumph of the computers over the top human chessplayers. It took about 50 years, but it happened. These contests also were probably essential for keeping the computer chess area sane and objectively judgeable – as opposed to most of AI so far.

Actually, the need for formal annual computer chess contests in specific *locations* at specific *times* no longer really exists thanks to the rise of the internet and "chess server" software, and the same would presumably now be true for the intelligence contests – it suffices to have a web site containing both a database of intelligence testers with standard interfaces, and tables of current performance records for each. Developing a good AI by successive improvements starting from the easy-to-program UACIs of the sort we can immediately build now, may also take 50 years – actually I have no idea how long it will take – but at least this way we will have a scorecard to keep track of where we are, what ideas work, and how quickly progress is occurring.

The accumulating contribution of the intelligence testers to this web site will itself be a significant contribution since in some sense they are the "training data" from which we must learn. Both intelligence tests *and* testees will be contributed annually. For example if 20000 intelligence tests get contributed to the site and a program whose total code-length is equivalent to only 2000 of them manages to do well on 17000 tests, then probably the quest to develop an AI could be said to have very substantially succeeded. I think[98] it will take an effort of roughly this magnitude.

It would also be possible for humans to enter the intelligence contests – they need not be solely open to computers.[99] As one impressive example of human capabilities on §9's sort of test, consider the Swedish mathematician Arne Beurling (1905-1986). Near the beginning of World War II, the Swedes were very worried about threats both from Nazi Germany and also from Stalin's Russia, the two of which had (they knew) signed a secret pact. After the Germans conquered Norway in April 1940, they often sent telegrams between Norway and Germany via lines passing through Sweden. The Swedes tapped the lines, but the Germans nevertheless felt secure because the telegrams had been "unbreakably" encrypted with the aid of "Geheimfernschreiber" (Siemens Corp. cryptographic teletype) machines. In a mere two weeks during June 1940, Beurling single-handedly deciphered and reverse-engineered an early version of the Geheimfernschreiber, doing it *without access to* and *without knowledge of the principles of operation of* any actual machine – purely by examining raw ciphertext (without plaintext) bit string intercepts. This enabled the Swedes to learn ahead of time about Nazi "Operation Barbarossa" plans to invade Russia.

24 About previous work, especially by Hutter

Precis. Our (2006) "mathematical definition of intelligence" (MDoI) and "UACI" discoveries both can be regarded as *re*discoveries of ideas by Marcus Hutter [79] during 2000-2006 but which he had expressed, explained, and motivated somewhat differently. We survey the relations and differences between our and Hutter's work.

Both Hutter's and this development exhibit some striking similarities, but we had both different attitudes and different terminology and in some cases investigated different topics or reached differing conclusions.

[96] Even if, say, the entire surface of the planet Mercury were to be covered with 7×10^{14} solar-powered computers all working on this (ten 1 GHz computers per square meter) they would only execute 2×10^{31} instructions per year, a rate still hugely insufficient to do 10^{48} instructions in the age of the universe.

[97] Cf. footnote **??**.

[98] This guess is based on some experience with computerized learners...

[99] It has been suggested that binary format is "biased" pro-computer and anti-human. That does not particularly bother me, but to decrease human time waste and increase understanding I would have no objection if human test takers got their problems in a more congenial format when possible, such as pictures and alphanumeric symbols.

Hutter	Smith
AIXI framework for optimal interaction policy of an agent with its environment	Mathematical Definition of Intelligence (MDoI)
Environment	Intelligence tester (PG & SC)
Incremental algorithms	Reent-algorithms
"agent"	"Entity under test" (ET)

Figure 24.1. Hutter vs. Smith terminology comparison. ▲

Hutter's AIXI framework indeed is *more general* than ours (ours is essentially the special case that arises from restricting some of Hutter's algorithms to be in P or NP and forcing the environment be what Hutter calls "passive") and the ideas he needed to reach that extra generality are more sophisticated than ours. Hutter's work indeed makes it clear that there is a *continuum* of successively more powerful kinds of intelligence, and that our MDoI and UACI (roughly) really constitute just the *lowest point* on that continuum. However, we shall argue that Hutter's extra generality and sophistication are *undesirable* for several reasons, the most important being that they make the problem of developing an AI much more difficult while yielding very little compensating benefit – even the lowest point on his continuum seems enough intelligence for anybody. For most uses, we advocate jettisoning them.

Three ways in which our development is superior to Hutter's are the following. Hutter later claimed [103] to have developed a "universal intelligence test" – as opposed to our notion of an infinite number of different intelligence tests but one UACI that is good at all of them – but we argue that this extension was a mistake because there actually is *no such thing as a universal intelligence test...* although as we shall also see, the question is a bit subtle. Hutter's publications did not investigate the experimental psychology literature and so all the confirmatory evidence for the HUH that we dug up from that literature in §16-20 is new. And Hutter did not invent the "faster than brute force" search in our §15.

Is our "definition of intelligence" new? Yes and no. I do not think anybody before myself and Hutter had formally stated "a mathematical definition of intelligence," but all the key ingredient idea-fragments were available to some degree within previous literature.[100] There is not a great deal to it, and further, one could say that there was even *less* to Turing's original proposal of *his* "IQ test." The present paper could in principle have been written in the early 1970s as soon as Cook and Levin's NP-completeness work became available. It is quite surprising to me that Turing himself did not invent both our definition of inteligence and the UACI; my best guess (although still it seems inadequate) for why he did not is that Turing died before Cook and Levin's NP work.

Nevertheless, I feel that our definition (§9) and consequent theorems (§12) and discussion (§22) has synthesized a previously unavailable level of total clarity, making it now for the first time reasonably clear both:

1. what an intelligence is,
2. showing how to build a theoretically good but practically useless AI,
3. showing many good reasons (§16-20) to believe human intelligence works in this manner,
4. and laying out a research program of how the field should proceed to try to engineer a practically useful AI.

There are **three**[101] **previous lines of research** which led up to the present work, which I will describe simply by listing the names of some of the most prominent contributors to that line of research ("prominent" in the sense that they either influenced me the most, or should have):

1. Herb Simon, Eric Baum, Igor Durdanovic: AI and thoughts about how to build one,
2. Jean Piaget: studies of development of human intelligence in children, and Arthur R. Jensen's[102] studies and expositions related to g.
3. Alan M. Turing, Ray Solomonoff, Leonid Levin, Marcus Hutter: universal algorithms techniques.

Actually I was unaware of the lattermost three researchers until my ideas had almost completely crystallized, and hence I rediscovered most of their ideas.

Solomonoff has been working on the same general line since the 1960s and his ideas evolved to be close to my own, whereupon they were further extended and developed by Marcus Hutter, who in 2004 described them in a book [79].

It was with mixed feelings that I discovered Solomonoff & Hutter's existence! However, the fact that I did all this independently of them until right at the end, has its advantages

[100] See footnote 19. We also mention that the "neural net backprop learning" literature was specifically intended to be applicable to a very wide class of problems, and Baum & Durdanovic in lectures on their "Hayek" artificial-economy approach to building an AI, also had the idea that the "problems" their system would solve could be posed in an extremely general fashion, without the solver "knowing what the problem was." In the actual system they built, though, they fell rather short of that ideal because the solver in fact was given a considerable amount of pre-programmed knowledge about the problem format and pre-programmed tools for manipulation designed specifically to seem relevant to that particular kind of problem – it was not just "here is a bitstring and the only preprogrammed tools you have are the ability to perform a few bit manipulations." These examples of previous thoughts in our directions are probably, at least if we are sufficiently generous in our interpretations, only the tip of the iceberg.

[101] Actually four, if you also count work on "competitive algorithms" [22], and five if you count the development of computational complexity and NP theory [53][129][148][187][61][6].

[102] If Huxley was "Darwin's bulldog" then Jensen might be called "Spearman's bulldog" since he has devoted his career to the confirmation and exploration of Spearman's g, apparently to a greater extent than any other person. However, this characterization is oversimplified because actually Jensen's work is more extensive and better than Spearman's, and also because not all of Jensen's work has been unidirectional, e.g. he was one of those who pointed out the anomalies in the work of Cyril Burt that later led to his exposure as a fraud. I have in the present work subjected both Jensen and the psychometric and g fields more generally to considerable and well-deserved criticism, and also I feel that Jensen has to some extent become an ideologue for the pro-g position (although not to as great an extent as Gould [68] and Kamin [90] have been ideologues against it) and that a less biased introduction to that area would be Deary [44][45] or Mackintosh [120]. But nevertheless I think Jensen would be pleased to learn that the present work's definition of intelligence leads to a hypothesis that explains and predicts the existence of g, which in turn hopefully will lead both to more understanding of it and which will suggest good directions in which to focus further g research; and also that computational complexity theory has been able to produce a definition of intelligence at all, despite Jensen's earlier description of that as "proved to be a hopeless quest."

because it caused us to work from rather different angles and with different emphases, as well as providing an "independent validity check" on each other's ideas. I believe, in fact, that Solomonoff and Hutter have, in some technical ways, the wrong ideas. (I hope they will interact with me and correct my ideas and impressions – which almost certainly are currently imperfect – too.) In the below discussion let me just focus on our differences.

Solomonoff's closest approach to our framework is something he calls "operator induction using algorithmic probability." The goal of that is, based on a sequence of question-answer pairs $(Q_k, A_k)_{k=1..n-1}$ to predict the next answer A_n given the next question Q_n. His proposed approach to accomplishing that is based on some search-over-all-algorithms ideas. So as you can see this is quite similar to our "intelligence test" framework in §9 as well as to our UACI construction in §12. The points where I believe Solomonoff has gone wrong are the following:

1. I believe it is not right to focus on *all* algorithms rather than just NP questions and polynomial time algorithms. I argue in the present work that NP is (aside from technical issues about randomness, see §14) actually both (a) adequate to build an "intelligence"' and (b) the only way (and most general possible way) to make the IQ test efficient enough so that it can really happen and be worth discussing. And this is good because it makes everything easier.

2. I believe intelligence and intelligence tests are not about algorithms, they are about "reent-algorithms," so one has to redo the Solomonoff framework with that in mind. (However this is not a crucial objection because an ordinary algorithm that simply redoes all the previous work every time can be made to simulate a reent-algorithm at a cost of roughly n times more.) Hutter also realized this under the name "incremental algorithms."

3. I believe it is wrong to assume there is a "unique answer" A_k which is known. Instead, I believe really intelligences deal with problems without unique answers, which are not necessarily known, and may only receive environmental feedback via a score function (utility) which indeed is a *quickly-evaluable* score function (i.e. in P).

4. Solomonoff in a recent working paper notes that he has been "working on a system" since 1986 (20 years!) to try to implement his UACI-like approach and build a decent AI. But if I understand his system right, I claim it won't work. That is because his system is essentially similar to the "brute force" UACI that we discuss in §22 and conclude will not be capable even of inventing the quadratic formula. Solomonoff has the idea that the search-over-all-algorithms should attempt to produce "weights" for each algorithm which should tend to get larger for the "more useful" algorithms – and that is an excellent sort of idea in rather germinal form (I fully agree the searcher will need to collect statistics in some fashion about what works better and more often, and then use those statistics to search in a stronger fashion – my own 'BPIP search' in the realm of computer game playing was another attempt in that direction, and we have mentioned evidence in §19 that human memory does something of that ilk) but by itself I cannot believe there is any way that is going to be enough to lift Solomonoff's system out of the realm of total incompetence.[103] Furthermore, I suspect there is no way that Solomonoff, or any other solo human for that matter, will be able, even with 20 years of work, to build a decent AI. I think it is going to take an entire competing and cooperating community of researchers working for ~ 50 years and kept sane as they go by means of the modus operandi in §23. My historical model is the ultimate success of computer chess, which *did* take 50 years starting from the work by its first visionaries (whom we can consider analogous to Solomonoff, Hutter, and I) to succeed in beating Kasparov, and which *was* kept sane by a mechanism similar to the one I'm proposing. Building a good quality computer chessplayer was a far easier task than building a good quality UACI will be, but we have the advantage of starting from a more advanced place.

Hutter's work [79] is really extremely conceptually similar to ours, and Hutter also had the advantage of being aware both of Solomonoff and of Schmidhuber. I finally obtained a copy of Hutter's book on 23 May 2006, and found that Hutter's work was both more general, deeper, longer, and more erudite than ours here. On the downside, though, it is significantly harder to digest, and Hutter's extra generality actually seems undesirable. (Also, the overlap with our work is not complete – we cover several topics Hutter ignores.) Hutter's book represents some deep and profound thinking, so it is not an easy matter to reach a full appreciation of, or to produce a critique of, it. (Doing so might take months.) It certainly deserves both, but we will only do an imperfect job.

Hutter's AIXI framework is more general than ours in two ways:

1. Hutter allows general algorithms where Smith intentionally restricts to NP and P algorithms.
2. Hutter allows the environment to back-react to what the agent does, whereas Smith intentionally restricts the power of his intelligence tester so that it *cannot* decide,

[103] Baum & Durdanovic, who were not concerned with mathematical rigor, had some interesting heuristic ideas about intelligence as an "artificial economic system" where "market prices" can again be thought of as a perhaps useful numerical method of estimating and keeping track of the "usefulness" of various possible sub-algorithms. That also may be a good idea, but I believe that Baum & Durdanovic, at least in their earlier work (some correspondence with Baum suggests he is now coming round more to my point of view) misplaced their emphasis in the following sense: the real action and focus has to be on engineering and tuning the *search* over algorithms (which they called "meta-learning" and a "homunculus") not on the algorithms. The Baum-Durdanovic means for meta-learning were exceedingly primitive; and I believe the market-means Baum had in mind were by themselves simply inadequate to make a practically good searcher; and Baum initially was trying to deny the existence of a homunculus both inside human minds and as a useful means for building an AI. However, the large competence of newborn horses and turtles combined with the work of Piaget indicates that there presumably *is* a highly sophisticated homunculus – comparable to the brain of a newborn horse – directing the development of the human intelligence from behind the scenes by (at the very least) serving as its PG/SC intelligence tester/scorer, and it also is clear from §22 that brute force naive algorithm searches will be highly ineffective. For example Baum & Durdanovic's "Hayek" systems (which we consider from our perspective here to be an advance over, but certainly not a great advance over and maybe in some ways actually worse than, "brute force" algorithm searchers) were incapable of learning to solve the Rubik cube despite great effort, whereas in §22 we've sketched how a system with a decently engineered homunculus would handle that fairly easily. It appears both from my correspondence with Baum and from recent statements he is making on the web-site associated with his book [12], that experiences of this nature have convinced him of similar conclusions.

based on the Intelligence's answer, to change its testing policy, i.e. (in Hutter's terminology) Smith's "agent" cannot affect his "environment."

We made these restrictions for several good reasons (aside from not being as brilliant as Hutter):

Why Hutter's extra generality #2 is bad: The trouble with (2) is that Hutter's more general policy allows "games" to be played between the tester and testee, and allows the tester to try to be "unfair" and "biased" (e.g. if you do well on the tests at first, then I will intentionally alter my IQ tests from now on in order to make you look like a worse intelligence... but then the Intelligence can try to "fight back" by pretending to be stupid, although I can force that to fail...)[104] Hutter's extra generality #2 would yield a much less "user friendly" AI. Specifically, I want a pet AI which, when I give it problems, just tries to give me answers that maximize whatever scoring function I say! But with Hutter, you are going to get an AI that tries to outhink you and predict your future actions, and intentionally tries to give you worse answers with smaller scores if it believes that that will influence you to keep awarding those scores for longer. I don't want that! (And also it would be much harder to develop that kind of AI anyhow.)

We have presented good arguments in §6-7 that even our restricted (re #2) kind of AI would still be capable of a heck of a lot of intelligence – far more intelligence than ought to be enough to satisfy anyone – so I contend that the world is not yet ready to go for Hutter's higher kind of intelligence. Still, one must admit that it *is* a higher kind of intelligence. Indeed, AIXI seems to allow a continuum of kinds of intelligence (which Hutter calls t and ℓ) and Smith's MDoI is in some sense just the lowest point on that continuum. Also, of course, one does want to consider one's effect on one's environment – I just believe for practical purposes we can and should leave that part to the supervisor/owner of the AI. It is possible to a considerable extent even for our sort of AIs to predict the response of an environment and react appropriately. For example, our AI could be trained to predict environmental change, and then the resulting predctive model could be used as part of the reward function for a second AI trained to act on the environment.

If we could take that further and prove that our AIs in a "passive environment" are capable of doing anything Hutter's can do in an "active" one, then that would prove Hutter's extention entirely valueless. However, such a proof is presumably impossible because, as §14 showed, the computational complexity classes of the fundamental problems faced by Hutter's AIs can include PSPACE and EXPTIME, which are thought to be larger classes than ours. The underlying reason is that in Hutter's setup, his intelligences aim to predict rewards into the *infinite* future (and in order to do that optimally, need to "solve games" of unbounded duration)[105] whereas our intelligences, when they attempt to model enviromental changes, only do so a *bounded* number of steps into the future.

Why Hutter's extra generality #1 is bad: It also makes it much harder to develop an AI, while bringing us little compensatory benefit, and while causing intelligence now to become effectively unmeasurable, and the IQ test-answer scores now to become unjustifiable to external referees and attackers in any reasonable amount of time.

Indeed Bruce Maggs once argued to me that really, exponential-time algorithms "do not exist." Why? His point was that effectively, if you think you are running an exptime algorithm, you are deluded, and really you are running an exptime algorithm with a polynomial-time kill-cutoff, which therefore is really a P algorithm. In fact we perhaps should go even further and declare that really, the only algorithms that exist are linear time algorithms (with large additive and multiplicative constants). But I did not have the gall to do the latter so I stayed with P. Besides a lack of gall, there are other good reasons: the strong Church-Turing thesis says all reasonable models of computation are polynomially equivalent but I think plenty of people would say that *not* all reasonable models of computation are equivalent up to constant factors, so if we want a MDoI valid in all reasonable models of computation we'd better focus on P not on linear time.

Also, in §4 we mentioned the arguments of Searle and Block which indicate that "intelligence" largely loses its meaning if consumption of exponential time and/or space resources is permitted. So – all in all I advocate restricting to P.

Conclusion: be general, but not too general: So my current suspicion is the right way to expose this area is to do it Hutter's way in order to be brilliant and get high generality, but then explain that we want to get rid of excess generality, resulting in the Smith MDoI as the right foundation for future AI research (I think the full-Hutter way is at least 50 years ahead of its time and not warranted now). On the other hand by just exposing it our way, we lose Hutter's extra generality entirely but get an easier-to-understand exposition.

Convergence theorems: Both Solomonoff, Willis, and Hutter have got "convergence theorems" which I currently do not fully appreciate. We also have our own such theorem here of course, but quite probably theirs are better or by combining everybody's ideas something still better results (I am not sure which).[106] These theorems seem very important for providing the rigorous backing for their approaches, and it is Hutter who has the latest and greatest such theorem.

Universal intelligence test? A controversy and its resolution: Legg & Hutter in 2005-6 [103] devised what Hutter considered to be a "universal intelligence test." This came as rather a shock to me since our §8 presented arguments indicating the nonexistence of any such thing, and our whole framework is based on there being an infinity of different kinds of intelligence test, but a *single* UACI is capable of scoring asymptotically optimally on any one of them.

After re-examination of both [103] and our own arguments, I came to the conclusion that I was right – Legg and Hutter

[104] For example, a Hutterian intelligence tester could decide to award high scores to entities that defeat it in chess, *but* any entity defeating it more than 100 times in chess, suddenly would be condemned to zero scores for the rest of eternity. (So the best strategy for getting a high test score is to win exactly 100 chess games then no more.) However, with our sort of intelligence tester, that kind of nonsense is simply impossible.

[105] And Hutter's combined version of PG and the answer-scoring agency SC, also may need to play such long games, which our PG and SC cannot do since we demand they be polynomial time.

[106] The asymptotic sense in which our UACI is "optimal" perhaps is weaker than the optimality notion achieved by Hutter and Solomonoff. True? And if so, does that have any important practical impact?

went down a wrong path, and there is no such thing as a universal intelligence test. But there is some subtlety involved. We now discuss all that:

1. Let us first explain why this is an important question. It would have been very useful if a universal intelligence test existed, because it would allow avoiding having humans type in 20000 intelligence tests; you could just program one test that is just as good as those 20000 for training purposes.

2. Legg & Hutter's alleged "universal test" essentially consists of all possible intelligence tests but probabilistically weighted with weight $2^{-\ell}$ where ℓ is the individual test's binary code length.

Let us now state some relevant facts and then counterargue.

3. The present paper's UACI will get within an asymptotic constant factor (in fact in appropriate computational models the constant is $< 1+\epsilon$ for any $\epsilon > 0$) of maximum possible rate of reward-eating on any given test selected from the infinity of them. So it also will do so on the Legg-Hutter "universal test."

4. I regard L&H's universal test as, for practical purposes, useless!

5. (a) L&H's "universal" test will behave drastically differently if the algorithms are coded in a different programming language or for a different computational model.
(b) Also, one can get within any fixed constant factor (arbitrarily near 1) of the maximum possible reward-eating rate on the universal Legg-Hutter test, *without* being intelligent (by our definition, or anybody's) at all!
(c) And since in (a) distortions of the distribution by exponentially large constant multiplicative factors happen naturally, I conclude that (b) suggests that L&H are totally busted!

6. For the coup de grace, here is an (admittedly rather informal) argument that there is no such thing as a universal intelligence test, and instead you just have to be satisfied with an infinite number of different kinds of intelligence tests: We can make a putative intelligence X that will get high scores on test U and low scores on test B, *and this seems true even if you try to design U to be "universal."* Why? Because then U will necessarily award arbitrarily low weight (probabilistic or otherwise) to some subset of tests. So we'll make X do poorly at that subset – which will not matter much from U's point of view – but we will make B emphasize the bad subset heavily.

Even this informal argument is quite convincing to me that there can be no such thing as a single useful "universal intelligence test."[107] However, it can be replaced with a formal theorem and proof! This totally resolves the controversy in our favor, *provided* you accept the present work's definition of "intelligence test":

Theorem (no universal IQ test exists): *Any intelligence test U of the form in §9 has the property that some entity ET exists, that performs asymptotically optimally on it,* but *performs pessimally on some other intelligence test B.*

Proof. To get asymptotically competitively optimal behavior on U, simply make ET be a UACI as in theorem 5 of §12. To create B so that the ET will always score *zero* on B, add a special modification to ET to detect test problems of B's form, and to deliver appropriately bad answers whenever that detection happens. (We can design B to always output problems in a special easy-to-detect format, such as an all-1s bitstring, and B's scorer will demand a certain easily-evadable kind of answer to get a nonzero score, such as an all-0s bitstring.)

To complete the proof, we need to argue that a suitable B and ET-modification both exist such that the UACI's asymptotically optimal cumulative score on test U, is only negligibly diminished asymptotically. We can accomplish that by a "Cantor diagonalization" argument. The successive problems P_k output by U (or the successive *probability distributions* of the P_k if U is randomized) are considered. We can easily see ala Cantor that there necessarily will exist an alternative sequence of $\widetilde{P}_k$ that will necessarily *not* be generated by U – or only generated with negligible probability, indeed with *total* expected number of equalities $\widetilde{P}_j = P_k|_{1\le j<\infty,\, 1\le k\le j^3}$ upper bounded by an arbitarily small *constant* c, since, e.g, the expected count of $\widetilde{P}_j$ equalities is $\le 0.5cj^{-2}$. In fact, were U deterministic one could construct B by simply making B *be* U but with a postprocessing step added to alter U's output away from U's (and away from all of U's previous outputs) whereas for randomized U one could try $2k^2/c$ Monte-Carlo experiments (for each running U up to $k = j^3$) and then add a postprocessing step to pick an easy-to-recognize output not in the resulting $2j^9/c$-element set. Q.E.D.

A subtlety – the slain dragon returns to life?: Nevertheless I conjecture that a universal intelligence test resembling Legg & Hutter's *can* be justified if we permit what Hutter calls an "active environment," i.e. what in our terminology would be called "a tester that is allowed to see the provided answers to its previous problems and to react by evilly altering future test problems." This goes outside the definition of "intelligence test" permitted in the present paper:

Conjecture (an "active environment" IQ test exists that is universal with respect to all passive tests): *An intelligence test U of the form in §9 but altered so that SC and PG are the* same *entity ("active environment") exists that is "universal," i.e. UACIs and only UACIs get asymptotically competitively-optimally high scores on U, and if any "passive" intelligence test B exists that some ET does asymptotically competitively-poorly on, then that ET will also do at least as asymptotically competitively poorly on our active universal test U.*

The conjectured construction: Initially, let U be, as suggested by Legg & Hutter [103], all possible intelligence tests but probabilistically weighted with weight $2^{-\ell}$ where ℓ is the individual test's binary code length. But now add the following modification. Simply have U, as it proceeds with tests, keep track *both* of ET's cumulative score *and* of a pet UACI's

[107] It may be possible to formulate a positive theorem about the Legg-Hutter universal IQ test (they apparently did not) but if so I believe it will necessarily be so weak as to be useless.

I do not want to be too negative since it looks to me like Hutter is generally ahead of me going in the same direction, but I do not think so in this particular instance; I think here Hutter went off the right path. At present I still think the right path is our framework with an infinite set of intelligence tests but one UACI that is optimally good asymptotically at all of them. We would however gain increased confidence in that if we can formalize the nonexistence proof sketched in our (6).

cumulative score on each Subtest. Have it alter its weights as it goes so that tests on which ET does much worse than the pet UACI, ultimately get larger weights, in such a way that each weight sum (over time) is ∞ (this assures an infinite number of samples will eventually be collected for each subtest), but the largest weights (corresponding to the worst relative performance of ET versus U's pet UACI) ultimately tend to 1. (Appropriate weight-sum-preserving updating schemes are easily devised, but it might be necessary to update them extremely slowly in order to avoid "confusing" the pet UACI; I conjecture doubly-exponential slowness suffices.[108]) Q.E.D.

Even assuming this universal test works, I still regard it as useless for practical purposes (points $\overline{1}$, 4, and 5 above still either apply, or still apply "for practical purposes") – and also it definitely only can be accomplished by going outside the present paper's definition of "intelligence test," which is a course we have deprecated.

I believe (and hope) that each of our (Solomonoff, Hutter, and this) work casts new and useful light on, and complements, the other.

25 Multiresearcher Consensus

I believe it would be useful – in the sense that it would prevent several years from being wasted – to get a number of prominent human- and artificial-intelligence researchers to sign the following short "consensus statement":

1. "Intelligence" and "intelligence test" both have mathematical definitions, which (perhaps up to minor alterations) can be taken to be the ones in §9 of the present work.
2. It is already known how – easily – to build a "universal artificial intelligence" that would meet this definition, but which unfortunately would perform poorly in practice.
3. The AI community should adopt the previous two points as the foundation for future research.
4. The AI community should organize a perpetually ongoing "intelligence contest" open to both human and computer intelligences as contestants, accepting standardized "intelligence tests" contributed by anyone, and posting scoring records of all contestants on all tests. This modus operandi should ensure that clear definable and measurable gains in machine intelligence happen every year.

So far (22 May 2006) nobody has signed it besides me.

References

[1] G. Scott Acton & David H. Schroeder: Sensory discrimination as related to general intelligence, Intelligence 29 (2001) 263-271.

[2] The AES cryptosystem can be used as a random number generator by encrypting a fixed sequence of bit-strings.

[3] George M. Alliger: Do zero correlations really exist among measures of different intellectual abilities?, Educ'l & Psychological Measurement 48 (1988) 275-280.

[4] A.Ames: CNS energy metabolism as related to function, Brain Research Reviews 34, 1-2 (2000) 42-68.

[5] D.Attwell & S.B.Laughlin: An energy budget for signalling in the grey matter of the brain, J.Cerebral Blood 21,10 (Oct. 2001) 1133-1145.

[6] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, M. Protasi: Complexity and Approximation, Combinatorial optimization problems and their approximability properties Springer Verlag 1999.

[7] Alan Baddeley: Your memory, a user's guide, Firefly Books (new illustrated ed. 2004).

[8] A.D. Baddeley & D.J.A.Longman: The influence of length and frequency of training sessions on the rate of learning to type, Ergonomics 21 (1978) 627-635.

[9] Renee Baillargeon: Object permanence in $3\frac{1}{2}$ and $4\frac{1}{2}$ month old infants, Developmental psychology 23 (1987) 655-664.

[10] Torsten Baldeweg, MD & 4 others: Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials, Annals of Neurology 45, 4 (1999) 495-503.

[11] Christoph Bandelow: Inside Rubik's Cube and Beyond, Birkhauser 1982.

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

[13] E.B. Baum & Igor Durdanovic: An artificial economy of Post production systems, IWLCS 3 (2000) 3-21. Q325.5.A344cs

[14] E.B. Baum & Igor Durdanovic: Evolution of cooperative problem solving in an artifical economy, Neural Computation 12 (2000) 2743-2775.

[15] W.Baur & V.Strassen: The complexity of partial derivatives, Theor. Comput. Sci. 22 (1983) 317-330.

[16] Bengt Beckman (translated by Kjell-Ove Widman): Arne Beurling and the Swedish Crypto Program During World War II, American Mathematical Society 2003.

[17] Camilla P. Benbow: Physiological correlates of extreme intellectual precocity, Neuropsychologia 24 (1986) 719-725.

[18] S. Ben-David, B. Chor, O. Goldreich, M. Luby: On the theory of average case complexity, J. Computer & System Sciences 44 (1992) 193-219.

[19] E.Berlekamp, J.Conway, R.Guy: Winning Ways (for your Mathematical Plays), Volume 2, Academic Press 1982.

[20] Ned Block: Psychologism and Behaviorism, The Philosophical Review LXXXX, 1 (January 1981) 5-43.

[21] A.Booth, G.Shelley, A.Mazur, G.Tharp, R.Kittok: Testosterone, and winning and losing in human competition, Horm. Behav. 23,4 (Dec 1989) 556-571.

[22] Allan Borodin & Ran El-Yaniv: Online computation and competitive analysis, Cambridge University Press 1998.

[23] T.J.Bouchard & M.McGue: Familial studies of intelligence, a review, Science 212, 4498 (1981) 1055-1059. See also Bouchard et al: Sources of human psychological differences: the Minnesota study of twins reared apart, Science 250, 4978 (1990) 223-228.

[24] B. Bouzy & T. Cazenave: Computer Go: An AI-Oriented Survey, Artificial Intelligence 132, 1 (2001) 39-103.

[108] This issue is the entire reason this presently is a "conjecture" and not a "theorem" – if uncomputable slowness is required then we are in trouble.

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

[26] Gunnar Brinkmann, Brendan D. McKay: Posets on up to 16 Points, Order 19,2 (2002) 147-179.

[27] P.A.Carpenter, M.A.Just, P.Shell: What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven Progressive Matrices Test, Psychological Review 97,3 (July 1990) 404-431.

[28] J.B.Carroll: Human cognitive abilities, survey of factor studies, Cambridge Univ. Press 1993.

[29] S.J.Ceci, J.E.Baker, U.Bronfenbrenner: Prospective remembering, temporal calibration, and context, 360-365 in *Practical Aspects of Memory* vol. 1 Wiley 1988 (M.M.Gruneberg et al. eds).

[30] W.G. Chase & H.A. Simon: Perception in chess, Cognitive Psychology, 4 (1973) 55-81.

[31] Bernard Chazelle: Triangulating a Simple Polygon in Linear Time, Discrete Comput. Geom. 6,5 (1991) 485-524.

[32] Paul M. Churchland: A neurocomputational perspective: the nature of mind and the structure of science, MIT Press 1989.

[33] J.Columbo: Infant cognition: predicting childhood intellectual function, Sage, Newbury Park CA 1993.

[34] Steve Cook: The Complexity of Theorem-Proving Procedures, Proceedings of the third annual ACM symposium on Theory of computing STOC 3 (1971) 151-158.

[35] Matthew Cook: Universality in Elementary Cellular Automata, Complex Systems 15,1 (2004) 1-40.

[36] Stanley Coren: The left-hander syndrome: the causes and consequences of left-handedness, Free Press New York 1992.

[37] Catharine M. Cox: The Early Mental Traits of Three Hundred Geniuses, in volume 2 of *Genetic Studies of Genius* (Lewis M. Terman, ed.). Stanford University Press 1926.

[38] R.W.Cranson & 5 others: Transcendental meditation and improved performance on intelligence-related measures, Personality & Individual Differences 12 (1991) 1105-1116.

[39] Francis Crick: Consciousness and neuroscience, Cerebral Cortex 8,2 (1998) 97-107.

[40] Joan Daemen & Vincent Rijmen: The design of Rijndael, the Advanced Encryption Standard, Springer-Verlag 2003.

[41] James M. & Mary G. Dabbs: Heroes, Rogues, and Lovers: Testosterone and Behavior, McGraw-Hill 2000.

[42] DARPA October 1983: Strategic computing: New-generation computing: a strategic plan for its development and application to critical problems in defense.

[43] V.Danthiir & R.D.Roberts: What the nose knows: olfaction and cognitive abilities, Intelligence 29,4 (2001) 337-361.

[44] Ian J. Deary: Intelligence, a very short introduction, Oxford Univ. Press 2001.

[45] Ian J. Deary: Looking down on human Intelligence, Oxford Univ. Press (psychology series #34) 2000.

[46] I.J. Deary, G. Thorpe, V. Wilson, J.M. Starr, L.J. Whalley: Population sex differences in IQ at age 11: the Scottish mental survey 1932, Intelligence 31,6 (2003) 533-542.

[47] A.D. de Groot: Thought and choice in chess (1st ed.) Mouton Publishers, The Hague 1965.

[48] J.E. Dennis, Jr. & Robert B. Schnabel: Numerical Methods for Unconstrained Optimization and Nonlinear Equations, SIAM (classics in applied math. #16) 1983.

[49] R.DeVries: Constancy of generic identity in the years three to six, Monographs of the society for research in child development 34(3, whole #127).

[50] Richard E. Dickerson: Exponential correlation of IQ and the wealth of nations, Intelligence 34 (2006) 291-295.

[51] Hubert L. Dreyfus: What Computers Can't Do: A Critique of Artificial Reason, Harper & Row, New York 1979; What Computers Still Can't Do, MIT Press 1992.

[52] J.R. Driscoll, N. Sarnak, D.D. Sleator, R.E. Tarjan: Making data structures persistent, J. Computer & System Sciences, 38,1 (Feb. 1989) 86-124.

[53] D-Z. Du & K-I. Ko: Theory of Computational Complexity, John Wiley & Sons 2000.

[54] Gerald M. Edelman: Bright air brilliant fire, On the matter of the mind; A Nobel laureate's revolutionary vision of how the mind originates in the brain, BasicBooks, New York 1992.

[55] S. Even & R.E. Tarjan: A combinatorial problem which is complete in polynomial space, J. Assoc. Computing Machinery 23 (1976) 710-719.

[56] Raymond B. Fancher: Spearman's g, A model for Burt?, British J. Psychology 76,3 (Aug. 1985) 341-352.

[57] E.A.Feigenbaum & P.McCorduck: Fifth Generation: Artificial Intelligence and Japan's Challenge to The World, Addison-Wesley 1983.

[58] P.M. Fitts: The information capacity of the human motor system in controlling the amplitude of movement, J. Exp'l Psychology 47 (1954) 381-391.

[59] R.Flin & 3 others: The effect of a five month delay on chilren's and adult's eyewitness memory, British J. Psychology 83 (1992) 323-326.

[60] A.S. Fraenkel & D.Lichtenstein: Computing a perfect strategy for $n \times n$ chess requires time exponential in n, J. Combinatorial Theory A 31 (1981) 199-214.

[61] M.R. Garey & D.S. Johnson: Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman, 1979.

[62] Thomas A. Gentry, Kristopher M. Polzine, James A. Wakefield, Jr: Human genetic markers associated with variation in intellectual abilities and personality, Personality & Individual Differences 6, 1 (1985) 111-113.

[63] C.C.A.M. Gielen & E. J. van Zuylen: Coordination of arm muscles during flexion and supination: Application of the tensor analysis approach, Neuroscience 17,3 (1986) 527-539

[64] James Gleick: Genius: The Life and Science of Richard Feynman, Pantheon, New York 1992.

[65] G.H.Golub & C.F.Van Loan: Matrix computations, Johns Hopkins Studies in Mathematical Sciences (3rd ed 1996).

[66] M.Gopnik & M.Crago: Familial aggregation of developmental language disorder, Cognition 39 (1991) 1-50.

[67] Linda S. Gottfredson: Mainstream Science on Intelligence: An Editorial With 52 Signatories, History, and Bibliography, Intelligence 24, 1 (1997) 13-23; see also Wall Street Journal 13 December 1994. This is a statement about "intelligence" purporting to represent "the mainstream view" and signed by 52 distinguished experts in the psychometric field including Bouchard, Carroll, and Jensen. The signatures of 131 distinguished experts were solicited with the result that 31 did not respond, 52 signed, and 48 refused to sign for various reasons (sometimes because of disagreement about content, sometimes disagreement about presentation, and sometimes out of political fears).

[68] Stephen J. Gould: The mismeasure of man, W.W.Norton 1981.

[69] Joy Paul Guilford: The nature of human intelligence, McGraw-Hill, New York 1967.

[70] Jan-Eric Gustafson: A unifying model for the structure of intellectual abilities, Intelligence 8 (1984) 179-203.

[71] J.Hass, J.Lagarias, N.Pippenger: The computational complexity of knot and link problems, J. Assoc. Computing Machinery 64 (1999) 185-211.

[72] John Haugeland: Artificial Intelligence: the very idea, MIT Press 1986.

[73] John R. Hayes: The Complete Problem Solver, L.Erlbaum Assoc. (2nd ed.) 1989.

[74] L.Hearnshaw: Cyril Burt, psychologist, Cornell Univ. Press 1979.

[75] R. Hecht-Nielsen: Neurocomputing, Addison-Wesley 1989.

[76] W.E.Hick: On the rate of gain of information, Quarterly J. Experimental Psychol. 4 (1952) 11-26.

[77] Roger A. Horn & Charles R. Johnson:, Matrix Analysis, Cambridge Univ. press, 1985.

[78] Marcus Hutter: The Fastest and Shortest Algorithm for All Well-Defined Problems, Int'l J. Foundations of Computer Science 13,3 (June 2002) 431-443.

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

[80] Most of Hutter's papers are cited in and/or incorporated into his book [79] and are available on his web page `http://www.idsia.ch/~marcus/ai`.

[81] N. H. Ibragimov: Elementary Lie Group Analysis and Ordinary Differential Equations, Wiley 1999.

[82] R. Impagliazzo & L. Levin: No better ways to generate hard NP instances than picking uniformly at random, Proceedings 31th Annual Symposium on Foundations of Computer Science FOCS (1990) 812-821.

[83] S. Iwata & T. Kasai: The Othello game on an $n \times n$ board is PSPACE-complete, Theor. Computer Sci. 123 (1994) 329-340.

[84] J.Jenkins, K.Dallenbach: Oblivescense During Sleep and Waking, American J. Psychology 35 (1924) 605-612.

[85] Arthur R. Jensen: Bias in mental testing, Free Press, New York 1980.

[86] Arthur R. Jensen: The g factor: The science of mental ability. Westport, CT: Praeger 1998

[87] Arthur R. Jensen: Individual differences in the Hick paradigm, pp. 101-175 in Philip A. Vernon (ed.) *Speed of information processing and Intelligence*, Ablex Pub. Corp, Norwood NJ 1987.

[88] Arthur R. Jensen: Straight talk about mental tests, Free Press, New York 1981.

[89] Wendy Johnson and Thomas J. Bouchard, Jr.: Constructive replication of the visual-perceptual-image rotation model in Thurstone's (1941) battery of 60 tests of mental ability, Intelligence 33,4 (2005) 417-430.

[90] Leon J. Kamin: The science and politics of IQ, Lawrence Erlbaum Assoc. 1974.

[91] E.Kamke: Differentialgleichungen: Lösungsmethoden und Lösungen, 2 vols. Teubner. Stuttgart Germany 1959.

[92] S.W. Keele: Movement control in skilled motor performance, Psychological Bulletin 70 (1968) 387-403.

[93] P.Kempel & 5 others: Second-to-fourth digit length, testosterone and spatial ability, Intelligence 33,3 (2005) 215-230.

[94] Christof Koch: Biophysics of Computation: Information Processing in Single Neurons, Oxford University Press 1999.

[95] Bryan Kolb & Ian Q. Whishaw: Fundamentals of human neuropsychology (5th ed.), Freeman-Worth, New York 2003.

[96] J.F. Korsh & P. LaFollette: Multiset Permutations and Loopless Generation of Ordered Trees with Specified Degree Sequence, J. Algorithms 34,2 (2000) 309-336.

[97] J.H.Kranzler & A.R.Jensen: The nature of psychometric g – unitary process or a number of independent processes? Intelligence 15 91991) 397-422.

[98] Robert Krueger & Wendy Johnson: The Minnesota Twin Registry: Current Status and Future Directions, Twin Research 5, 5 (October 2002) 488-492.

[99] Anand Kumar & Shanker Krishnan: Memory Interference in Advertising: A Replication and Extension, J. Consumer Research 30,4 (2004) 602-611.

[100] Ray Kurzweil: The singularity is near: When Humans Transcend Biology, Penguin 2005.

[101] E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy-Kan, D.B. Shmoys: The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization, John Wiley & Sons 1985. A web page by David Applegate, Robert Bixby, Vasek Chvatal, and William Cook on their Concorde TSP-solver and some of the record TSPs it has solved, including 24978 cities in Sweden, is `http://www.tsp.gatech.edu/`.

[102] A.Leahy: Nature-nurture and intelligence, Genetic Psychology Monographs 17,4 (August 1935) 235-308.

[103] Shane Legg & Marcus Hutter: A Universal Measure of Intelligence for Artificial Agents, IDSIA technical report 04-05 (Galleria 2, CH-6928 Manno-Lugano, Switzerland, April 2005); A Formal Measure of Machine Intelligence, IDSIA TR 10-06 April 2006 (8 pages) presented at Annual Machine Learning Conference of Belgium and The Netherlands (Benelearn-2006) and both are available on Hutter's web page `http://www.idsia.ch/~marcus/official/publ.htm`.

[104] Doug Lenat & E.A.Feigenbaum: On the threshold of knowledge, Artificial Intelligence 47 (1991) 185-250.

[105] Leonid A. Levin: Universal sequential search problems, Problems of Information Transmission 9 (1973) 265-266; original Russian version: Problemy Peredaci Informacii 9,3 (1973) 115-116.

[106] S-C. Li, M. Jordanova, U. Lindenberger: From good senses to good sense: A link between tactile information processing and intelligence, Intelligence 26,2 (1998) 99-122.

[107] M. Li & P.M.B. Vitanyi: An introduction to Kolmogorov complexity and its applications, Springer (2nd edition) 1997.

[108] Robert L. Linn: A Monte Carlo approach to the factors problem, Psychometrika 33 (1968) 37-71.

[109] M.Linton: Real world memory after six years: An in vivo study of very long term memory, pp.69-76 in M.M.Grunberg et al. (eds) Practical aspects of memory, Academic Press London 1978.

[110] E.L.Loftus: Eyewitness testimony, Harvard University Press 1996.

[111] Elizabeth F. Loftus & Hunter G. Hoffman: Misinformation and Memory, The Creation of New Memories, J. Experimental Psychology: General 118,1 (1989) 100-104.

[112] E.F.Loftus: Creating false memories, Scientific American 277,3 (September 1997) 70-75.

[113] E.F.Loftus & J.C. Palmer Reconstruction of automobile destruction, J. Verbal Learning & Verbal Behaviour 13 (1974) 585-589.

[114] Joan M. Lucas: The rotation graph of binary trees is hamiltonian, J. Algorithms 8,4 (1987) 503-535.

[115] J.M. Lucas, D. Roelants van Baronaigien, F. Ruskey: On Rotations and the Generation of Binary Trees, J. Algorithms 15,3 (1993) 343-366.

[116] Richard Lynn: Eugenics, a reassessment, Praeger, Westport CT 2001.

[117] R.Lynn & D.C. Rowe: Skin Color and Intelligence in African Americans, Population & Environment 23 (2002) 365-375; Lynn: Skin Color and Intelligence in African Americans: A Reply to Hill Population & Environment 24,2 (2002) 215-218.

[118] Richard Lynn & Tatu Vanhanen: IQ and the Wealth of Nations, Praeger Westport CT 2002.

[119] N.J.Mackintosh: Does it matter? The scientific and political consquences of Burt's work, ch.7 in N.J.Mackintosh (ed.): *Cyril Burt: Fraud or Framed*, Oxford University Press 1995.

[120] N.J.Mackintosh: IQ and human Intelligence, Oxford University Press 1998.

[121] Gary F. Marcus & Simon E. Fisher: FOXP2 in focus: what can genes tell us about speech and language? Trends in Cognitive Sciences 7,6 (June 2003) 257-262.

[122] M. K. Mason: Learning to Speak After Six and a Half Years of Silence, Journal of Speech Disorders 7 (1942) 295-304.

[123] John McCarthy: Review of *The Emperor's New Mind* by Roger Penrose, Bulletin Amer. Math'l Society (October 1990).

[124] B.D. McKay: Isomorph-free exhaustive generation, J. Algorithms 26,2 (1998) 306-324.

[125] I. Chris McManus: Right Hand, Left Hand, the origins of asymmetry in brains, bodies, atoms, and cultures, Harvard University Press 2002.

[126] Donald Michie: On machine intelligence, Ellis Horwood (2nd ed.) 1986.

[127] G.A. Miller: The magical number seven plus or minus two: some limits on our capacity for information processing, Psychological Review 63 (1956) 81-91.

[128] Henryk Minc: Nonnegative matrices, Wiley 1988.

[129] Marvin L. Minsky: Finite and infinute machines, Prentice-Hall 1967.

[130] Neil Moray, A. Bates, T. Barnett: Experiments on the Four-Eared Man, J. Acoust. Soc. Amer. 38,2 (1965) 196-201.

[131] Erich Neumann: The origins and history of consciousness, Princeton Univ. Press 1954, Bollingen reprint, has been reprinted at least 12 times.

[132] Allen Newell: Unified theories of cognition, Harvard Univ. press 1990.

[133] Allen Newell & H.A.Simon: Human problem solving, Prentice-Hall, Englewood Cliffs, NJ 1972.

[134] A. Newell & P.S. Rosenbloom: Mechanisms of skill acquisition and the law of practice, 1-55 in J. Anderson (ed.) Cognitive Skills and Their Acquisition, Lawrence Erlbaum Associates, Hillsdale, NJ 1981.

[135] Ulric Neisser & 10 others: Intelligence: knowns and unknowns, American Psychologist 51 (Feb. 1996) 77-101; also `www.lrainc.com/swtaboo/taboos/apa_01.html`. This is an attempt to generate a consensus statement.

[136] U. Neisser (ed.): The rising curve: Long term gains in IQ and related measures, Amer. Psychological Assoc. 1998.

[137] T.Nelson, M.Fehling, J.Moore-Glascock: The nature of semantic savings for items forgotten from long-term memory, J. Experimental Psychology: General 108 (1979) 225-250.

[138] A.Newell & H.A.Simon: Program That Simulates Human Thought, in Computers and Thought (eds. Feigenbaum & Feldman) McGraw-Hill, N.Y., 1963. See also A.Newell, J.C. Shaw„ H.A. Simon: IFIP 60 (Paris 1960) 256-264; CACM 19,3 (1976) 113-126.

[139] E.Newport: Maturational constraints on language learning, Cognitive Science 14 (1990) 11-28.

[140] Nils J. Nilsson: Eye on the Prize, AI Magazine 16,2 (1995) 9-17.

[141] Helmuth Nyborg (ed.): The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen, Pergamon 2003.

[142] M.W. O'Boyle & C.P. Benbow: Handedness and its relationship to ability and talent, pp. 343-372 in S.Coren (ed.): *Left-handedness: behavioral implications and anomalies*, North-Holland 1990.

[143] M.Oaksford & G.D.A. Brown (eds.): Neurodynamics and Psychology, Academic Press 1994.

[144] Octopus twists for shrimps An octopus in a German zoo has learned to open jars of shrimps by copying staff – and is now showing off her skills to visitors. BBC News story, 25 February 2003.

[145] Peter J. Olver: Applications of Lie Groups to Differential Equations, Springer (2nd ed. GTM 107) 2000.

[146] David Owen with Marilyn Doerr: None of the above: the truth behind the SATs, Rowman & Littlefield Publishers Lanham, Md. 1999.

[147] 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.

[148] Christos H. Papadimitriou: Computational Complexity, Addison Wesley 1994.

[149] F.Partensky, W.R.Hess, D.Vaulot: Prochlorococcus, a marine photosynthetic prokaryote of global significance, Microbiol. Molec. Biol. Rev. 63 (1999) 106-127.

[150] A.Pellionisz & R.Llinas: Brain modelling by tensor network theory... Neuroscience 4 (1979) 323-348; Space-time representation in the brain, the cerebellum..., Neuroscience 7,12 (1982) 2949-2970.

[151] Roger A. Penrose: The emperor's new mind, Oxford Univ. Press 1989.

[152] Steven Pinker: The language instinct, William Morrow 1994.

[153] William H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling: Numerical Recipes: The Art of Scientific Computing, Cambridge University Press.

[154] Hilary Putnam: Review of Penrose's *Shadows of the mind*, Bull. Amer. Math'l Soc. 32,3 (1995) 370-373.

[155] Hilary Putnam: Much ado about not very much, pp. 269-281 in Stephen R. Graubard (ed.): The artificial intelligence debate, MIT Press 1988.

[156] J.R.Quinlan: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Francisco 1993.

[157] Patrick Rabbitt: Does fast last? Is speed a basic factor determining individual differences in memory? 161-168 in Practical Aspects of Memory (M.M. Gruneberg et al eds.) Wiley 1988.

[158] Allan Ramsay: Formal methods in Artificial Intelligence, (Cambridge tracts in TCS 6) Cambridge Univ. Press 1988.

[159] John E.G. Raymont: Plankton and productivity in the oceans, volume I (phytoplankton), 2nd ed. Pergamon 1980.

[160] Controversial Theory Linking Reading Ability to Specific Brain Region Gets a Boost, Scientific American News, 20 April 2006

[161] T. Edward Reed, Philip A. Vernon, Andrew M. Johnson: Confirmation of correlation between brain nerve conduction velocity and intelligence level in normal adults, Intelligence 6 (2004) 563-572.

[162] S. Reisch: Hex ist PSPACE-vollständig, Acta Informatica 15 (1981) 167-191.

[163] Elaine Rich & Kevin Knight: Artificial Intelligence (2nd ed.) McGraw Hill 1991.

[164] J. M. Robson: N by N checkers is Exptime complete, SIAM J. Computing, 13,2 (1984) 252-267.

[165] Susan A. Rose, & Judith F. Feldman: Prediction of IQ and Specific Cognitive Abilities at 11 Years From Infancy Measures, Developmental Psychology 31,4 (July 1995) 685-696.

[166] P.S. Rosenbloom, J.E. Laird, A. Newell: The chunking of skill and knowledge, 391-410 in Working models of human perception (B.A.G. Elsendoorn & H. Bouma eds.) Academic Press 1989.

[167] D.E.Rumelhart, G.E.Hinton, R.J.Williams: Learning internal representations by error propagation, pp. 318-362 in D.E.Rumelhart & J.L.McClelland (eds.) Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1, Cambridge, MA: The MIT Press 1986.

[168] J. Phillippe Rushton: Race, Evolution, and Behavior, Transaction Publishers 1995.

[169] Stuart J, Russell & Peter Norvig: Artificial Intelligence: A Modern Approach, (2nd Edition) Prentice-Hall 2003.

[170] Russ Rymer: Genie: a scientific tragedy, Harper Perennial, New York 1994.

[171] Carla Savage: A survey of combinatorial Gray codes, SIAM Review 39,4 (1997) 605-629.

[172] Max Scharnberg: The non-authentic nature of Freud's observations, Acta Universitatis Upsaliensis, Uppsala (Uppsala Studies in Education #47 and 48) 1993.

[173] Uwe Schöning: A probabilistic algorithm for 3-SAT and constraint satisfaction problems, SFOCS 40 (1999) 410-414.

[174] Daniel Schutzer: AI an applications-oriented approach, Van Nostrand NY 1987.

[175] G. Seibel: Discrimination reaction time on a 1023-alternative task, J. Exp'l Psychology 66 (1963) 215-226.

[176] E.Seneta: Nonnegative matrices and Markov chains, 2nd ed Springer 1981.

[177] Evelyn Sharp: The IQ cult, Coward, McCann, Geohegan, New York 1972.

[178] Jianhong Shen: Singular values of Gaussian random matrices, Linear Algebra and its Applications 326 (2001) 1-14.

[179] Thomas H. Shephard II & S.M.Gartner: Increased incidence of nontasters of phenylthiocarbamide among congenital athyreotic cretins, Science 131, 3404 (Mar. 1960) 929.

[180] Jerome M. Siegel: Why we sleep, Scientific American (Nov. 2003) 92-97.

[181] Jerome M. Siegel: Clues to the functions of mammalian sleep, Nature 437 (2005) 1264-1271.

[182] R.S. Siegler & M.W.Alibali: Children's thinking (4th ed.), Prentice-Hall 2005.

[183] Robert S. Siegler & D.D. Richards: The development of intelligence, 901-971 in (R.J. Sternberg, ed.) *Handbook of human intelligence* Cambridge 1982.

[184] Marian Sigman, Sarale E. Cohen & Leila Beckwith: Why does infant attention predict adolescent intelligence, Infant Behavior & Development 20,2 (1997) 133-140.

[185] David Singmaster: Notes on Rubik's magic cube, Enslow, 1981.

[186] M.Sipser: A Complexity theoretic approach to randomness, Proceedings 15th ACM Symposium on Theory of Computing STOC (1983) 330-335.

[187] Michael Sipser: Introduction to the Theory of Computation, PWS publishers 1997.

[188] M.Skodak & M.H.Skeels: A final followup study of 100 adopted children, J.Genetic Psychol. 75 (1949) 85-125.

[189] N.J. Slamecka: Retroactive inhibition of connected discourse as a function of practice level, J.Experimental Psychology 59 (1960) 104-108.

[190] W.D.Smith: Information content of human intelligence and life, in preparation.

[191] Ray Solomonoff: The Universal Distribution and Machine Learning, The Computer Journal 46,6 (Nov. 2003) 598-601.

[192] Ray Solomonoff: Three Kinds of Probabilistic Induction: Universal Distributions and Convergence Theorems, To appear in Festschrift for Chris Wallace.

[193] Ray Solomonoff: Progress in Incremental Machine Learning; Revision 2.0, 30 Oct. 2003, Given at NIPS Workshop on Universal Learning Algorithms and Optimal Search, Dec. 14, 2002, Whistler, B.C., Canada.

[194] Ray Solomonoff: Two Kinds of Probabilistic Induction, The Computer Journal 42,4 (1999) 256-259.

[195] Ray Solomonoff: A Formal Theory of Inductive Inference, I: Information and Control 7,1 (1964) 1-22; II: 7,2 (1964) 224-254. [All of these Solomonoff papers are available on his web page `http://world.std.com/~rjs/pubs.html`.]

[196] Norman E. Spear & David C. Riccio: Memory phenomena and principles, Allyn & Bacon 1994.

[197] Charles Spearman: "General Intelligence," objectively determined and measured, American J. Psychology 15 (1904) 201-293.

[198] George Sperling: The information available in brief visual presentation. Psychological Monograph 74,11 (1960) [whole no. 498].

[199] Saul Sternberg: High-Speed Scanning in Human Memory, Science 153, 3736 (Aug 1966) 652-654.

[200] Larry J. Stockmeyer: Planar 3-colorability is NP-complete, SIGACT News 5,3 (1973) 19-25. Stockmeyer's reduction is also briefly described in the easier-to-obtain Jörg Rothe: Heuristics versus completeness for graph coloring, Chicago J. Theoretical CS (2000) article 1.

[201] G.J. Sussman: A computer model of skill acquisition, American Elsevier 1975.

[202] Max Tegmark: The importance of quantum decoherence in brain processes, Physical Review E 61 (2000) 4194-4206. Also discussed in 4 February 2000 issue (#287) of *Science*.

[203] R.L.Thorndike: Stability of factor loadings, Personality & Indiv. Differences 8,4 (1987) 685-686.

[204] Louis L. Thurstone: Vectors of the mind, 1935, Thurstone later redid and expanded this book as Multiple Factor Analysis 1947 (both University of Chicago Press).

[205] L.L. Thurstone: Primary Mental Abilities, 1938.

[206] Louis L. & Thelma G. Thurstone: Factorial studies of Intelligence, 1941.

[207] John Todman, Iain Crombie, Leona Elder: An individual difference test of the effect of vitamin supplementation on non-verbal IQ, Personality & Individual Differences 12,12 (1991) 1333-1337.

[208] "Too Smart to be a Cop!" (2000 news story), Associated Press as published by ABC News 09/08.

[209] M.Tsukamoto: Program stacking technique. Information Processing in Japan (Information Processing Society of Japan) 17,1 (1977) 114-120.

[210] A.M.Turing: Computing machinery and intelligence, Mind 59 (Oct. 1950) 433-460.

[211] A.M.Turing: On Computable Numbers, With an Application to the Entscheidungsproblem, Proc. London Math. Soc. 2,42 (1936) 230-265; 43 (1937) 544-546.

[212] R. Venkatesan & L. Levin: Random instances of a graph coloring problem are hard. Proceedings 20th Annual Symposium on Theory of Computing STOC (1988) 217-222.

[213] Philip A. Vernon (ed.; this is a survey book with many authors): Biological Approaches to the Study of Human Intelligence, Ablex Pub. Corp, Norwood NJ 1993.

[214] Philip A. Vernon: The generality of g, Personality & Indiv. Differences 10,7 (1989) 803-804.

[215] Amy Wallace: The Prodigy / a Biography of William Sidis, America's Greatest Child Prodigy, E.P.Dutton NY 1986.

[216] David L. Waltz: The prospects for building truly intelligent machines, pp. 191-212 in Stephen R. Graubard (ed.): The artificial intelligence debate, MIT Press 1988.

[217] Betty U. Watson: Some Relationships Between Intelligence and Auditory Discrimination, Journal of Speech and Hearing Research 34 (June 1991) 621-627.

[218] David Wechsler: Manual for the Wechsler Adult Intelligence Scale, Psychology Corp. NY 1997.

[219] R. Wegmann: The asymptotic eigenvalue-distribution for a certain class of random matrices, J. Math. Anal. Appl. 56 (1976) 113-132

[220] J.M.Wicherts & 6 others: Are intelligence test measurements invariant over time? Investigating the nature of the Flynn effect, Intelligence 32 (2004) 509-537.

[221] Frank Wilczek: A call for a new physics (highly negative review of Penrose book "Shadows of the mind"), Science 266, 5191 (1994) 1737-1738.

[222] S.W.Wilhelm, M.G. Weinbauer, C.A.Suttle, W.H.Jeffrey: The role of sunlight in the removal and repair of viruses in the sea, Limnology and Oceanography 43 (1998) 586-592.

[223] D.E.Wilkins: Practical planning, Morgan Kaufmann Publishers 1988.

[224] D.G.Willis: Computational Complexity and Probability Constructions, J. Assoc. Computing Machinery (April 1970) 241-259.

[225] Patrick H. Winston: Artificial Intelligence, Addison-Wesley (now 3rd Edition) 1992.

[226] Dan J. Woltz & 3 others: Memory for Order of Operations in the Acquisition and Transfer of Sequential Cognitive Skills, J. Experimental Psychology: Learning, Memory, and Cognition 22,2 (March 1996) 438-457.

[227] Martin Wood: Soil biology, Blackie London 1989.

[228] Eugene B. Zechmeister & Stanley E. Nyberg: Human memory, Brooks/Cole 1982.

[229] D.Zwillinger: Handbook of Differential Equations, 3rd ed. Boston, MA: Academic Press 1997.