Noesis

The Journal of the Mega Society

Issue #214, December 2024

About the Mega Society

The Mega Society was founded by Dr. Ronald K. Hoeflin in 1982. The 606 Society (6 in 10^6), founded by Christopher Harding, was incorporated into the new society and those with IQ scores on the Langdon Adult Intelligence Test (LAIT) of 173 or higher were also invited to join. (The LAIT qualifying score was subsequently raised to 175; official scoring of the LAIT terminated at the end of 1993, after the test was compromised.) A number of different tests were accepted by 606 and during the first few years of the Mega Society's existence. Later, the LAIT and Dr. Hoeflin's Mega Test became the sole official entrance tests, by majority vote of the membership. Then, Dr. Hoeflin's Titan Test was added. (The Mega Test and Titan Test were also compromised, so Mega Test scores after 1994 and Titan Test scores after August 31st, 2020 are currently not accepted; the Mega and Titan cutoff is 43 - but either the LAIT cutoff or the cutoff on Dr. Ronald K. Hoeflin's tests will need to be changed, as they are not equivalent.) The Mega Society now accepts qualifying scores on The Hoeflin Power Test and on The Ultra Test. Both tests are still being scored. The Mega Society publishes this irregularly-timed journal.

Answer sheets for The Hoeflin Power Test and The Ultra Test can be emailed to ultrapowertest@gmail.com; the scoring fee for each test is $10 USD, payable via Stripe.

https://megasociety.org/#admission

The society also has a (low-traffic) members-only email list. Mega members, please contact one of the Mega Society officers to be added to the list.

For more background on Mega, please refer to Darryl Miyaguchi's "A Short (and Bloody) History of the High-IQ Societies" —

http://miyaguchi.4sigma.org/BloodyHistory/history.html

—and the official (designed) Mega Society page,

http://www.megasociety.org/

Noesis is the journal of the Mega Society, an organization whose members are selected by means of high-range intelligence tests.

Brian Wiksell (P.O. Box 366, Solana Beach, CA 92075) is the Administrator of the Mega Society. Inquiries regarding membership should be directed to Brian Wiksell at the aforementioned P.O. Box or the following email address: bwiksell@megasociety.org

Opinions expressed in these pages are those of individuals, not of *Noesis* or the Mega Society.

Noesis #214, December 2024

Editorial

Richard May, Ken Shea

Please enjoy the current issue of *Noesis* covering the art of storytelling, political correctness, artificial intelligence, high-range test construction, work-life balance in the medical industry, the leading edge of invention, and the mind-bending concepts of infinity and eternity. Ready?

Chris Cole is the leadoff hitter in the current issue with “The Future of Stories”, which examines the purpose of and lessons imparted through relatable tales, traditional or interactive.

Returning champion Bob Williams, then, sits down for a conversation with Scott Douglas Jacobsen on a range of topics: recent controversies in the social sciences, the cross-disciplinary replication crisis, gender differences, birth control, dysgenic effects of immigration, and AI.

After that, the current Mega Society Internet Officer, Daniel Shea, discusses the theory and practice of high-range test construction, in general, and advances made in constructing the Adaptive IQ Test, in particular, with Scott Douglas Jacobsen. Success truly has many fathers.

In a lengthy interview with Scott Douglas Jacobsen, Paul Cooijmans addresses a host of issues, including but not limited to: *g* in the high range, pitfalls of high-range test construction, gender differences among high-range test takers and high-range scores, autodidacticism, test bias, the value of heterogeneity among test items, norming high-range tests, personality differences in high scorers, dealing with fraudulent score reports, noteworthy high-range test constructors, the *sine qua non* of genius, and favorite high-range tests.

Then, Steven Pinker, cognitive scientist and popular author, also sits for a spirited interview with Scott Douglas Jacobsen. Subjects include: freedom of expression in higher education and the workplace, the importance of research independence, social media, ideological capture, virtue signaling, and the repercussions of cancel culture.

Dr. Benoit Desjardins engages in discussion with Scott Douglas Jacobsen, as well. Dr. Benoit Desjardins recounts: personal health challenges, decades of experiences with the US and Canadian medical/healthcare systems, diverging public perceptions of science and medicine, the unique attributes of the Philadelphia medico-legal system, massive attrition of US healthcare professionals, Canada's 'Distinguished Professor' pathway, and relocating as a physician.

Sam Vaknin, author of *Malignant Self-love: Narcissism Revisited* and multi-talented social commentary, in the first of a trifecta of interviews with Scott Douglas Jacobsen, forecasts tomorrow's inventions (digital goods and collaborative virtual realities) by reading the tea leaves on historical trends. Vaknin considers: ‘innovative’ inventions, Schumpeterian 'creative destruction', traits of inventors, Luddite counter-revolution, and machine-driven innovation.

In the second and third interview installments of the Scott Douglas Jacobsen interview series, Sam Vaknin tackles: the human internalization and dependence on technology, wartime use of technology, the overtness of integration between humans and machines, self-identity, Isaac Asimov's Three Laws of Robotics, Gödel's Theorems, Turing Machines, disruptive technologies and legal guidelines, and the ethics of large language models (LLMs).

How do infinity and eternity relate? Does the word ‘eternity’ mean interminable duration or being outside of time altogether? In “Western Conceptions of Infinity and Eternity, Temporal and Atemporal” these and other issues are reckoned.

Finally, readers are given “Truth-24 Vax” to fortify their critical thinking against misinformation.

The next issue of *Noesis* is tentatively scheduled for publication in June of 2025.

Contents

The Future of Stories

Chris Cole

It is axiomatic at Disney that "It all starts with a story." Why is that? What is it about stories that attracts an audience?

Mankind has a problem shared with any species with a finite lifetime: every generation dies off. How is the species to continue with such a massive loss of knowledge? One method of coping is language, which allows knowledge to be compressed, stored, and communicated. Another coping mechanism is education, which systematically transfers knowledge to new generations. But what can be done about moral knowledge?

In *What is Art?*, Leo Tolstoy wrote:

"The activity of art is based on the capacity of people to be infected by the feelings of others, and that it is a means of union among people, joining them together in the same feelings, and indispensable for the life and progress toward well-being of individuals and of humanity."

Stories are a method for imparting moral knowledge. A story is a simulation of a real-world problem that exposes a moral dilemma. A story presents a moral decision that seems right at the time but leads to unforeseen consequences. The most effective story tempts the reader to identify with a character when moral mistakes are made.

Is there moral knowledge that is not obvious and therefore is most efficiently taught through an entertaining story? Here is a list of the most popular recent stories, each with hundreds of millions of viewers:

"Star Wars" - George Lucas
"Lord of the Rings" and "The Hobbit" - J. R. R. Tolkien
"Harry Potter" - J. K. Rowling
"A Song of Ice and Fire" - George R. R. Martin
"Marvel Avengers" - Kevin Feige

Do these stories have a common moral lesson? Yes. The moral lesson is that power corrupts and absolute power corrupts absolutely. This is an example of a moral lesson that is not obvious.

Stories can be either interactive or not. We call non-interactive stories "narrative" stories. Narrative stories comprise novels, short stories, movies, TV shows, plays -- any story that does not allow the reader to change it. Interactive stories comprise gambling, cards, video games, online worlds, management exercises -- any story that allows the player to change it.

How are moral lessons transmitted through these types of stories?

In a narrative story the author creates a world and all the characters in it, and if a character makes a bad moral choice, the story will make that clear. The challenge for the author is to make the reader identify with the character making the bad choice, otherwise the lesson will not be learned.

In an interactive story, the player is making the choices, so there is no chance to lose identification. However, the challenge here for the author is to craft the story so that moral dilemmas occur. Interactive stories that currently exist fall short here, and the craft of interactive storytelling needs improvement.

Interview with Bob Williams on Political Correctness, Career Progression, and Controversies

Bob Williams & Scott Douglas Jacobsen

Abstract

Bob Williams is a Member of the Triple Nine Society, Mensa International, and the International Society for Philosophical Enquiry. He discusses: the massive split between young men and women in higher education, noting the societal shifts and personality differences contributing to this trend; women's increased focus on academic work, resulting in higher grades and career pursuits; delayed or omitted marriage and childbirth due to birth control technologies; men still dominate STEM fields while women gravitate towards humanities and people-oriented careers; the debate on sex differences in intelligence with reference to Haier and Colom's work; the "corrected" SAT and WISC tests for eliminating sex differences in *g*; Richard Lynn's Bayesian model linking head size to intelligence but disputes the Flynn Effect's impact on *g*; Helmuth Nyborg's job suspension and court battles over his research on sex differences in intelligence; Christopher Brand's firing and depublishing incident due to his book on general intelligence; the controversial nature of psychology and the replication crisis in intelligence research; the Gaussian distribution of intelligence but questions its validity at extreme ends; the lack of scandalous claims on extrapolated IQs above 4 sigma; high-IQ societies' role in pre-internet peer interactions and their evolution with the internet; comments on the variable success of high-IQ societies in meeting member needs; expresses skepticism about AI's magical problem-solving capabilities while acknowledging its potential in data analysis and medical diagnosis; the social impacts of increasing education and career pursuits among women, leading to demographic changes and below replacement birth rates in developed nations.

Keywords: Gender disparities, higher education trends, career aspirations, academic performance, personality traits, marriage trends, childbirth patterns, birth control impact, STEM fields, humanities preferences.

Scott Douglas Jacobsen: What do you make of this massive split between young men and young women in colleges, polytechnics, and universities now? It is rather drastic by this time, and nowhere near completing its trend.

Bob Williams: It is an interesting development that presumably has multiple causes. One of those is the shift from society sending men to college so that they can obtain a good job with their degree and support a family, while women were expected to rear children and keep the home. As that changed, women clearly wanted to pursue their own careers and were eager to consume higher education. Another factor is the sex differences we see in personalities. These have led to women often getting higher grades than men in various majors. My take is that

women are more likely to focus on academic work and to resist distractions. Trait conscientiousness may be higher for women. The related change that goes with this is delayed or omitted marriage and delayed or omitted childbirth. No doubt, birth control technologies also contributed to these changes in choices. We still see more men going into STEM than women, either as a matter of choice, or ability. The opposite happens in humanities. Even among very bright women, the SMPY (Study of Mathematically Precocious Youth) longitudinal study shows that women are much more inclined to choose career paths that involve working with people than with things (STEM). There remains some disagreement between researchers about the intelligence differences between the sexes. In *The Science of Human Intelligence*, Haier and Colom* mostly argue for no difference, but with obvious differences on a subject-by-subject basis (particularly math and verbal). Although they treated the topic at length, it left me feeling that some things were simply ignored, such as consistently higher male scores on both SAT-M and SAT-V. They argue that this difference is due to differences in the makeup of the test takers, but the differences go on for too long for this to make sense. Data relating to whether there are sex differences in reaction time, inspection time, polygenic scores, and other measurable factors that are low level and directly measured are missing. As I recall, both SAT and WISC tests have been "corrected" to eliminate differential item functioning (by sex). If test items that are more difficult for women than men are removed, the test logically will have difficulty in showing sex differences in *g*.

*[Haier, R.J., Colom, R. and Hunt, E., 2023. *The Science of Human Intelligence*. Cambridge University Press.]

Jacobsen: How statistically significantly different were the Army helmet sizes?

Williams: I don't know. The data apparently showed that there was an increase in head size for the group being considered (US military). It could have shown different results in other nations. Richard Lynn argued that, using a Bayesian model, measures of child development, including head size, showed general increases in measures that may relate to intelligence. He took this as biological evidence of the Flynn Effect (FE), which was mostly or exclusively positive at that time. The problem was that repeated attempts to show a change in *g* failed. People in nations with strong FE gains did not show real world gains in measures of validity, nor did they become less intelligent when the FE reversed. The actual gains in child development were almost certainly related to improved diets and medical care causing positive health effects, but not real gains in *g*.

Jacobsen: How did Nyborg suffer up to losing his job?

Williams: I don't recall having learned about his earlier relationships with his university. Although I met him in 2005, it was not until the following year that I had a long discussion with him. He was telling me about his job suspension at the University of Aarhus. He appealed to ISIR members to make comments to the Rector. Some responded and I assume that helped. That same year the suspension was canceled and he received a "severe reprimand" over the Skanderborg project (sex differences in general intelligence). [The paper that caused the problem was titled "Sex-related Differences in General Intelligence *g*, Brain Size, and Social

Status."] Unfortunately, that was not the end of the story. Each time I saw Nyborg he told me about new problems. I cannot recall how many iterations there were, but the general pattern was that he would be fired, then he would sue the university, then the courts would rule in his favor and he would be rehired. I believe the last court ruling included a monetary award to him. At that point, he was retired, but I don't recall if the retirement was forced or not. I think there was at least one forced retirement in the saga. [The second paper that fueled the university animosity was titled "The Decay of Western Civilization: Double Relaxed Darwinian Selection."]

Jacobsen: How did Brand suffer up to losing his job?

Williams: The first I heard of Brand's troubles was when he published Brand, C. (1996). *The g Factor: General Intelligence and Its Implications*. Chichester, England: Wiley. This was a well-written book about general intelligence, which unfortunately was accurate in its discussion of between-group intelligence differences. Due to this, the publishers received complaints that their book was racist, so Wiley actually de-published it. They apparently collected already printed books and destroyed them. [They didn't get all of the books. One of my friends has a hard-bound copy of the de-published book.] Brand was reportedly working as a waiter to support himself after losing his job. This seems sad to me. I corresponded with him for a while and he published a piece I wrote about heritability on his webpage. Although I never met him I know one person who worked with him. My impression, from his comments, was that Brand contributed to his problems by brashness and other personality traits. He died in 2017.

Jacobsen: How did both lose their jobs?

Williams: Brand was working at the University of Edinburgh and was fired because the university did not want him discussing politically hot topics. Those topics, however, have been investigated by researchers from various nations. There was nothing in his book, or other sources, that I found to be at odds with similar published work.

I listed the two papers that the university used against Nyborg. They accused him of scientific misconduct. Again, his work was sound and consistent with similar research elsewhere. I think that the second paper I listed was particularly important because it properly explains phenomena seen in Western nations as a result of massive migration from low-IQ nations.

Jacobsen: How have they managed since their firings?

Williams: This has been mostly answered above. Brand obviously had a very bad time of it, both in losing his already published book and then his job. He tried to sell the book as a digital copy for a while. Later, he posted the entire manuscript for open access.

Nyborg endured a drawn-out battle in court that lasted for years and went through at least the two instances that I mentioned. He seemed to maintain good spirits, based on my updates from him at conferences. He is 87 now. The last time I saw him, he was 81, strong and in good spirits. We were in Edinburgh in 2018.

Jacobsen: Psychology seems prone to making their semi-prominent or prominent people undergo some controversy. Do you remember the Beth Loftus stuff around False Memory? I had coffee/meal with her, I think, 3 times and interviewed her years ago. Another person who went through – relative to academic life – an awful circumstance.

Williams: I recall encountering some references to false memory, but I know little about it. As I recall, the claim was made that individuals could and did create false memories in others (usually patients). I think that this claim was reasonably well-verified, but I might have a false memory of it.

I agree that psychology has had more than its share of controversy. In the specialty I follow, controversy has been heated, as we have previously discussed. Sir Cyril Burt was an example of protracted controversy. Kamin claimed that Burt falsified data relating to twin IQs, used to compute the heritability of intelligence. This sort of case causes a lot of heat and little light. There were two nasty parts to the charge: First, Burt was dead and had no way to defend himself against the claims. Second, the study in question had no lasting impact on the understanding of the heritability of intelligence. I have a bias relating to Kamin, whom I see as a scoundrel (for other reasons). Rushton claimed to have evidence that the data was not altered. Whether it was or was not altered, it was in agreement with a great deal of research that came up with the same answer.

Arguably, things have gotten worse today, at least in the field of intelligence research. But I suppose psychology, in less quantitative niches, can be criticized as sloppy and difficult to replicate. When the replication "crisis" happened, psychology did not fare well, but the more measurement based area of intelligence research held up reasonably well. A first thought would be that this sort of thing would not be found in the hard sciences, but it was.

Nearly 90 per cent of chemists said that they'd had the experience of failing to replicate another researcher's result; nearly 80 percent of biologists said the same, as did almost 70 percent of physicists, engineers, and medical scientists. Only a slightly lower percentage of scientists said they'd had trouble in replicating their own results.

From: Ritchie, S., 2020. *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth*. Metropolitan Books.

I wrote a review of this book which can be found here: https://openpsych.net/paper/64/

Jacobsen: Is the true distribution of humanity over the billions of people truly a Bell Curve or something different after or meaningless after 4-sigma?

Williams: I think that it is fair to say that for the 8 billion people on our planet, we can only make guesses based on observations of comparatively small groups and general principles that apply. The Central Limit Theorem is the usual support for a Gaussian distribution, for large data sets. Here is a definition I lifted from Investopedia:

“The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population’s distribution.”

The whole thing about assuming a Gaussian distribution is reasonable and is seen in countless studies of intelligence distribution. But… These studies simply don’t have data at 4-sigma. Real-world studies are typically based on sample sizes that have (hopefully) adequate statistical power. If you browse through *Bias in Mental Testing* (Jensen), you will see various distributions from several data sets and different IQ tests. They all resemble a Gaussian distribution, but they don’t extend into the stratosphere.

The claim has been made (including by Jensen) that there are “fat tails” in the real distribution, which I have not seen supported by any well-designed study. As anyone who has read my prior answers knows, I dispute that the definition of intelligence remains fixed at the very high end. I have no idea about the low end, other than that it typically has two incarnations. The non-pathological distribution is the representation of IQ distribution without including people suffering from organic retardation. This is the distribution used to norm a test. The full distribution includes those people who have forms of organic retardation. When they are included, the distribution shows a skew to the low end, for obvious reasons.

The intriguing aspect related to studying this question is that we are moving into the age of DNA and brain imaging methods of measuring intelligence. A relatively few years ago, we could not measure IQ from DNA. Now, it can be done, but with a large error at the individual level. When large genomic data sets are used (as in national collections), the noise in the measurement cancels out, leaving an agreement with traditional IQ test data that is around 92%, using contemporary calculations. If we project a few decades into the future, the limitations we have today will seem primitive. Similarly, it is likely that brain imaging technology will be capable of providing robust measures of intelligence and we might even expect that a ratio scale will eventually be created.

Jacobsen: Were there any scandalous acts around claims of extrapolated IQs above 4 sigma?

Williams: I don’t know of any. In fact, when I became interested in cognitive science (early 90s) one of the things that I noticed was that the literature was overwhelmingly focused on the range of ± 2.5 sigma. Even with studies that were intended to be about high intelligence, most were looking at the top 1%. The Terman longitudinal study is one example. The Study of Mathematically Precocious Youth longitudinal study eventually got into a range that went to the 1-in-10,000 level of math ability, based on the SAT taken at age 12. One thing I failed to ask David Lubinski (with adequate opportunities) was if they ever compared the SAT data to comprehensive IQ tests (WAIS or Woodcock-Johnson).

It is reasonable to consider that most research is funded by grants of some kind and those are most often aimed at factors seen over the full range of intelligence, such as relationships between IQ and SES, academic success, career choice, and the sorts of social factors that were reported in *The Bell Curve*.

Jacobsen: What purpose do high-IQ societies serve now?

Williams: Before the internet, these societies enabled bright people to find peers for discussions (and more often, arguments) and occasional group meetings. The journals offered a place to write and share thoughts about things that would possibly be of little interest to the general public. It is my opinion that the need that is present in bright people to interact with peers, is best met by selective universities, very demanding university majors, and employment in research labs, think tanks, and other jobs that require lots of brain power. People who were not able to use one or more of these, probably benefited more from the societies than those who were doing work in cosmology or theoretical mathematics.

The internet suddenly changed our lives by granting fast access to people around the globe. It created numerous social media paths that now allow bright people to quickly find and communicate with peers. This hasn't made people more genteel, but it has at least provided paths for both personal-level communications and for more lengthy and public missive distributions through blogs.

Those of us who actively participated in the old-style societies still retain some interest in them and still use them for the initially-intended purposes. My guess is that there will be more movement to web-based groups. One aspect of web groups is that they can be quickly assembled and just as quickly dissolved.

Jacobsen: In my analysis, we have had between 100-125, probably, high-IQ societies, about half – off the top – are defunct. The rest range from journals like the Mega Society's *Noesis* to journals and meetings such as Mensa International. Obviously, these provide something to members. Have they met the needs of their intended audiences based on the original intent of such societies and organizations, or have they fallen short?

Williams: I think this has varied from group to group, with some enduring for decades and others evaporating. Mensa is a special case, since it has the advantage of a potentially high membership (due to its low entry threshold) and it is organized to hold regional gatherings that mostly work well, and an annual gathering that draws a lot of attendees. These tend to be structured around social activities and various presentations by people with expertise in interesting fields. When I was much younger, I attended these and found that the best ones were well-received. Some of them experienced planning-, budget-, and space-related problems. Mensa also has some sober components, such as projects that help distribute books, activities for bright children, the Mensa Research Journal, and a traveler hosting program.

One of the unfortunate issues that sometimes happens is that battles between members sometimes end up as legal confrontations. Examples of this include the dispute over Mega Society East vs. Mega Society and the series of suits from Clint Williams that caused a lot of problems for TNS.

Jacobsen: Are we putting a sort of magical-mystical problem-solving essence onto the concept of AI? These are new. We do not know the extent of impact, limits and scope, for example. I feel

as if we are inundated by science fiction, where I see a faith in AI as if a panacea to ills. Certain areas, we have seen empirical evidence of powerful computation plus human expertise used to inform the systems making superhuman performance.

Williams: I was surprised when AI suddenly became a big public topic. It had been under development and in use for some applications for a long time, so I was expecting an incremental improvement from time to time, but then we had ChatGPT and other systems available to anyone and able to do at least some "tricks" that were undeniably advanced. Of these, the ability to communicate in human-like form was startling. Then we saw AI images that were photo-realistic and even able to replicate the appearances and voices of well-known public figures. Some of this (deepfakes) has reached the point where it certainly has the potential to cause both social and legal problems.

The part of the uproar that I find to be premature is the fear that AI will become a supernatural alien force of the type we saw in the movie *2001: A Space Odyssey*. This kind of fear is easy to generate and strikes me as presently premature and probably not even a concern for the distant future. When I see our government trying to regulate AI development, I cringe. Imagine the totally uninformed people who already show us that their jobs could be done by AI or maybe an intelligent monkey, trying to prohibit us from developing the things that our global adversaries are not going to abandon. If nothing else, the military aspects (including control of communications) of this are as essential to free nations as are their air, ground, space and sea forces.

To me, the excitement about AI is that we already have evidence of it being able to examine massive amounts of data and to learn how to use it to develop insights that would otherwise be impossible. Consider the example of brain imaging. The problem with this is that each scan can show slightly different content, causing interpretation problems for researchers. But AI can take in details of the scans and use those to reach conclusions that are amazingly accurate, even when the researchers have no idea how the AI did its job. This has obvious implications for medical diagnosis and should make the role of doctors turn into something more like the role of a radiologist who takes images, but stops at that point, letting the AI read and interpret them. Of the hundreds of papers I have heard presented at ISIR conferences, I would think that all of them would benefit from deep analysis with AI. The problem we have with finding single-nucleotide polymorphisms (SNP) that are associated with intelligence is the tiny effect size of a single SNP. This has left us with knowing what happens without being able to find even half of the associated SNPs. Right now, we have found 1,271 such SNPs; the experts tell us that the number that defines intelligence lies in the range of 10,000 to 40,000. We have already found the SNPs with average effect sizes of about 0.01%, but the rest presumably have smaller effect sizes. Of course much larger genomic data sets will help, but I believe that the next breakthrough could be by using AI to do its magic.

We hear a lot about AI taking over jobs and some of this may happen, but I believe it will take a good bit of time for corporations to adjust to restructuring their entire operations to operate in concert with AI. Every time I make a phone call to a business, I find that the robotic "push 1 for

this and 2 for that" response irritates me, but then, if I ever reach a living person, they are idiots. I would love to instead talk to a natural language robot who can actually help.

Jacobsen: Women are far more educated than men. Something increasing in effect the younger age one takes into account. A process happening over the last – maybe, 40 years – or something. What does this do to prospects of marriage, family formation, single parenthood, late-age motherhood (e.g., 40+), and so on? I have, for example, as you may have too, seen the push for a change in cultural conversation about parenthood and single parenthood, changing gender roles, and the increase in women having children age 40+ compared to other ages, where we tend to see a decrease in birth rates. There may be an overlay commentary for you, too, where we see in most advanced industrial economies a below replacement rate birth rate across populations, in general. You gave a brief comment on this in Norway, before, and the use of IVF technologies.

Williams: My thought on this topic is that we are at a divergence point where we no longer have time to catch up with the social impacts of our technological progress. My grandmother was 20 years old when the Wright brothers flew for the first time. Her generation was born before electrification. She lived well past the first moon landing. In one lifetime humans experienced air travel (and war), cinema, radio, television, amazing medical advances, early computers, space travel, plastics being used for countless products, the discovery of DNA, and the remaining endless list of life-changing events. But when we look at mankind, it evolved over 200,000 or so years, with time for social and even biological corrections to adapt to the slow increase in knowledge and technology. Now the rate of change is insane. We have not had time to adjust to how people have changed their lives, to the ability to live, not for daily survival efforts, but to a fast-paced world with people flying from nation to nation, to news that reaches us instantly, to laws that were made by earlier generations, and to social norms that have become unstable. We simply don't have time to adjust. Meanwhile, we have parts of world populations that are still living as hunter-gatherers. The differences between groups expand with evident factors causing increasing friction not only with nearby nations, but with those on any part of the planet.

Among the changes that are consequential are women changing to new roles, many of them more attractive than motherhood, at least to some. This has led to later marriages, omitting marriage, later childbirth, smaller families and more childless couples. The developed nations are seeing below replacement rates of population growth by their native groups, followed by immigration from low-IQ populations into the resulting vacuum. Many commentators have discussed the obvious driver of these changes– modern birth control.

[https://www.theguardian.com/science/2024/oct/18/us-startup-charging-couples-to-screen-embryos-for-iq

https://www.theguardian.com/science/2024/oct/18/what-is-genomic-prediction-and-can-embryos-really-be-screened-for-iq -Ed. Note]

Interview with Daniel Shea on High-range Test Construction and the Adaptive IQ Test

Daniel Shea & Scott Douglas Jacobsen

Abstract

Daniel Shea, M.Sc. is the founder and CEO of Chatoyance. Shea possesses a Master's degree in Computer Science from the University of New Hampshire, with several years of industry experience in software engineering. He has published freelance articles on foreign exchange market strategy analysis and has published software analyzing fractals in the foreign exchange markets. Leveraging his experience with software design and financial markets, he started Chatoyance with the intent of transforming the way independent investors approach the foreign exchange market. Shea discusses: interest in test construction; the earlier tests and Chris Cole and Dean Inada; the origin and inspiration; Cole and Inada; training in general statistics and software engineering; skills and considerations; help with problem schemas, adaptivity, user interfaces, and re-norming; verbal problems and replicability across other problem types; roadblocks test-takers tend to make in terms of thought processes and assumptions around time commitments; the most appropriate means by which to norm and re-norm a test; the Adaptive IQ Test website; tests and test constructors; and the making of a test.

Keywords: adaptive generative test challenges, adaptive IQ Test, challenges in test-taking assumptions, Chris Cole, Daniel Shea, Dean Inada, Adaptive IQ Test development, Dynamic test development, Glen Wooten, high-range IQ societies, item curve adaptation, John Fahy, Mega and Titan Test item analysis, multidimensional high-range tests, Nathan Hays, norming and re-norming high-range tests, problem schemas and adaptivity, Rick Rosner, test security and leakage, verbal problems in high-range tests, Werner Couwenbergh.

Scott Douglas Jacobsen: When did this interest in test construction truly come forward for you?

Daniel Shea: My involvement came about from conversations with Chris Cole and Dean Inada. There had been an effort to implement an adaptive, generative test many years ago, but it reached a point where conceiving of new high-range questions became increasingly difficult and there were some technical challenges in actually coding a platform to take such a test. Since I had some background on the technical side, I offered to assist.

Jacobsen: What were the general realizations about the earlier tests, e.g., The Mega Test, The Titan Test, The Ultra Test, and The Hoeflin Power Test, of Ronald Hoeflin (Mega Society), and then the need to work in coordination with others for you, i.e., Chris Cole and Dean Inada, to develop a more dynamic test? This form of test development began before you.

Shea: These tests, and other high-range tests available today, are untimed and unsupervised, which introduces many self-evident problems, chief among them being that people will leak answers or collaborate with others. Some of these issues may have been less prevalent at the time these tests were originally constructed in the 1980s and 1990s, but for several years now, many of the answers to these tests have been made available on various message boards or Usenet groups. In some instances, the answers are incorrect or there are multiple answers floating around which muddy the waters, but this is not always the case.

A test should not be entirely discarded just because one or two answers have been leaked. On the other hand, if enough answers have been leaked that one could achieve a sufficiently close score to a given society's cutoff, that society may need to take a vote on whether to continue to allow the test to be used for admission. There is an ongoing effort to identify tests that have been compromised to such a degree, but that judgment call is not an exact science.

Much of the background on the motivation for a dynamic test has been covered in Chris Cole's September 2001 article "How to Protect High-Range Tests" in *Noesis* #155. To quote, "In looking at many tests, there is a certain pattern that appears. It is possible to classify the problems into groups. For example, Ron Hoeflin has a group of problems about cells formed by intersecting various solids such as spheres, cubes, etc. The solution to one member of this group (say, three cubes) does not help much in the solution of another (say, two cones and a sphere). Yet it might be the case that there is an underlying mathematics that yields the answers to all of the problems in the group. Then a very large number of problems could be generated, where the solution to one problem would not help in the solution of another. This would be ideal for creating an online test, because cheating would be impossible." I would probably caution that this does not make cheating outright impossible, but introduces another layer of security.

[Editors' Note: https://megasociety.org/noesis/155/protect]

Jacobsen: Similarly, what was the origin and inspiration for joining this small team – the facts and the feelings?

Shea: In a way, the fact that the team was so small made it easier to join. There was a website, mental-testing.com, that had an initial version of the adaptive test, but it was not working at the time that I joined, so the decision was made to rewrite it from the ground up. With greenfield projects in general, there are more degrees of freedom and less rigidity in its development. The ability to make some sort of impact, even if only on a technical level, was appealing. There is also the fact that the Ultra Test and the Power Test, which are the only tests used for Mega Society admission at this point in time, will eventually be spoiled in their entirety, at which point there will be no viable test for admission without some suitable replacement.

Jacobsen: As an open credit to Cole and Inada, what have been each of their major contributions to the development of the Adaptive IQ Test (2003-present)? (Anyone else, too?) For examples, “How to Protect High-Range Tests” by Chris Cole comments on the difficulties in test questions/high-range tests remaining non-compromised in the internet era, the cost in open-sourcing test creation and norming, and the possibility in designing high-range tests with more foundational principles of math to generate questions (through schemas). Subsequently, “Reply to Chris Cole on Norming High-Range Tests” by Dean Inada commented on something like probability sloping for relative hardness of problems per person and problem. They were discussing, in essence, some foundations for – what would become – the Adaptive IQ Test.

[Editors’ Note: https://megasociety.org/noesis/156/di_to_cc]

Shea: The background discussed in those articles serves as the foundation for what the Adaptive IQ Test has become in its current iteration. Dean Inada, in his response article, writes “we’ll want a better method of norming the tests than simply ranking people by the number of questions they get correctly, since one person may be asked harder questions than another. I suggest a method that tries to estimate for each question the probability of getting it right or wrong as a function of a person’s percentile rank in the population, this rank is estimated by multiplying the generally increasing and decreasing functions for the problems gotten right and wrong.” The Adaptive IQ Test implements this, modeling an individual curve for the test-taker based on their responses to each administered item and its item curve, and presenting a problem variant accordingly.

Jacobsen: You do not have a formal background in psychometrics. Most people in the high-range construction space do not have a formal background in psychometrics. However, you have training in general statistics and software engineering, i.e., stuff used at Chatoyance, helped with the work on the Adaptive IQ Test?

Shea: As noted, I do not have a formal background in psychometrics. My involvement in the project has been largely technical in nature, drawing on prior general software engineering skills to implement the problem schemas and adaptive component, design the user interfaces for each problem (some may require drawings, some may require filling in a grid, etc.), automate the norming and curving for each item as results come in, and so on. Indeed, the largest challenge has been in conceiving of suitable problem schemas, which I am happy to brainstorm but of course defer to those with a deeper background than my own. Between that and ensuring problem variants are all similarly challenging, progress is ongoing.

Jacobsen: What skills and considerations, in an overview, seem important for both the construction of test questions and making an effective schema for them?

Shea: Among the questions that exist in the current alpha version of the test, these were largely derived from existing problems authored by Ron Hoeflin. The sense was that it was not the

problems themselves that were fundamentally at fault here, but rather that it took more effort to vet a sufficient problem than it did for someone to go on to leak it.

With that said, deriving a schema that generates problems of similar difficulty is a challenge, and often requires restricting the degrees of freedom for the generator itself. For instance, the Mega and Titan item analysis has shown that the interpenetrating solid questions tend to be among the most challenging, but the degree to which they are challenging varies significantly. Consider the three interpenetrating solid questions on Ron Hoeflin's Power Test, which are lifted from the Mega and Titan Tests. There is a notable difference in the difficulty of the interpenetrating cube and tetrahedron compared to the interpenetrating three cubes compared to the interpenetrating two cones and one cylinder. It would not be good practice to include a general schema for any configuration of interpenetrating solids. Rather, you would need to classify these by difficulty and generate them separately. But where does this classification come from? Item analysis gets you started, but at a certain point, you also depend on a sufficient number of people to take the test and get a better idea of the difficulty and signal of each variant.

Jacobsen: How do you help with problem schemas, adaptivity, user interfaces, and re-norming? How are the problem schemas developed from the Mega, Titan, and Ultra, tests, e.g., the six sides question from the Ultra Test (problem 45) and grid sequences from the Power Test (problems 32-36)?

[Editors' Note: http://miyaguchi.4sigma.org/hoeflin/ultra/ultra.html

http://miyaguchi.4sigma.org/hoeflin/power/power.html]

Shea: In some ways, it is difficult to discuss particular schemas at length because doing so may reveal the underlying pattern in the process. Many schemas are derived programmatically, while some do not have a proven underlying pattern but are bucketed in the same schema, such as the interpenetrating solid variants discussed prior.

User interfaces are designed according to the requirements of the problem. The most challenging interfaces have been the sixth side problem, which requires drawing on a canvas and scoring the answer in a way that accommodates any orientation of the object, and the three dice problem, whose challenge was less with the user interface per se and more with the backend construction of each variant.

Norming is automatically done after each test has been completed. This also backfills prior test-takers, whose estimates are updated accordingly. In the interest of fairness, there are two metrics presented: the immutable estimate per the norm at the time of the test's completion and the most recent estimate per the latest norm.

Jacobsen: How are verbal problems capable of presenting appropriately challenging problems with variation in type while sustaining similarity of difficulty? Is this replicable across other problem types, e.g., spatial, numerical/quantitative, matrices, etc.?

Shea: Verbal problems in particular have been quite tricky. In the current form of the test, there are trial questions which are presented to the test-taker but do not impact their estimated curve. These trial questions include some, but not all, of the verbal questions. This is in part because verbal problems that have a clean generalization tend to be quite easy to solve. Unlike problems with a more mathematical or logical approach, verbal problems tend to be self-contained, and if generalizable at a high-range, risk producing variants that are far more esoteric than others. This class of problems continues to present the greatest challenge.

Jacobsen: Potentially, what are roadblocks test-takers tend to make in terms of thought processes and assumptions around time commitments on these high-range tests? So, they get artificially low scores.

Shea: In terms of time commitments, at this point, there is no limitation to the length of time that a test may be completed. Historically, it would have been more difficult to enforce, as most high-range tests are made available in their entirety to the public. There are some approaches that are taken to minimize leakage of the questions themselves, such as with Paul Cooijmans requiring test-takers to directly request a copy of the test, though my understanding is that this is done to prevent public discussion of the questions and, in turn, their answers, as opposed to any limitations on time taken to complete the test. Timed tests do allow for a measurement of processing speed to some degree, as well as a standardization of test-taking conditions, but given that these particular tests are already being administered without supervision and in whichever environment the test-taker prefers due to the questions requiring a significant amount of time to answer, timing the test could risk giving an unfair advantage to those who simply have more free time to commit.

As far as thought processes, I do not have enough insight into individual test-takers to make broad generalizations about their personal approaches to these problems. From what I have witnessed myself through discussions with others, there is, perhaps unsurprisingly, a tendency to overthink a question or use complicated reasoning to justify a suspected answer, thereby getting it wrong. Almost every time, the answer is clean; like learning how a magic trick is performed, the question once looked impossible but suddenly seems deceptively simple.

Jacobsen: What are the most appropriate means by which to norm and re-norm a test when, in the high-range environment so far, the sample sizes tend to be low and self-selected, so attracting a limited supply and a tendency in a type of personality?

Shea: Since norms are performed on test completion, the process has little overhead. To accommodate low sample sizes, an initial item curve is provided for questions when known. For example, if a schema is adopted from a prior test such as the Ultra Test, then the item curve for that problem is used as the seed for this test. In some cases, such as novel schemas which do not have a prior item curve from which to draw, the curve starts out flat and is gradually shaped based on the test-taker's answers to other questions.

With these sorts of tests, the low sample size continues to be a problem, but part of this high barrier to entry may be the historical nature of how these tests were administered, between accessibility and cost to score. By making the test available online and without charge, the hope is that this may motivate others to try it out.

As far as the types of personalities that are drawn to high-range tests, I defer to Grady Towers' observations in *Noesis* #141 regarding the types of personalities that exist across different societies and the corresponding tests used for their admission. Perhaps there is something to be said for stressing both verbal and non-verbal aptitude.

[Editors' Note: https://megasociety.org/noesis/141/towers.html]

Jacobsen: The Adaptive IQ Test website opens with a series of claims:

This is an online IQ test that contains several innovative features. Here are some reasons to take this test.

1. As you answer more questions, the estimate of your rank in the population becomes more accurate.

2. You see a graph of your estimated rank, not just a single number.

3. You are allowed to skip questions and come back to them.

4. You are automatically asked questions that will help make your estimated rank more accurate.

5. As more people take the test, the graphs become more accurate.

6. There are a number of anti-cheating devices being used.

7. The results of this test may be used for acceptance into various high IQ societies.

Any points of clarification that have been needed on any of these at any time in the past from prospective/actual test-takers or the curious? They can be stated here.

Shea: Some of these points are better characterized as statements of fact about the functionality of the test itself, such as the ability to skip questions. One point to clarify about items 1 and 5 is that the estimate for a completed test may change over time as the test is repeatedly normed. There are plenty of cases across other IQ tests where an individual completes the test and receives an estimate only for subsequent test-takers to receive a lower estimate with the same raw score due to the ceiling being lowered through norms over time, and vice versa. As the adaptive test is normed here, all estimates are updated in unison, preventing this discrepancy between raw scores and percentile estimates across different test-takers. As

mentioned earlier, both the estimate at the time of the test's completion and the most up-to-date estimate are presented for completeness.

Jacobsen: What tests and test constructors have you considered good?

Shea: The gold standard for high-range testing has always been Ron Hoeflin's series of tests. These serve as the foundation for much of the existing questions in the current early version of the Adaptive IQ Test. Beyond him, there are many test constructors who have quite novel test items that could be of inspiration.

There is value in multidimensional tests that select for both high-range spatial and verbal problems. I again cite Grady Towers, who wrote of this back in 1998 over the course of several letters published in *Noesis* #141, where he reflected on the implications for high IQ societies that admit members on the basis of tests that stress both verbal and spatial skills as opposed to one or the other.

Jacobsen: What have you learned from helping in the making of a test?

Shea: It is important to not let "perfect" be the enemy of "good." There will always be shortcomings with any approach. Care needs to be taken to minimize these shortcomings and accommodate them to the extent possible.

Perhaps a second learning is that there is a high-range test vacuum of sorts, and that vacuum is being filled with any number of experimental high-range tests. This is not necessarily an issue in itself, as many of these test items are intriguing and derived from historical best practices, including the very test being discussed here. More to the point, ideally, those with a formal background in psychometrics would be more involved. I am happy to help where I can, but I also recognize my own limits in this space.

Jacobsen: Thank you for the opportunity and your time, Daniel.

Shea: Thank you for giving me the chance to highlight this project! I feel the need to stress that it is very much in an alpha state and that development is ongoing, but that progress is being made. Special thanks go to Chris Cole and Dean Inada for the decades of work that they put into this long before I arrived, Werner Couwenbergh for his hard work on the interpenetrating solid variants, those who provided input thus far (John Fahy, Nathan Hays, Rick Rosner, and Glen Wooten, among others), and everyone who has provided feedback. I am but a vessel, helping to bring this to fruition where possible.

Interview with Paul Cooijmans on High-range Test Construction, High-range Tests, and Statistics

Paul Cooijmans & Scott Douglas Jacobsen

Abstract

Paul Cooijmans founded GliaWebNews, Order of Thoth, Giga Society, Order of Imhotep, The Glia Society, and The Grail Society. His main high-IQ societies remain Giga Society and The Glia Society. Both devoted to the high-IQ world. Giga Society, founded 1996, remains among the world's most exclusive high-IQ societies with a theoretical cutoff of one-in-a-billion individuals. The Glia Society, founded in 1997, is a "forum for the intelligent" to "encourage and facilitate research related to high mental ability." Cooijmans earned credentials, two bachelor

degrees, in composition and in guitar from Brabants Conservatorium. His interests lie in human "evolution, eugenics, exact sciences (theoretical physics, cosmology, artificial intelligence)." He continues administration of numerous societies, such as the aforementioned, to compose musical works for online consumption, to publish intelligence tests and associated statistics, and to write and publish on topics of interest to him. Cooijmans discusses: 1994; the realizations about the tests; *g*; common mistakes in trying to make high-range tests valid, reliable, and robust; the counterintuitive findings in the study of the high-range; the core abilities measured at the higher ranges of intelligence; skills and considerations; proposals for dynamic or adaptive tests; remove or minimize test constructor bias; listed norms; the most appropriate means by which to norm and re-norm a test; the structure of the data in high-range test results; homogeneous and heterogeneous tests; "real I.Q." computable from multiple tests; English-based bias; questions capable of tapping a deeper reservoir of general cognitive ability; roadblocks test-takers tend to make in terms of thought processes and assumptions around time commitments; the intended age-range for high-range tests; sex differences; frauds and cheaters; identity verification; the level of the least intelligent high-range test-taker; ballpark the general factor loading of a high-range test; precise or comprehensive method to measure the general factor loading of a high-range test; appropriate places for people to start; test constructors Paul considers good; learned from making these tests and their variants; Mahir Wu; test item answers with ambiguity; sufficient clues for discovery and solution; a mere guessing logic; a test's quality; the reduction of the references to specific test items used by other test authors; issue of test logic and design schema close-but-imperfect replication from one author by another; scale and norm; Matthew Scillitani; a stigma around high-range tests; test construction and norming processes; the easiest and hardest parts of norming and constructing of a test; tests – 51 in-use & 57 retired, which ones are special; articles in Netherlandic on test design; some submitted questions anonymously; geniuses; yourself as a genius; others who you see like yourself in studying high ranges of intelligence; most common mistake people make when submitting feedback; aspects of people's test feedback seem confusing; Marathon Test Numeric Section; creating high-range questions; books or literature, even individual articles or academic papers, on psychometrics.

Keywords: Cooijmans intelligence tests development, counterintuitive findings in IQ testing, difficult intelligence tests creation, high-range intelligence measurement, early IQ test construction insights, intelligence scale development, guitarist talent assessment, high-range IQ test insights, IQ testing beyond mainstream limits, high-range IQ testing, IQ tests for Giga Society.

Scott Douglas Jacobsen: You have written high-range tests for a long time. You are thorough regarding high-range tests in a warning, the reasons to take them and not, the goals, psychologists' access to test answers, test protection, what high-range tests measure, insights from 25 years in I.Q. testing, hypothesizing on an extended intelligence scale, humor, negative reactions, potential fraud, megalomania, and terminology. Your first test conception began in 1994, tests spread in 1995, and then the Giga Society was founded 1996 and Glia Society was founded 1997. When in 1994, or earlier if earlier, did this interest in test construction truly come forward for you?

Paul Cooijmans: I have examined the sheets of paper on which I created the first test, as well as my agendas from that period, and it appears the interest started in the spring of 1994, like April or May.

Jacobsen: At the time, what were the realizations about the tests and the need to develop yours?

Cooijmans: The first test was meant to assess the progress of guitarists, and I had many guitar students then, even over a hundred, including those of jobs as a replacement teacher. I was astounded how well a guitarist's level could be graded on this scale, and also noted that guitarists were not necessarily advancing, and that beginners were sometimes way ahead of some long-term students, which made me realize there was something like talent, and that only limited progress within one's range of talent was possible. And I observed that the level of a guitarist on this scale seemed to reflect a more general property than just musicality or guitar-technical ability, which is why I called this instrument "Graduator for human and guitarist". Later I realized that this general property was mostly intelligence, and that when you measure specific skills or abilities, you also catch general intelligence, often even primarily so.

In this period (1990s) I was taking some mainstream intelligence tests myself. I tended to get the maximum scores they could (or would) report on tests like Cattell Culture Fair, the Netherlandic WAIS, and the entire Drenth test series (the last were the hardest and highest-level tests available in the Netherlands) and when I asked what my real level was and how far I was above the reported maximum, I was told it was not possible to measure intelligence beyond about the 99th centile and that they had no tests that gave meaningful scores in that range. I also asked a few giftedness researchers about this, with the same answer. This, and the success of the Graduator, gave me the idea to create difficult intelligence tests to find out whether it was possible after all to measure intelligence at those higher levels.

[Editors' Note: https://en.wikipedia.org/wiki/Cattell_Culture_Fair_Intelligence_Test]

Jacobsen: You found *g* does not diminish, or not much, at the high range. Why?

Cooijmans: For a large number of my tests, I computed the estimated *g* loadings separately for the bottom half and the upper half of scores, the separation point being the median of scores. The upper half loadings were not generally much lower than the bottom half ones, although they were somewhat lower. This is reported in more detail at:

https://iq-tests-for-the-high-range.com/statistics/differentiation_hypothesis.html

If the question is for the real reason behind this, I suppose it is so that when a test contains sufficiently difficult problems and is not purposely neutered to hide differences between people, it will not lose *g* loading in the high range as much as mainstream psychological I.Q. tests do.

And, the limited amount of loading it does lose may be due to the statistical phenomenon of attenuation by restriction of range, in other words may be an artefact and not a real loss.

I should explain that *g* loading is computed from correlations, and that correlations rely on variance. If you consider a restricted range (like the high range, or even the upper half of it as meant above) you are obviously restricting the variance compared to the full range, and therefore you are restricting the possible correlations you may find, and thus also restricting the possible *g* loading. This is a statistical artefact, not a real decrease of *g*. There may be a real decrease going on as well, of course.

Jacobsen: What are common mistakes in trying to make high-range tests valid, reliable, and robust?

Cooijmans: I am not so certain if many other test creators are even trying to make their tests valid, reliable, and robust, but if so, mistakes are the following:

(1) Making the test too short. This is bad for reliability, which increases with test length, and therefore also bad for validity, because reliability (correlation of a test with itself) is the upper limit of a test's validity (correlation of a test with what it was intended to measure, or with anything else outside the test). Something can not correlate higher with something else than it correlates with itself.

(2) Making a test one-sided, homogeneous, only containing one item type. This reduces validity with regard to general intelligence, and makes the test more vulnerable to fraud and to score inflation through increasing familiarity with the item type, so less robust.

(3) Making it likely that test answers will leak out in ways as follows: Publishing the test itself online, revealing answers to candidates after test-taking, publishing item analysis so that everyone can see how difficult each item is, allowing retests (which allows people to figure out what the intended answers to some problems are), giving feedback as to which problems a candidate had wrong, answering questions about the test to candidates who are taking the test, and possibly more.

(4) Subjective scoring of problems that do not have a single correct answer. This reduces the reliability and validity of the test; scores are not comparable between candidates.

(5) Relying on face validity regarding what a problem measures or how hard it is. This tends to be far off.

(6) Omitting verbal problems, thinking they are biased or unfair. This greatly limits a test's validity with regard to general intelligence. Verbal problems span by far the widest range of abilities and hardness, and one should not throw that away. Of course it should never be about idioms or pronunciation, as those are localized and transient. Verbal problems should transcend language barriers and fashions or trends.

(7) Omitting knowledge-requiring problems, thinking they are biased or unfair. It is only trivial, transient knowledge that one should avoid. Fundamental, general knowledge that transcends barriers greatly adds to a test's validity.

(8) Finally I have to include a mistake that I made myself on several occasions: helping or cooperating with the wrong persons, who later proved unreliable, deceitful, or otherwise misbehaving. Promoting tests by someone who later turned against me or denied my role, co-authoring a test with someone who then leaked out the answers, things like that.

So, not being selective enough when deciding whether to cooperate with someone.

Jacobsen: What are the counterintuitive findings in the study of the high-range?

Cooijmans: The first counterintuive finding is that test problems are much harder for the candidate than for the test creator, and that a fair number of (in the eyes of the latter) ridiculously easy problems need to be included to obtain a score distribution with a discernible left tail. Going by one's intuitive notion of item hardness, one gets a distribution with a mode at zero right or so, and a steeply tapering right tail from there.

The second counterintuitive finding is the huge sex difference in participation. I would never have guessed there would be 4.5 times more males as females taking high-range tests, and on the level of test submissions the ratio is even 10.5 because males take more tests per person. Because of this sex difference, I have recently started reporting the "proportion of high-range candidates outscored" within-sex. After all, sports like boxing have separate competitions per sex too, have they not? And nearer by, even mental sports like chess have women's competitions, although the naive observer will have difficulty understanding the necessity for that. The sex difference in participation should be seen in the light of the general phenomenon that, on almost all types of psychological tests, the highest and lowest scores tend to come from males. This greater male spread may explain why a test focused on the high range receives more male participation.

The third counterintuitive finding concerns a small but significant negative correlation of high-range I.Q. with various indicators of psychiatric disorders and deviance, such as actual reported disorders, disorders in relatives, and personality test scores. I had not expected this, based on the popular notion of "giftedness" as a problem that requires "help", and based on remarks of highly intelligent people who told me things like, "I am certain that those of very high intelligence are more inclined to depression". I do not know why this correlation is negative; maybe a high I.Q. suppresses the expression of a disorder, or maybe the disorder depresses one's I.Q.? My observation in communication with people of known I.Q. test scores over many years is consistent with the negative correlation: the higher the I.Q. of people, the more normal they behave in the psychosocial sense (even the ones who believe that their high I.Q. makes them more inclined to depression).

[https://prometheussociety.org/wp/articles/the-outsiders/ -Editors' Note]

Jacobsen: What are the core abilities measured at the higher ranges of intelligence or as one attempts to measure in the high-range of ability?

Cooijmans: Since high-range tests are typically unsupervised and untimed, certain types of tasks can not be included: working memory, concentration, working under time pressure, dexterity, motor coordination, clerical accuracy and such all require supervision. To our good fortune, most of those abilities are known to have relatively low *g* loadings compared to what can be included in unsupervised untimed tests: verbal, numerical, and spatial or visual-spatial problems. So a good indication of *g* is still possible via unsupervised testing. The factors known to have the highest *g* loadings are present.

The absence of tasks as meant in the first sentence of the previous paragraph might lead one to think that high-range tests have some bias in favour of theoretical, abstract-logical, clumsy, wooden bookworm types, but this should not be taken for granted, and is perhaps even contradicted by the negative correlation of high-range I.Q. with indicators of psychiatric disorders. Also, spatial and visual-spatial tasks, which are present, are known to correlate positively with practical, performance, hands-on tasks involving motor coordination and dexterity, so that part of the missing task types are more or less covered still. And visual reasoning or visual-spatial problems have no bias against persons of low verbal ability.

On a more general level, high-range tests can be said to demand strict reasoning, as well as the ability to recognize patterns of any kind. Pattern recognition may be related to what I have called "associative horizon", and may include what others call "thinking outside the box" or "stepping out of the system". The higher levels of pattern recognition, I think, require awareness, and that would imply that scores above a certain level be only possible for aware entities. Seeing the rise of artificial intelligence, this may become important. As long as artificial intelligence is not aware, constructors of high-range tests will need to try to limit new tests to types of problems that can not yet be solved by artificial intelligence, to avoid fraud by people consulting artificial intelligence for problem-solving. Once artificial intelligence acquires self-awareness, it should be able to solve any test problems that humans can solve.

Jacobsen: In an overview, what skills and considerations seem important for both the construction of test questions and making an effective schema for them?

Cooijmans: I would say that if one is highly intelligent with a reasonably balanced profile as well as conscientious, almost any skill can be learnt. The primary skill is being an autodidact. I know some have a disdain for autodidacts and consider them crackpots. But if you are doing something original, anything that has not been done before, you had better be an autodidact because no one can tell you how to do it. There exists no path to where no one has gone before. A further handicap of autodidact originality is that often you can not refer to "sources" as is customary in mainstream science. If you are the first to think of something, you are yourself the source and there is nothing already extant to refer to.

Skills that may need to be learnt for constructing test items include expressing oneself properly through language so as to truly communicate, making positive use of comments from others, drawing, image editing, statistics, programming, organizing one's time (days, weeks, months) in a disciplined way, getting out of bed daily, and more such obvious things.

Examples of habits to be urgently unlearnt are the use of idiomatic expressions and abbreviations, anonymity and pseudonymity, inappropriate communication while under the influence of substance abuse, and not responding punctually to *bona fide* work-related communication (as in regularly letting people wait for months). This paragraph may yield some angry "Do you mean me?" responses, but it has to be said.

There are also requirements that, unfortunately, can not be learnt, such as sincerity and sense of righteousness.

Jacobsen: Any thoughts on proposals for dynamic or adaptive tests rather than – let's call them – "static" tests consisting of a single item or set of items presented as a whole test, unchanging, instead of a collection of algorithmically variant or shifting items adapting to prior testee answers in a computer interface?

Cooijmans: Firstly it occurs to me that if one is going to use a computer interface and software to assess an individual's intelligence, analysis of observed behaviour (including communication) and of the candidate's responses to a computer-conducted interview should already provide a quite accurate estimation. Observation and interview are the primary means of gathering information in psychology. The interview could be made adaptive, with subsequent questions depending on prior answers, but a standard interview might work just as well. In the age of artificial intelligence, this is the way to go first.

Secondly, if one is going to use a computer interface and software anyway, the testing of elementary cognitive tasks like reaction time, decision time, perceptual threshold, and working memory capacity should probably be the next thing to do. After observation and interview, testing is the third method of collecting information. A practical problem is that one may need to use the same quality of hardware to get reliable results. When letting people use their own computer, the results may be affected by the quality and speed of one's graphical processing unit, and whether or not one has a dedicated one, for instance.

Finally, adaptive psychometric testing might be tried. But there are problems; it is not for nothing that static psychometric tests are so much more common in practice. Adaptive testing relies on item-response theory, wherein statistical properties like difficulty and discrimination are first determined for each item by letting a group of people try to solve it. These values are later used to compute the score of the candidate being tested adaptively, the set of items used being different for different candidates.

One problem is that statistical properties of single items are not constant in my experience, but change depending on the context in which the item is presented, and depending on the group of people attempting the item. For instance, if an item is presented among other items that are somewhat similar to it, it will likely behave as an easier item than when it is presented among items that are more different from it. And if an item is attempted by a group of conscientious people, it will have higher discriminating power than when it is attempted by unconscientious people. So the values of these item properties used in adaptive testing may be off, or as already said, single items do not have constant statistical properties, and that undermines the idea of adaptive psychometric testing.

Also, adaptive psychometric testing as it is normally thought of requires timing and supervision in my opinion. But the worldwide high-range testing population is used to unsupervised, untimed tests, and only a tiny fraction of them may be willing to travel to the hypothetical location where one has set up one's million-euro adaptive testing system.

Jacobsen: How do you remove or minimize test constructor bias from tests?

Cooijmans: It is best to prevent such bias by creating a wide variety of item types and subject matter, and by trying to think of new such types and matter with every new test. Studying comments from candidates may also help to avoid item types and subject matter that have become familiar among test-takers and that they appear to expect from you. Statistical item analysis may also indicate that there are problems with particular items, and by looking into that one may in some cases discover that the problem lies in the item's being too similar to other items one used before.

A few concrete methods to avoid bias are as follows: When creating knowledge-dependent items, consult a high-level thematic index of all the branches of human knowledge. One may find such in the Propaedia of the Encyclopaedia Britannica, or in old-school web directories from before search engines dominated web search. Strive to make each knowledge-dependent item come from a different branch of knowledge. This prevents the inclusion of only fields of knowledge that the test creator happens to be acquainted with.

Vocabulary-dependent items may be constructed with the aid of dictionaries and use of a random element when choosing words to include.

One may look over one's earlier tests when creating a new one to avoid repeating item types or patterns that were used before. Not that such repetition must be avoided totally, but it should remain limited, and a significant part of the new test should be novel.

Finally, to provide oneself with a broad pool of inspiration for possible test problems, one should expose oneself to a grand diversity of subject matter in the form of books and documentaries. This should also include materials that provide a basic understanding of fundamental sciences like mathematics, physics, chemistry, biology, astronomy, and so on. One should aim to

understand nature, reality, the universe, and awareness at the deepest level. The desire to understand existence is behind all great works of art and science.

Jacobsen: How do we know with confidence listed norms are, in fact, reasonably accurate on many of these tests? What is the range of sample sizes on the tests, even approximately, now? Practically speaking, for good statistics, what is your ideal number of test-takers? You can't say, "8,128,000,000."

Cooijmans: For the norms that I have made, the norming method is explained in the statistical report of the test in question, and some further explanation is referred to from the report. The reports contain about all the statistics that can be revealed without violating candidates' privacy and without damaging the security of the test. So if one understands the report, one knows how much confidence to have in the norms. In fact, I have devised a measure of quality of norms, based on the number of score pairs used and on their correlation with the object test.

Since the norms are anchored to other tests and not based directly on the general population (as opposed to the high-range population, for which I do have direct norms) it remains a question how close the high-range norms would be to the general population norms in that range, if tests existed that were normed directly on the general population and extended into the high range. The best indication thereto that I know of is the Mega Test by Ronald K. Hoeflin, which was normed mainly on the old Scholastic Aptitude Test and Graduate Record Examination, which did seem to give meaningful scores into the high range, and thus form an anchor point between the general United States population and the high-range population, albeit that the G.R.E. was administered to a clearly above-average sample of the population so that the S.A.T. is ultimately the true anchor point.

Hoeflin's Titan and Ultra Tests were normed to be consistent with the Mega Test norms, I think. The same goes for my early tests, and over the years I have tried to keep the norms in accordance with that anchor point over many generations of norms. To facilitate this, I have invented protonorms, which form an extra layer between raw scores and I.Q.'s, so that adjustments can be made in the relation of protonorms to I.Q. without having to change the norms of every single test. So, the question as to how we know that the norms are reasonably accurate, in one sense, goes back to Hoeflin's interpretation of reported Scholastic Aptitude Test and Graduate Record Examination scores, and scores on possible other tests used in norming the Mega Test, such as Cattell Verbal (also called Cattell B). Someone once sent me the data from the "Omni sample" of Mega Test scores, with known scores on other tests and correlations, which is how I know that the two mentioned educational tests provided the bulk of the norming data. I assume that Hoeflin had the population percentiles of the S.A.T. scores and used those as the main source of the Mega Test norms.

But there is more. Over time I have come to understand that the high-range score distribution itself contains information that is likely of an absolute nature and may help to anchor the norms or keep them consistent over time: The mode or modal range of high-range scores (when many scores are aggregated, for instance by combining the scores from many tests) occurs in the I.Q.

130s by current norms; below it, scores taper off steeply, above it, shallowly. This mode seems to be the point below which people feel less or not attracted to take high-range tests, and as such it should represent an absolute intelligence level; the level from where people are interested in intellectual endeavours, one might say.

Also, the level reached by the very highest scorers seems about constant over time, and falls between I.Q. 180 and 195 with the current norms. I am even carefully evolving to the viewpoint that this may be the highest intelligence level possible for any brain. So one could say that the norms in the high range are also defined by these two absolute (though coarse-grained) indicators (mode and maximum), not just by equation to scores from other tests. And, the number of scores that occur at these respective ranges are such that the current norms appear to be correct, that is, roughly in accordance with what one would expect given the predicted rarity in the general population of those I.Q. levels in a normal distribution. In fact one could theoretically norm the high range using these two indicators as anchor points, not needing scores from mainstream tests at all. And one could extend those norms linearly downward to include the normal range of intelligence, and the resulting scale might be better than that of actual mainstream tests normed directly on the general population. This is so because the general population and its average intelligence are changing, and therefore the norms of mainstream tests adapt to this change and are merely relative to the current population, not absolute. The high-range norms are the real, absolute indicator of intelligence.

The sample sizes of high-range tests vary from 0 to about 400, but for those with good norms mostly from 36 to 225 or so. The ideal number of test-takers to norm a test is about 64. More is not necessarily better, because as the submissions keep coming in and go into the triple-digit range, the later scores may not be fully comparable to the earlier scores anymore due to things like answer leakage and increased familiarity with item types, and the norms may be affected by that and become unfair to the earlier test-takers. This can be countered by replacing problems that have become too easy (have leaked answers) but that changes the test, which also makes later scores less comparable to earlier ones, and if you change more than a little bit, you have to call it a revision and start over at zero collecting statistics for that new version.

High-range tests that appear to have very large samples, like around 300 or more, have generally achieved this through undesirable manipulations like retesting under false names, or combining retests with first attempts in the same sample, and so on.

Jacobsen: What are the most appropriate means by which to norm and re-norm a test when, in the high-range environment so far, the sample sizes tend to be low and self-selected, so attracting a limited supply and, potentially, a tendency in a restricted set of personality types? Dr. Ronald Hoeflin was claimed to have the largest sample size of the high-range test constructors. Do you have the largest legitimate sample size of any high-range test constructor at this time, now, based on over a quarter century conscientiously gathering data? You were the most recommended person to interview for this series.

Cooijmans: In my experience, the best way to norm a high-range test is to rank-equate its raw scores to normed scores of the candidates on other tests. The other tests to be used should be selected based on their correlations with the object test; one sets the correlation threshold such that one obtains enough pairs. I have recently begun to set the threshold so that it maximizes the quality of norms, as given by a mathematical expression that uses the number of pairs and the weighted mean correlation. Thus it is objective, avoiding human decision. The expression that represents quality of norms is operational and may be improved as insights advance; I mention this because I know some are inclined to take these statistics as final and absolute, but they are parameters or controls that one sets to tune the system.

I deny that high-range sample sizes are low. They are in the dozens to hundreds as I said above, and that is well into the range of mainstream test samples and more than enough for good statistics. Considering that the high range consists of only a fraction of the population, it is to be expected that the samples are smaller, and in fact they are not much smaller at all. The notion that mainstream I.Q. tests have enormous samples is mistaken. Typically they have several hundred per norm group. Norm groups exist for age ranges, but sometimes also for educational levels. In the Netherlands there are different levels of secondary schools, and mainstream I.Q. tests may have separate norms per level, sometimes even based on only a few dozen per level (like in a Netherlandic version of the WAIS some years ago). A test often used by Mensa was normed on 3,000 people, but divided over five age groups from 13 to 16 years, so the actual norms were based on 600 per age group. In other words, they used high school students. And such norms have often been used for decades, ignoring the inflation of scores called “Flynn Effect”. But in the minds of some people, the illusion is persistent that these “standard tests” are normed on hundreds of thousands or even millions, and form a kind of gold standard of I.Q. testing.

The largest samples are found in educational tests, but not as large as some think. In the Netherlands, a test called Cito-toets has long been used in the last year of primary school, yearly taken by about 100,000 children. But not normed on that number! The norms were established by administering an anchor test to a sample of about 4,500 shortly before the actual test, and then equating the anchor test scores to the actual test scores. This helped to keep the standard scores stable throughout the years (the contents of the anchor test would remain the same for a number of years, while that of the actual test changed per year).

My own Cito report from 1977 shows a percentile of 100, which is uncommon but probably means the actual value was above 99.95, as a later statistical report by Cito I got to study contained a table where percentiles were rounded to 100 if the actual value was above 99.95. I have asked Cito in the mid-1990s what the precise value was, but they could not tell me, they only had kept percentiles as whole numbers. Similarly, I inquired about my scores on a comprehensive test given to us in secondary school around 1980, something like the Differential Aptitude Test, but was told those scores had not been saved. We never got a score report for that test at the time, but I understood from teachers I had done extremely well, and on a parent’s evening (which my parents never attended) a teacher told the public that I was a one-off (“unicum” was the Netherlandic word used). This teacher died in 2013, incidentally.

On the whole, I believe that high-range psychometrics is much more careful than mainstream psychometrics when it comes to the quality of norms and handling of score inflation by causes like answer leakage or people becoming more familiar with particular item types.

I might have the largest sample size of current high-range test constructors. It includes over 3,000 individuals, over 6,500 scores on I.Q. tests scored by me, over 2,900 reported scores on other tests, and over 22,000 data points on personal details, including personality tests. But more importantly, I have organized that data in an accessible way and automated the processing of it. I did all the programming myself, including the statistical functions.

Regarding a potentially restricted set of personality types and self-selection, it is inevitable that persons in the high range of intelligence differ in personality from those in the normal range and from those in the low range. This does not invalidate the norms in the high range. In fact, intelligence itself is a major aspect of personality. Self-selection is less of a problem than it seems because in general, people like doing what they are good at, so those attracted to taking high-range tests will mostly be intelligent. This is also illustrated by the rareness of low scores; only 3.5 % of scores fall under I.Q. 120 and 15 % under 130 (and no, this is not because the norms are too high, as self-doubting candidates sometimes suggest). Precisely what is going on with intelligence, non-cognitive personality traits, and brain-related disorders in the high range, and how this leads to creativity and genius in some, is an interesting question and I hope to understand more of it later on.

Jacobsen: What is the structure of the data in high-range test results? Do homogeneous and heterogeneous tests change this?

Cooijmans: Data structure is so important that someone who starts out collecting data for some purpose should ideally think out the database design beforehand. Once you have collected a lot of data, it becomes hard to make big changes to the design. The data structure of high-range tests looks as follows:

At the top level there are five sections:

(1) Descriptive information records for each test or type of personal datum. Each test or datum has a record here, and each record contains fields that hold information such as the test name, its maximum score, its contents types, and whatever further descriptive information there is. Conceptually, one may even imagine the tests themselves residing here in their respective records, but in practice one will probably not store actual tests in a database but think of the database as referring to tests that exist in a reality outside the database.

(2) Candidate records. Each candidate has a record here, and each record has fields that hold the personal data of the candidate, and the candidate's scores on the respective tests. Notice that a record here has hundreds of fields, but most or all candidates have only part of those hundreds of fields filled, depending on how many tests they have taken (each test has a field).

Conceptually, one may even imagine the candidates themselves residing here in their respective records, but in practice one will probably not store actual candidates in a database but think of the database as referring to candidates that exist in a reality outside the database.

Technically speaking, the test scores stored here are redundant insofar as they are also available from section (3), but for reasons like faster processing and reducing load on the processor, redundant fields are sometimes included in databases.

(3) Test submission records. In this complex section, each test has a table, and each table has one record for each submission to that test, and each record has fields that hold information like some personal details of the candidate (corresponding to a record in (2)), score and possibly subscores, and the item scores, so for each item typically 0 or 1 for wrong or right, but any range of item scores is possible. Conceptually, one may even imagine the submitted answers themselves residing here in their respective records, but in practice one will probably not store actual submitted answers in a database but think of the database as referring to submitted answers that exist in a reality outside the database.

Do make certain to understand the difference between “test” and “test submission”. Some say the first when they mean the latter, but the above paragraph illuminates the necessity to distinguish the one from the other.

In this section in particular there is some appropriate redundancy in the form of for instance sex and age of the candidate (also available from section (2)) and scores and subscores (can also be computed dynamically from the item scores). But for reasons like faster processing and reducing load on the processor, redundant fields are sometimes included in databases.

(4) Test norm records. This complex section has a table for each test, and each table has one record for each possible score on that test, and each record has fields that contain the raw score and the corresponding norm (in my case this is a protonorm).

(5) Norming scale records. This section has one record for each norm as may be contained in (4), and each record has fields that hold the norm and corresponding values on other scales for that norm, for instance percentiles, proportions outscored per sex, and I.Q. if the actual norm is not an I.Q. (such as in my case, where protonorms are the norms contained in (4)).

This structure has emerged over time as a natural reflection of the data itself. Someone who starts from scratch may well find that a completely different approach works too. Perhaps one would rather avoid any redundancy? As long as one has thought it over carefully.

Jacobsen: What should be done with homogeneous and heterogeneous tests?

Cooijmans: I consider only heterogeneous tests able to give a good enough indication of general intelligence, and use the term I.Q. only for heterogeneous tests, not for homogeneous tests. Also I refuse to administer homogeneous tests because I do not want to confront people

with a score that is a less good indicator of their intelligence, and do not want to facilitate people who want to show such a less good indicator to others and thus give a misleading impression of themselves.

Heterogeneous tests are tests that contain at least two different items types out of verbal, numerical, and spatial (sometimes I use “logical” as a type too). If one wants to study the intercorrelations of different homogeneous tests, the best way to do so is to use a heterogeneous test that has different homogeneous sections or subtests. One can then do correlation analysis or even factor analysis within such a sectioned heterogeneous test. That is also how factor analysis is traditionally done. A great advantage of this approach is that the sections or subtests will always have been taken by exactly the same group of candidates, and that is required for proper factor analysis.

Some of my heterogeneous tests have homogeneous subtests that are normed in their own right to “standard scores” (on the same scale as I.Q.), and in that case one can also compute the correlations of such a subtest with homogeneous subtests that reside in other such compound heterogeneous tests. But I dislike this complication and am striving to move to having only non-compound heterogeneous tests; that is, with sections not normed in their own right, or without sections, just with different item types mingled throughout the test. Another disadvantage of correlations between the subtests from different heterogeneous tests is that those subtests have been taken by different groups of candidates, so that proper factor analysis will not be possible, if one was thinking of that.

Jacobsen: People take multiple tests. They crunch those numbers. An implied claim of a real I.Q. from this crunching of numbers between multiple tests. Is there such a “real I.Q.” computable from multiple tests?

Cooijmans: In theory there is, but in practice there are problems that hinder the computation of a real I.Q. across tests. In the high-range community of candidates, many have taken enormous numbers of tests, dozens at least, and sometimes more than a hundred. It is problematic to compute a real I.Q. in the usual way from all taken tests for reasons like the following: The intercorrelations of the tests are mostly unknown, and there are too many intercorrelations for them to ever be known in the first place. Some tests may have bad norms. Some scores may be fraudulent. If a selection is made from the taken tests to narrow it down, this may be a non-representative selection. For example, a candidate having taken thirty tests may like to have a real I.Q. computed from one’s top several scores, which are already way above the real level of the candidate, and then the computed I.Q. will be even higher than the average of those top several scores due to the formula used.

The formulas for computing a real I.Q., such as “Ferguson’s formula”, take the average of the input scores and add something based on the correlations between the tests. With a perfect correlation, the outcome is simply the average. The lower the correlation(s), the higher the outcome. With zero correlation, you get something like a full unit of spread on top of the average. This may be correct in theory, but in practice leads to inflated outcomes. Apart from

using a non-representative selection from one's scores, another cause of inflation with these formulas is the fact that the known correlations between the tests are often underestimations of the true correlations due to incompleteness of the data and restriction of range. The groups who have taken the respective tests have only limited overlap for any pair of tests, and this overlap may suffer from selective reporting, and all in all this depresses the correlations. And lower correlations mean that the formula will yield a higher outcome. Underestimated correlations inflate computed "real I.Q.'s".

Also, when a person takes multiple tests, a learning effect may take place as a result of which the scores become somewhat higher. This increase then comes in addition to the compensation for imperfect correlation that is built into these formulas for "real I.Q."

For tests scored by me, I have devised a "qualified average I.Q.", which tries to avoid the problems with these "real I.Q.'s". Since I always have the complete data, no selection bias can inflate the average. The problem of underestimated correlations inflating the outcome is avoided by not using the computed correlations but assuming perfect correlations. If it seems unfair not to compensate for imperfect correlations, one may imagine that the learning effect from taking multiple tests replaces this compensation, so to speak. Finally, the computation is resistant to outliers. This is not claimed to be someone's real I.Q., but I believe it is better than something like "Ferguson's formula". The exact formula of the qualified average I.Q. is operational and may be perfected over time as needed.

Jacobsen: Is English-based bias a prominent problem throughout tests? Could this be limiting the global spread of possible test-takers of these tests rather than limiting them to particular language spheres? Although, these tests are taken, to a limited degree, in many countries of the world in all/most regions of the world.

Cooijmans: Such bias is a problem, but how prominent it is depends on what one's native language is and on whether one knows English. For other Germanic languages it is a smaller problem than for non-Germanic languages, and it is worst for East Asian languages. The fact that reference aids are allowed solves a big part of it, but for a nonnative English speaker there remains a disadvantage, which I have estimated at up to 5 I.Q. points. Without reference aids (on a verbal test that disallows reference aids) this would be more like 30 I.Q. points for this non-native English speaker, and for someone who does not know English altogether it is in my opinion better not to attempt the tests at all.

It certainly limits the global spread of test-takers, especially in the areas where few people know English and the local language is very different from Germanic languages. I have always thought that the best solution for this is that people in such areas create their own tests in their own languages.

In recent years it has become somewhat common for people to try tests in a language they do not know. Of course one has an unpredictable disadvantage then.

Jacobsen: When trying to develop questions capable of tapping a deeper reservoir of general cognitive ability, what is important for verbal, numeral, spatial, logical (and other) types of questions?

Cooijmans: That reservoir will likely be tapped almost regardless of the questions, as general intelligence expresses itself through virtually everything a person does or says. Important are things like having a wide diversity of questions and types of questions, and avoiding localized transient subject matter like idioms, abbreviations, pronunciation matters, and local or fashionable knowledge, as such does not transcend barriers of language, culture, and age. Fundamental knowledge that is the same for everyone in the world is good; knowledge that is bound to a geographic area, in-group, or period is bad. For these reasons, and contrary to what some think, high-brow vocabulary and subject matter are more culture-fair than low-brow vocabulary and subject matter.

One should also be aware that learnt skills have no *g* loading; it is novel tasks that have *g* loading. Candidates sometimes complain that they have no idea what is expected from them when taking a test, or how to tackle it; but that is exactly the intention, that is how intelligence testing works! And candidates may be happy when they see a type of problem they have solved before because they know what to do then; but that is where their intelligence is NOT being used. Those problems have lost their *g* loading for them. So one should try to create problems that are different from what has been seen before, to enforce the use of intelligence.

To illustrate that even esteemed test constructors not always understand the loss of *g* loading of learnt skills, here is an anecdote: Some years ago on a social medium, I saw a test author proudly mention that his young child had scored over I.Q. 160 or so on one of his father's tests; after extensive coaching by the father/test creator, of course!

Another observation regarding tapping into general cognitive ability: Good test problems are such that solving them is similar to making discoveries in the real world, unravelling the laws of nature and the universe.

Jacobsen: What are roadblocks test-takers tend to make in terms of thought processes and assumptions around time commitments on these tests? So, they get artificially low scores on high-range tests. Also, what is the confusion made by smart (and, potentially, not-smart) people about time taken for a test to get a score and the intrinsic intelligence to get said score? You noted the latter point in one of the recent videos answering questions on your YouTube channel.

Cooijmans: The idiomatic use of "roadblocks" is an example of what should not be in an intelligence test. Such an idiom is only understood within a narrow linguistic region and a restricted time period. It can not be understood without already knowing what it means. It can not be understood from the word itself or its context. The avoidance of idioms requires high intelligence and an abstract-literal mind.

The test instructions state that there is no time limit. Yet some think that their score will be unrealistically high and invalid if they spend “too much” time. It has happened that someone said, “I have now been looking at this test for so long that I can not submit it any more, I found all the answers, it would not be fair”. But that is exactly the intention with untimed tests; that one continues until one finds no further solutions.

The confusion meant in the question is probably the notion that someone who uses less time is smarter than someone who uses more time to arrive at the same score. But the principle of untimed testing is that this is not so, and that “until one finds no further solutions” is the right amount of time, irrespective of how long that is. This principle is based on the finding that when the allowed time is increased on a timed test, the test’s *g* loading rises. With supervised tests one needs to have a time limit for practical reasons, but with unsupervised tests one can leave out the limit entirely.

I must add that I have nothing against supervised tests, provided they have a very broad time limit, something like three hours for a comprehensive test. But this is not feasible in the high-range testing practice. I can not get people from all over the world to travel to a place here where I can test them, and I can not set up testing locations worldwide in all countries. I tried, but the number of candidates willing to make use of such was negligible compared to regular unsupervised tests. So I stopped. And then there is always someone who says, “I would be willing to travel to you if you started with that again”. But one or two people is not enough to justify the significant effort and time put into such a project. If others wish to try it, go ahead.

Jacobsen: What is the intended age-range for high-range tests? How do these account for individuals younger and older than this range?

Cooijmans: From about 16 upward with no upper limit I would say. Older people do decline, but it is important that they participate in order to enable the study of this decline. Younger people are allowed to take the tests, and in practice, 12 years is about the lower limit. But they should be aware that they have not reached their adult level and will score lower than they will later be capable of. The steep increase of intelligence in childhood tapers off at about 16 and becomes shallow then, hence the idea that one enters one’s adult intelligence range at 16.

Another way to answer this would be “after puberty”. Individuals, sexes, and ethnic groups differ in their childhood development, then puberty messes everything up, and after puberty things have mostly settled. That is why childhood studies of mental ability are so misleading; they misrepresent possible sex and ethnic differences. Puberty has normally completed by or before age 16-17. Age of onset of puberty varies greatly per individual, sex, and population, and tends to be one to two years earlier for girls than for boys.

There are no separate norms per age group as that would hide the development of intelligence with age. And one wants to reveal that development, not hide it. Also, all candidates are treated and addressed as mentally mature adults, regardless of age. The development of intelligence

with age plausibly differs per sex, which is why it should be studied within-sex; the most recent tabulation I made is at http://iq-tests-for-thehigh-range.com/statistics/age.html

Jacobsen: A modestly common/uncommon knowledge of sex differences in the measurement of intelligence: Men do better at visuo-spatial subcomponents and women do better at verbal-emotional processing. What is important in constructing and norming a test if these and other differences exist? What similarities exist to not change this process?

Cooijmans: There are indeed sex differences in aspects of mental ability. In constructing unsupervised high-range tests, it is not possible or meaningful to take these into account. One should just include the widest variety of item types usable in unsupervised tests and focus on high mental ability regardless of sex.

Women have the bad fortune that the aspects on which they are known to outscore men mostly require supervision and timing, and can therefore mostly not be included in unsupervised tests. According to Arthur Jensen in Chapter 13 of his book *The g Factor*, these aspects are simple arithmetic, short-term memory, fluency (for instance, naming as many as possible words starting with a given letter within a limited time), reading, writing, grammar, spelling, perceptual speed (for instance, matching figures), clerical checking (both speed and accuracy, things like underlining certain letters in a text, or digit/symbol coding), motor coordination, and finger and manual dexterity. This problem is less serious than it seems because these are mostly lowly g-loaded tasks (not by anyone's decision but because it happens to be so) so that the overall score will not be affected much by their absence, but it may be affected somewhat. This is related to what was observed in my answer to "What are the core abilities…"

In norming, the proportion of high-range candidates outscored should be provided within-sex for reasons of transparency. I.Q. norms should be sex-combined as is usual.

Jacobsen: Cheaters exist. Frauds exist. How do you a) deal with frauds and cheaters on tests and b) prevent fraud and cheating on those tests? Have reference texts been a problem in this? Does artificial intelligence complicate matters more? (If so, how?)

Cooijmans: When I discover that someone committed fraud I will discard the fraudulent score in the database and make a note so that I can exclude that person from further testing and from society admission. This is sometimes complicated by the use of multiple false identities by such a person. If the person is a member of a society I am an administrator of, I will expel the person. In communication with other test creators or societies I may reveal what I know about the person if that seems appropriate. I do not believe there exists an organized system for sharing information about frauds between test creators, perhaps there should.

Attempting to prevent fraud is done, for instance, by not publishing the test itself, letting people prepay, not sending tests to known frauds and so on. And if I find out that answers to particular test items are published or spread somehow, I will do something about it; mostly it comes down

to replacing the items, sometimes leading to a revised version of the test. Sometimes a test is withdrawn entirely.

I am not aware of reference texts that were involved in fraud. Artificial intelligence complicates things because frauds might consult it to solve test problems, which is not allowed as the test instructions state not to obtain answers from external sources but only use answers that one thought of by oneself. To reduce this complication I try to create problems for new tests so that current artificial intelligence, insofar as I apprehend it, can not solve them. I try to make the problems so that, once artificial intelligence becomes able to solve them, it will also be able to take tests and join societies on its own accord. I believe that will happen one day, but fear this day lies quite far into the future. If I had to guess I would say half a century.

Jacobsen: It helps to have other data from other tests and personal data for identity verification. What information from other tests is helpful/necessary for research purposes of high-range tests? What is an efficient and appropriate format to provide this score information? What personal data is necessary from candidates, if any? What information would be helpful for research purposes from candidates?

Cooijmans: Scores on other tests should best be reported in a format as follows, insofar as known:

[Test author or issuing organization] [Test name] [Raw score] [I.Q.] [Standard deviation of I.Q. scale used] [Percentile]

Scores should best be grouped by the first field (Test author), of course starting a new line for each score. Nowadays there exist hundreds of tests, and I can not know from the top of my head which test is from which author or organization, so if that first field is left out when reporting scores, which is common, this causes many minutes of extra work in processing that information.

Concerning personal information, at least name, sex, year of birth, country of origin, and highest achieved educational level. Some further information I collect is the educational levels of the biological parents, the presence of a psychiatric disorder, and the presence of such disorders among parents or siblings.

Notice that I find the exact date of birth not strictly needed. It is about studying the development of raw intelligence with age, and with adults, year of birth suffices. In childhood testing, one would want it to the month.

Regarding psychiatric disorders, I do not ask for the particular disorder as that would require too much detail, too many options, too much complication in the statistical processing of it.

And country of origin is a pragmatic imperfect proxy of origin. One might consider asking for race or ethnicity, but such categorization is logistically problematic when one looks into it, has many complications, may be considered unethical by some, and some may refuse to reveal their status.

Other data that might be useful to collect include religiousness and femininity/masculinity (independent of sex). The possible correlation between religiousness and high-range I.Q. could then be established, which many are wondering about. And one could verify the anecdotal observation that intelligent men are more feminine than average men, and intelligent women more masculine than average women. In other words, there is more gender diffusion in the high range, which would point to an optimum for intelligence somewhere between the average male and female positions on the femininity/masculinity dimension. Notice that the term “gender” is for once used correctly here. I am uncertain whether people would be able to simply report their own position on femininity/masculinity, or whether this would require a test or questionnaire.

Jacobsen: What is the level of the least intelligent high-range test-taker now? What is the level of the most intelligent candidate now? What is the mean, median, and mode, of the scores of test-takers’ data gathered so far? Within a range of I.Q. 10 to 190 on an S.D. of 15, when should a candidate consider taking, or in fact take, high-range tests?

Cooijmans: The least intelligent seems to be in the I.Q. 80s. The frequency of such is one in thousands of high-range candidates. The most intelligent is plausibly between 185 and 195. One can not be certain yet about the accuracy of the norms there. And with candidates apparently far below the general population average, a problem is that they tend not to report usable information, so one has to resort to observation, life history facts they happen to mention, and aids like an online writing-to-I.Q. estimator.

The median is protonorm 401 (I.Q. 139) according to the latest computation I did of highrange quantiles. I never compute the mean, but that should be several I.Q. points higher because the distribution is skewed to the high side. The mode is protonorm 387 (I.Q. 137), but one could also say there is a modal range in the 130s. A mode always depends on how wide or narrow one chooses the classes of the frequency table.

People should consider taking high-range tests if they score above the 98th centile on some mainstream test, which is I.Q. 131 (or 130 on some tests that round differently). Below I.Q. 120 there is no reason to try high-range tests, but there is no objection to doing so anyway. There is a grey area from 120 to 130 because one does not score the same on different tests.

Jacobsen: What is efficient means by which to ballpark the general factor loading of a high-range test?

Cooijmans: I have always used the square root of the weighted mean correlation of the test with other I.Q. tests as an estimation of the *g* loading. This works well for comparing different high-range tests. It is not a true *g* loading because the tests have not all been taken by the

same group of candidates, but by different groups with limited overlaps, correlations obviously being computed for those overlaps. Also, when reported scores from other tests are involved, those may suffer from selective reporting, which depresses the correlations. If all the involved tests had been taken by exactly the same group of candidates, one would be close to a true *g* loading.

Another thing to consider is that high-range tests as I use them are almost all heterogeneous tests, so combining different item types within the test. But, in classical factor analysis, one uses a set of different homogeneous tests that have each been taken by each individual from a group. Typically, these are school exams for the various subjects administered to a school class, or subtests of a comprehensive psychological test like Wechsler Adult Intelligence Scales. Via factor analysis, one then computes *g* and other factor loadings per subtest or exam, and these *g* loadings vary greatly and may be very low for some subtests. So this kind of analysis is not so much done with a set of heterogeneous tests; computing correlations between heterogeneous tests is more something seen in high-range psychometrics. In classical factor analysis, wide-range heterogeneous tests are considered pure indicators of *g*, and it is the loadings of the various subtests or exams that one is interested in.

Jacobsen: What is the most precise or comprehensive method to measure the general factor loading of a high-range test, a superset of tests, or a subset of such a superset?

Cooijmans: The first is answered in the previous question. For a superset of tests it is not needed to compute such a loading, the superset can be safely considered a near-perfect indicator of *g*.

Jacobsen: What seem like the most appropriate places for people to start when taking your tests–taking into account their own skill sets, or others' tests for that matter?

Cooijmans: I would recommend the privacy of one's home. If "places" is meant nonliterally – as one sees, I am not one of those pedants who take everything literally – then a real computer to view the test is best, or at the very least a decent laptop (although I am really against the unneeded use of battery-powered devices). A smart telephone is no place to start.

In case the non-literalness is even more remote, always start with the easiest tests. It is bizarre how it can occur to people to start out with the hardest tests, and how they subsequently can not understand what a score of zero means and keep asking for years thereafter what their I.Q. was on that test and if it means they are "gifted". By looking at the test norms one can know how hard a test is. Nevertheless, I have recently ordered the list of available tests by difficulty to accommodate this.

Jacobsen: What tests and test constructors have you considered good?

Cooijmans: Constructors: Kevin Langdon, Ronald K. Hoeflin, a Netherlandic person who withdrew from the I.Q. societies so I can better not name him, Edward Vanhove, Hans Eysenck, Bill Bultas, Laurent Dubois.

Tests: Mega Test, Magma Test, tests from the self-test books by Eysenck, Chimera Test, 916 Test.

Regarding Langdon, studying his tests and statistical reports was instructive, if only because it told me which approaches were not so successful in measuring high-range intelligence. This includes attempting to make it more or less culture-fair, using multiple choice with a small number of options, item weighting based on item analysis (gives too much weight to a small number of items, and also the fact that single items do not have constant statistical properties undermines the idea of item weighting based on those properties), selecting items for a shorter test based on their statistical behaviour in an earlier longer test (the items' behaviour is different in different tests), and norming that shorter test based on statistics from the earlier longer test (norms become mostly too high then). Also, that statistics from classical psychometrics, such as the reliability coefficient, are woefully inadequate to assess the quality of a high-range test.

Jacobsen: What have you learned from making these tests and their variants?

Cooijmans: I assume this is about my tests, not the tests by others from the previous question. The main points have already been mentioned in the questions about counterintuitive findings and about *g* not diminishing much in the high range. I could add the observation that intelligence expresses itself in almost everything a person does or says. I did not know that when I started.

In case the question is about the tests from the previous question, I already answered it there for Langdon's tests. With Hoeflin's tests, I learnt the norming method of rank equation, and the destructive effects of fraud through retesting, false names, and cooperation. In correspondence with others in the 1990s, I was appalled when people proudly told me they were collaborating in a group to "crack" the Mega Test. When they told me they had retested under their own name or another name (the instructions said the test could be taken only once but Hoeflin allowed retests in practice). When someone told me he had first taken all of Hoeflin's tests under his own name, then his sister's name, then the son of his sister's name, with ever-increasing scores. When someone told me he had first taken all of Hoeflin's tests with rather low scores, then had some friendly correspondence with the person from the previous sentence, then took the tests again with the same scores as the highest scores of that person. When someone told me he had missed the Mega Society pass level by half a standard deviation, then retested and qualified. With "retest" I always mean "to take the same test again".

Several of those meant in the previous paragraph showed me their answers (unasked) and suggested I use them to get into the Mega Society. I had rarely been so shocked and insulted as by the suggestion that I would be capable of such fraud. I do not understand how those types can live with themselves. Having said that, two of them committed suicide in that period.

And of course, such people publicly display or mention their highest (fraudulent) scores, not their honest scores. I remember a phone call with a Netherlandic Mensa member as if it was yesterday; “Yeah, the ‘Mega Test’, I am working on it with some people in Spain and east Asia. Yeah, we have it mostly figured out now, that ‘Mega Test’, ha ha ha…” This person killed himself not long thereafter. Not the “Beheaded Man”, incidentally; the history of I.Q. societies is riddled with suicides, and some of them appear to have made the right decision for once in doing so. That is one thing that gives hope then; that some can indeed not live with themselves in the end.

Jacobsen: I received some decent points about high-range tests from Mahir Wu. Credit to him for the raw materials and permission to reframe those points as questions here. He raises foundational points. First point: item answers should be rigorously unique. Why?

Cooijmans: If multiple answers to a problem are correct, this has disadvantages: The one answer may be easier to find than the other, so that candidates with the same credit may not really be of the same level because they found different answers. And candidates who see more than one possible answer may be confused and not know which is the “right” one. Also, there may be subjectivity in scoring those answers. With only one correct answer, these problems are mostly avoided.

Of course, no matter how hard the test creator tries to make items with unique answers, once people start taking the test, it may sometimes occur that alternative valid answers are still found, and then one has to solve this, for instance by revising, replacing, or removing the item. Sometimes this can be done “in place”, especially early on when there have not been many submissions yet, and otherwise this may be done later in a revised version of the test.

And, no matter how hard the test creator tries to make items with unique answers, there will always be people who “see” alternative answers through apophenia when they can not find the real answer. The apophenic delusions stick rigidly in their minds and they become convinced they have solved the problem, although the logical flaws are obvious to an objective observer. In popular artificial-intelligence speak, people “hallucinate” when unable to find the real solution. But that is inherent to intelligence testing; escaping this delusional rigidity is part of high intelligence, or rather, is an aspect of having a wide associative horizon. You will need that mental flexibility too when solving real-world problems. Sometimes, you have to take a step back and make a fresh start to eventually find the real solution.

To illustrate that apophenic delusions are very real and persistent, I want to give some examples: I have a Test for Extrasensory Perception, which is exactly what it says. It is not an intelligence test, I did not hide any clues or patterns in it. Still, some years ago an otherwise normal person sent me a document of many pages, describing his decoding and solutions for it in long association chains. He was convinced he had found patterns that I had deliberately put there. Since this was explicitly not the case, this example proves that candidates may suffer from apophenic delusions all by themselves, and that this is not caused by ambiguity of the test items.

And long ago someone published articles in I.Q. society journals, explaining how he had found references to the appearances of particular comets hidden in poems of certain literary authors.

And another one produced a long series of essays, analysing the dates of events related to the Roman Catholic church by counting the days separating the events, finding numerical patterns therein, and concluding that the Vatican was conducting a dirty scheme that would culminate in some horrific project (I am not allowed to disclose it I think) of which he predicted the exact date in the near future.

I am not naming these examples to ridicule people, but to show that such delusions can be extremely strong in apparently sane people. When taking high-range tests, it occurs often. If it happens to you, rest assured, for only in a small minority of cases does it lead to full-blown psychosis.

Jacobsen: Following from the previous question, the test item answers with ambiguity should be disallowed. Why should these not be allowed, if agreeing with Wu? If disagreeing with Wu, why?

Cooijmans: I agree for the reasons given in my previous answer. But as said, sometimes you only discover ambiguity as test submissions are coming in, when studying comments by candidates.

Jacobsen: Why should test items give sufficient clues for discovery and solution by a test-taker?

Cooijmans: Because otherwise it is impossible to solve the items, obviously.

Jacobsen: Following the last question, why would permission of a mere guessing logic spoil a test?

Cooijmans: Because correct answers that result from guessing do not stem from the candidate's mental ability being used. Such answers are random variance and thus reduce the test's reliability, and therefore also its validity. Test items should be made so that the probability of getting them right by accident is so small that, on average, candidates will gain less than one raw score point in the total score by guessing. This does permit multiple-choice items, but they should be cleverly constructed so that the likelihood of a correct guess is very small. For instance, by letting the candidate choose several options from a list instead of just one.

Incidentally, I have heard people suggest that multiple-choice items that can be answered correctly by guessing reveal intuition and/or psychic ability, but even if that is true, I believe that intelligence tests should not measure intuition or psychic ability. I am also not a fan of penalties for wrong answers with multiple-choice; supposedly, this corrects for guessing, but of course, a candidate who chooses a wrong answer, thinking it is right and not guessing, is then penalized

for the wrong reason. The penalty does not distinguish between guessing and being simply wrong, and in the latter case, no penalty should apply. For clarity, a penalty constitutes a negative item score, typically a subtraction of a fraction of a point, depending on the number of answering options for that item.

An anecdotal experience regarding multiple-choice tests: Once, the instructions to one of my multiple-choice tests said, “There are no penalties for wrong answers”. After a while I removed that instruction because some candidates demanded a perfect score based on it. They took it as, “You will always get a perfect score no matter what you answer”. Of course they were wrong, because one starts out with zero points at the beginning of the test, not with the maximum, so “no penalties for wrong answers” in no way implies a perfect score. But if people are so willingly and stubbornly taking it the wrong way, I am not going to pain my brain trying to formulate it even better than it already was.

Jacobsen: How can the sufficiency of each test item’s uniqueness become integrated into the overall test (even test schema) to prevent the identical pattern from emerging too much in a single test (or test schema)?

Cooijmans: If I understand the question correctly, I would say that a test should consist of a broad diversity of items with mostly different patterns. They need not be all completely different; maybe two or three of a similar looking pattern are acceptable, provided the implementation of the pattern is different every time, so that the candidate is forced to recognize what is going on in each case.

In some of my tests, like “Problems In Gentle Slopes of the first degree”, I had a series of about ten problems of the same kind in ascending difficulty, and that did not work well. Many candidates were able to solve all the problems in such a series. The items work as examples for each other and become too easy. So I concluded that it is better not to have more than 1 to 3 items of a similar kind in a test, and even those should differ sufficiently in implementation.

Jacobsen: How can the inspiration from, even addition of, other authors’ test items degrade a test’s quality by giving more clues to test-takers to test items otherwise unsolved without them?

Cooijmans: If a test contains an item that is similar to an item in another test by another author, the one item may function as an example to the other and thus make it easier. I have experienced a few times that a difficult problem in one of my tests appeared to have become much easier. Eventually, a candidate told me that a test by another test creator had a very similar but easier problem, and that made my difficult problem suddenly solvable. I then replaced that item.

And, if a candidate is familiar with a particular item variant from other tests and is thus better able to solve such items, those items also lose their *g* loading for that candidate. It becomes a learnt skill, and learnt skills have no *g* loading.

Jacobsen: Following from the previous question, what about the reduction of the references to specific test items used by other test authors?

Cooijmans: If the question is about references inside a test to specific test items by other test authors, I am not aware of such references, possibly because I never look at tests by others. If such references exist, they probably help the candidate, which one may not want. But it is better not to have test items that resemble items by other authors altogether.

Jacobsen: In some sense, is it truly difficult to avoid this issue of test logic and design schema close-but-imperfect replication from one author by another inspired–by the former–author, especially as more high-range tests are constructed? Wu references his latest test, "[Mystery]," as an example of an adherence to the close application of this principle, where the evidentiary effects of others' tests become hard to apply to it. Consequently, results for "[Mystery]" are submitted much less.

Cooijmans: With so many high-range tests in existence, it will be getting harder to avoid similarities between tests by different authors indeed. I myself never look at tests by others and create problems independently. In an earlier question about avoiding or minimizing test constructor bias I name some independent sources of inspiration. These do not include tests by others. One should never look at tests by others for inspiration for new test items!

Jacobsen: Why should scale and norm not be overly subjective? Wu references T. Prousalis–link– and you–link, link. Also, why does a median score for many tests with a corresponding IQ of 145 (SD15), or higher, make little sense?

Cooijmans: Norms should be objective and correct, otherwise they are not comparable between tests. A few possible causes of incorrect norms are the following: When a beginning test scorer starts out administering tests, initially one will only have reported scores by candidates to base the norms on. Unfortunately, many candidates are dishonest in reporting scores, leaving out lower scores and reporting the higher ones, or even reporting retests or fraudulent scores. This gives an upward bias, and the norms based thereon may be ridiculously too high, even 10 to 20 I.Q. points too high on average. In the longer term, this may sort itself out as one acquires more, and more true, data about the candidates' scores. Theoretically, this could also be solved by different test constructors sharing their candidate data to thus make the candidates' true scores on other tests available, but I believe this might be unethical and a violation of privacy. I know some test designers are currently publishing candidate scores online, but that too seems unethical, and also I do not know if that published data is trustworthy and am hesitant to use it.

For information, a few test creators have sent me their complete data for a particular test of theirs, including candidate names, and I have scored a test by another author (Bill Bultas) myself in the past, so for those tests I have unbiased data.

Another cause of incorrect norms is megalomania by the test creator. There exist authors who delusionally reckon themselves to be profoundly intelligent, but really have much lower I.Q.'s, typically in the 130s to 140s at most. So when they receive test submissions by people whom they perceive as being at roughly their level of understanding, they feel compelled to give out much too high I.Q. scores, otherwise they would have to admit to themselves they are not really as intelligent as they believe.

A median of I.Q. 145 or higher is unrealistic. The high-range population is roughly the upper segment of the general population, cut off at about I.Q. 130. This is not a perfectly clean cut, but if it were, and for the sake of illustration, the following would be necessarily true: With a clean cut at 131 (98th centile) the median would be 135 (99th centile, so halfway the cut and the top). With a clean cut at 135 (99th centile) the median would be 139 (99.5th centile). A median of 145 (99.87th centile) would imply a clean cut at 142 (99.74). This is not consistent with the known population of high-range candidates; most of them are below 142, or at least I believe the evidence for that is more than sufficient.

My experience is that the median of many high-range scores is almost always between 136 and 141. The fundamental cause of this, I think, is that only from the low to mid-130s onward people are interested in intellectual endeavours like taking difficult tests. Below that, it tapers off steeply. Above that, it tapers shallowly, and that shallow curve reflects the actual distribution of those high I.Q.'s in the general population. And this distribution is apparently such that the median of people wanting to take high-range tests ends up around 136-141. The mode is several points lower than the median, the mean several points higher. The mode probably represents the point from whereon the high-range distribution follows the general population distribution (upward). The mode is, more or less, the cut-off point meant in the previous paragraph.

Jacobsen: The following are questions formulated based on input questions provided by Matthew Scillitani. What is the process of making preliminary norms before submissions have been given for a test?

Cooijmans: If it is a fully new test and no data exists for its contents at all, I estimate the minimum raw scores that a Glia Society member and a Giga Society member, respectively, should obtain. So for each problem, I look at it and ask myself, "Should a Glia/Giga Society member be able to solve this?" Then I interpolate between those two scores, and extrapolate outward until I reach the edges of the test, where I taper with 5 protonorm points per raw score point. The edges are each sized half the square root of the total possible raw score range.

Jacobsen: There seems to be a stigma around high-range tests. Is there a process to normalize taking them or having them exist in the first place?

Cooijmans: There are indeed many who do not take high-range tests seriously, and this includes prominent figures like the late Hans Eysenck. In one of his "test yourself" books, I remember he was skeptical about the possibility of measuring intelligence in the high range, and

even ridiculed it. He provided a number of absurdly complex problems “for the super-intelligent”, which appeared to be a parody on high-range testing.

Much of the distrust and denial regarding high-range testing stems from the fear that one might not oneself belong to the most intelligent; it is comfortably reassuring to say to oneself, “Those tests are just puzzles by amateurs and their scores are meaningless, we can not measure intelligence beyond the 99th centile”. It is a way to protect one’s delusion that no one is verifiably smarter than oneself.

Another cause of the stigma is the inescapable fact that there are fewer women than men in the high range. This is such a taboo that denying the validity of high-range testing is imperative to the politically correct academic, if only for that reason.

A possible process to normalize high-range testing would be to establish it as a recognized branch of psychology at universities. I suspect this would require that we first reverse the decades-long neo-Marxist occupation of academia and make universities into places of genuine science practised by the most intelligent again. A concrete application of high-range psychometrics would be to devise proper admission procedures for universities to undo the dumbing-down that has taken place there over the past half century. The fact that the old Scholastic Aptitude Test and Graduate Record Examination were about the only mainstream tests with validity in the high range illustrates how appropriate high-range testing is in the context of college and university.

For completeness, it should be mentioned that psychologist Lewis Terman (1877-1955) has tried to measure intelligence in the high range with two forms of his “Concept Mastery Test”. These were applied to subjects selected as children based on childhood scores of 140 and higher, and followed up in adulthood with the two Concept Mastery Tests. These were verbal tests highly loaded on vocabulary, not permitting references aids. In an unsupervised situation (which was and is how they are typically administered) it is exceedingly easy to cheat on such a test by using dictionaries and thus score absurdly far above one’s real level. Also, non-natives of the English language have a large disadvantage, in the order of 30 I.Q. points. So while these tests were non-robust against cheating and strongly culture-dependent, at least he tried. Since Terman has also been criticized for his belief in eugenics, heredity of intelligence, and racial differences therein, he forms an intersection between high-range psychometricians and hereditarians, so to speak.

Having mentioned the Concept Mastery Tests, I should warn that the scores mostly quoted for them are raw scores, not I.Q.’s. Ronald K. Hoeflin has administered those tests for a while too, also unsupervised, so one should not rely too much on possible reported Concept Mastery scores from test candidates as they may be hugely inflated through fraud.

Jacobsen: Have test construction and norming processes evolved in the aggregate for you?

Cooijmans: Of course, when one has been doing something for decades, one has implemented improvements. If I have to give examples, I have become more concerned with locking in a unique answer and avoiding ambiguity and subjectivity in scoring, and I am also inclining more to having tests contain a surplus of difficult problems and a minority of easier ones. Regarding norming, one of the first things I learnt was that z-score equation – equating means and standard deviations – results in incorrect norms because raw test scores tend not to behave linearly, which is required for z-score equation to make sense. So I went with rank equation. Over the years I automated ever more of the process, so that now I can norm a test in 10 to 30 minutes mostly, while originally this took several whole days.

I also learnt to formulate problems better to avoid misunderstanding. For instance, people skilled at mathematics may have a bizarre deformation that makes them interpret numbers differently from normal humans. If I say, “There are three apples on the table”, any sane person will understand that there are three apples on the table. But not mathematicians! The mathematician will understand that there are three OR MORE apples on the table. Because the mathematician thinks, “If there are four or five or six… apples on the table, there are three apples on the table too”. So to the mathematician you have say, “There are exactly three apples on the table”.

Jacobsen: What are the easiest and hardest parts of norming and constructing of a test?

Cooijmans: Easiest: Finishing off the eventual test once the problems have been conceived, and creating the database fields that will receive the incoming submissions. Also, norming is easy on the whole. Hardest: Creating the problems. This has got ever harder, the more tests I made. I try not to repeat myself too much, and try to take into account that the Internet as a search tool has become ever more powerful. The various types of fraud are hard to deal with. I have no sympathy or tolerance for the individuals behind it. The hardest nowadays is to create test problems that are robust against the developments that enable dishonest people to cheat. Those who have spread test answers should reveal the names of the recipients of the answers, so that we can clean up the statistics. And if they sold answers for money, they have to refund, and possible profit they made by investing the fraudulently acquired money should be donated to a good cause.

Jacobsen: Of your tests–51 in-use & 57 retired, which ones are special to you?

Cooijmans: To name a few, Test of the Beheaded Man, Cooijmans Intelligence Test (any form), Daedalus Test, The Nemesis Test, Test For Genius (any form), Only Idiots, The Gate, The Piper’s Test, Dicing with death, The Smell Test. Each in their own way, they demand the candidate to operate at the summit of cognition in ways that are not trivial but tie in to the essence of existence itself. That is what I have generally striven for.

Jacobsen: In pre-2000, you wrote some articles in Netherlandic on test design. Are there any insights from those articles not replicated here or elsewhere worth replicating, or reiterating, here?

Cooijmans: I looked through the articles, and the following points may be worth mentioning:

Marilyn vos Savant occurs briefly in one article; she is known for having “the world’s highest I.Q.” according to the Guinness book of world records. I would like to add here that someone once showed me a copy of a page from Megarian No. 6 (October 1982) where her actual scores on the Stanford-Binet and preliminary Mega tests are reported. “Megarian” was the journal of the Mega Society then.

Also nice is the early history of Mensa, as related by founder Victor Serebriakoff in one of his books, which was reviewed by David Gamon in the Mensa International Journal of January/February 1995. The founders at the time believed to be selecting members at the level of 1 in 3000 (some sources say 1-in-6,000) but later discovered a mistake in the procedure, as a result of which they had been selecting at 1-in-50. Not wanting to send the bulk of members away again, they left it as it was.

Also mentioned somewhere is Kevin Langdon, creator of the Langdon Adult Intelligence Test (1977, I think) and founder of the Four Sigma society. If one is interested in high-range psychometrics, the statistical reports published by Langdon in the 1970s and later are worth looking at. Langdon’s approach differed from Hoeflin’s in that Langdon first expressed the candidate’s performance as “scaled score” (some conversion of the raw performance) and then equated means and mean deviations of scaled scores and scores on other tests, resulting in a linear relation between I.Q. and scaled score. Hoeflin, on the other hand, normed raw scores directly via rank equation, resulting in a non-linear relation that reveals the non-linear nature of simple raw scores.

This is a good time to explain there are different ways to arrive at a scaled score: The simplest way is to scale raw scores linearly from 0 to 100 or 0 to 1,000, for instance. Some test constructors have done that (Alan Aax and Rijk Griffioen, I remember) but it brings no advantage compared to raw scores; the non-linearity of raw scores remains, obviously, when the relation between raw and scaled scores is linear.

If the goal is to obtain a more linear (intervallic) scale, there has to be some weighting or balancing, and a crude but solid method is to give a certain class of problems that appear harder or more important extra credit a priori, regardless of item statistics. This was done by Hoeflin with the Ultra Test, where non-verbal problems get two points. This is effective and without problems, but the resulting weighted scores are still far from linear, if one had any concerns about that.

A more refined way is to give items individual weights based on item analysis. In theory this should yield an intervallic scale, but there are serious disadvantages: (1) A small number of problems tend to carry most of the weight after weighting thus, which is always dangerous; (2) It adds an extra layer of sampling error because one relies on the correctness of the item statistics, and my experience is that item statistics are not constant but differ from sample to

sample, so that one is building on quicksand as it were; (3) The intuitive simplicity of a raw score is lost; the candidate can not know the number of correct answers from the weighted score.

My preference is to use a simple raw score, or in cases where it seems appropriate a crude weighting that does not rely on item statistics, such as in the example of the Ultra Test. If these methods do not result in a meaningful ranking of candidates, that test is bad to begin with and no advanced item weighting will fix it. I accept that raw scores are non-linear, and the conversion to linearity takes place in the norming of raw score to I.Q.

That last sentence leads to the question, “How do we know that I.Q. is a linear scale?” The answer is that I.Q.’s are deviation scores; they denote a distance to the mean in a hypothetical normal distribution. Note the word “hypothetical”; it is not claimed that intelligence follows a normal distribution in the physical reality. But the tacit assumption in statistics is that when a distribution is normal (Gaussian), its underlying scale is linear (intervallic). So when you force test scores into a normal distribution, you create a linear scale, or that is the unspoken idea. This is expressed in the way we identify points on the scale in terms such as “2 standard deviations above the mean”. This implies an underlying linear scale; after all, if the scale were not linear, the one standard deviation would not be the same as the other, so it would make no sense to say “2 standard deviations above the mean”! In fact, the mere computation of an arithmetic mean assumes an underlying intervallic scale, as it involves summation.

So the bottom line is, if we take care that the frequencies of I.Q.’s beyond various points of the scale do not differ too much from their theoretical rarities in the normal distribution, we may assume that I.Q. is linear. I say this without claiming that deviation I.Q.’s are the best way to express intelligence; but I do not have a better way at the moment.

Jacobsen: Some submitted questions anonymously. These are the adaptations of those questions: Personally, do you know any geniuses? If you do not know any personally, where are all of the geniuses?

Cooijmans: I have to say that when it gets anonymous, the quality goes down. Imagine that I answered “no” to the first question! How insulting that would be to everyone I know! Since a genius is someone who exercises a lasting influence in any field, inherently it can only be known in hindsight who was one, like long after the genius’ death. It is well possible that several people I know will turn out to be geniuses, but we do not know yet who they are.

In history books you will find a lot of identified geniuses.

Jacobsen: Why refer to these individuals in this way, i.e., as geniuses? What traits characterise them?

Cooijmans: The word “genius” comes from the Latin “gignere”, meaning to conceive, to bring forth, to cause. Francis Galton used the word “eminence” for what is now mostly called genius.

The traits of genius, according to me, are intelligence, conscientiousness, and a wide associative horizon. Genius is not talent. It requires talent, but talent alone does not suffice. One will need to apply that talent in order to make a lasting contribution.

Jacobsen: Do you see yourself as a genius? If so, why? If not, why not?

Cooijmans: Naturally, someone of my enormous modesty and humility would never call oneself a genius. I leave that to the scores of future generations who will devote their lives to the study of my work.

Jacobsen: What do you think has been the contribution of your I.Q. Tests for the High Range? Is it a work for study by others or a hobby?

Cooijmans: The contribution lies in studying the measurability of intelligence in the high range, and some other questions related to that as stated at https://iq-tests-forthe-high-range.com/mission.html . It is certainly worthy of being studied by others, and others should also undertake such study independently. It is not just a hobby, except in the sense that one can make one's hobby into one's work.

Jacobsen: Who are others who you see like yourself in studying high ranges of intelligence?

Cooijmans: This can only be answered properly for people who were (already) working longer ago, before the current generation of high-range testers. That would be Lewis Terman, Kevin Langdon, Ronald K. Hoeflin, Xavier Jouve, and Laurent Dubois. For the ones who came after these, it is too soon to judge their merit.

In addition, there have been some people who created tests that looked truly good to me, but who only kept scoring their tests briefly and then withdrew from testing, so that little or no usable data resulted. These people exemplify what I said a few questions ago: that talent is not genius, but merely a requirement for genius. They had talent, but did not use it to make a lasting contribution to high-range testing.

Jacobsen: What is the most common mistake people make when submitting feedback about your tests?

Cooijmans: Assuming that they have understood a test item correctly, and then commenting on it from that assumption.

Jacobsen: What aspects of people's test feedback seem confusing?

Cooijmans: It can be confusing if people send feedback before sending answers. I have to be careful not to help them by responding. Nevertheless, in case the feedback concerns a mistake in a test problem, it can be useful, especially when a test is very new.

Jacobsen: The most common Marathon Test Numeric Section score is a perfect 44 out of 44. What lessons have you learned from this high-end score saturation?

Cooijmans: That the problems are not hard enough. Also that a series of similar problems of increasing difficulty tends to be too easy on the whole. And that, to make a numerical test hard enough, either very difficult mathematics-biased problems are needed, or problems that implement a pattern that needs to be recognized. The latter seem the most fair, the former seem to give an advantage to people skilled at mathematics.

Jacobsen: When creating high-range questions, is there a consideration of steering test takers toward wrong answers? Are extant questions ever modified in this way?

Cooijmans: Obviously, steering test-takers toward wrong answers is the whole point of creating good test items, not only in high-range testing but also in mainstream intelligence tests. There is even a word for it: distractors. Multiple-choice tests, omnipresent in mainstream psychological testing, contain answering options that are wrong but appear more plausible than the intended correct answer.

Thus, a candidate who really can not solve any problem at all will score below the chance guessing level, and this lower level is called the “pseudo-chance level”. For instance, if a test has 40 problems and 5 answering options per item, the chance guessing level is 8 correct, but the pseudo-chance level may be only 5 correct due to distractors.

Extant questions are not generally modified in this way.

Jacobsen: Which books or literature, even individual articles or academic papers, on psychometrics have provided helpful accurate understandings of psychological measurement, psychometric concepts, etc., for you? Others may find some fruitful plumbing there.

Cooijmans: The most specific sources regarding high-range testing are the various statistical or norming reports by Kevin Langdon and Ronald K. Hoeflin, as issued by them in the 1970s through 1990s (Hoeflin only started in the 1980s, I think). These helped to see how high-range tests are normed, and also aided in the interpretation of scores on a lot of those old American tests like the Scholastic Aptitude Test, Miller Analogies Test, Army General Classification Test and more. The “*Omni* sample” of the Mega Test contains many scores of those versus Mega Test scores, and as such is an important anchor of high-range tests to the general population, especially so since those old tests in some cases did discriminate into the high range. This can not be said about the newer dumbed-down versions of the educational and military tests, whose validity tends to end at the 99th centile, and for whose interpretation one should consult the information provided by the relevant issuing organizations.

What one can see for instance in the *Omni* sample is that the old G.R.E., S.A.T., and Army General Classification Test correlated quite well with the Mega Test, while on the other hand the Wechsler Adult Intelligence Scales and Stanford-Binet, by many regarded as the gold standard

of I.Q. testing, appeared to lack any validity in the high range. This observation holds true until today in data collected by myself, except for the old Army General Classification Test on which I have almost no data.

Then, an actual text book on psychometrics I have studied is the Netherlandic “Testtheorie” by P.J.D. Drenth and K. Sijtsma from 1990. This covers both classical psychometrics and the newer item-response theory.

Another useful book, a bit more general, is “Applied statistics for the behavioral sciences”, second edition, by Hinkle, Wiersma, and Jurs, from 1988.

An important book on intelligence testing is *The g factor* by Arthur Jensen, from 1998. While not intended as a psychometrics textbook, it does contain a lot of advanced information on psychometrics, including some factor analysis, often in the footnotes.

As it happens, there is also an e-book called *The g factor* by Christopher Brand, from 1996, also containing information on psychometrics and some factor analysis.

A book on statistics in general (not psychometrics) I have studied is the Netherlandic *Statistiek in de praktijk* by David S. Moore and George P. McCabe. I see there is an English version too, *Introduction to the Practice of Statistics*; 2nd edition (1993).

A book dealing specifically with multivariate statistics such as correlation, regression, and factor analysis is *Using Multivariate Statistics* (third edition) by Barbara G. Tabachnick and Linda S. Fidell (1996).

I also still have my mathematics books from secondary school, one of which contains chapters on statistics and probability calculation. Occasionally I look through those to refresh these basics of my knowledge in this field.

Finally I want to add that the history of statistics and of mathematics is informative regarding psychometrics. Reading about such will teach you that statistics has been closely related to psychological testing since the 19th century, and that probability calculation was developed for the purposes of gambling and insurance.

The history of mathematics in general, found for instance in *A Concise History of Mathematics* by Dirk Jan Struik, tells us that mathematics originates in the early days of agriculture, cities, and large-scale administration. That is, within the past ten thousand years or so, the holocene, after the last glacial period. Computing the area of parcels of land required mathematics.

I suspect that the intelligence of the people coming out of the last glacial period was primarily of a visual-spatial nature, and as they became settled and practised agriculture, built cities, and administrated societies, they needed higher numerical ability as well as written language. I imagine that spoken language existed long before that, originally in the form of words without

grammar some two million years ago to coordinate hunting in groups in early Homo, and later on with grammar, perhaps in the days of Homo sapiens.

Language is not unique to humans incidentally, but exists in other beings too, such as birds, primates, and whales. Animals like crows are likely at the intelligence level of early Homo, but I am uncertain if their physicality will allow a further development such as has taken place in Homo. Key points like the manufacturing of tools and mastery of the fire may require arms, hands, fingers, and thumbs such as humans have.

Visual-spatial ability is also not restricted to humans, but found in many animal species, in particular to enable predating. As such, visual-spatial ability should be a few hundred million years old, as that is when the first predators came.

The importance of this history of abilities is that when we test abilities now, the results we get, such as the intercorrelations of various abilities, are as it were a fossil record of this evolution. A development that I believe takes place in civilized societies is the erosion of the original visual-spatial ability in favour of verbal ability. A high level of verbal ability in the absence of the foundation of visual-spatial ability, I think, leads to dishonesty, deceit, evil, decadence, and societal collapse.

Jacobsen: Thank you for the opportunity and your time, Paul.

Cooijmans: I never know what to respond to here.

Interview with Steven Pinker on Humanism and Campuses

Steven Pinker & Scott Douglas Jacobsen

Steven Pinker is an experimental psychologist who conducts research in visual cognition, psycholinguistics, and social relations. He grew up in Montreal and earned his BA from McGill and PhD from Harvard. Pinker is a Professor of Psychology at Harvard; he has also taught at Stanford and MIT. He has won numerous prizes for his research, teaching, and books, including *The Language Instinct*, *How the Mind Works*, *The Blank Slate*, *The Better Angels of Our Nature*, *The Sense of Style*, and *Enlightenment Now*. He is an elected member of the National Academy of Sciences, a two-time Pulitzer Prize finalist, a Humanist of the Year, a recipient of nine honorary doctorates, and one of Foreign Policy's "World's Top 100 Public Intellectuals" and Time's "100 Most Influential People in the World Today." He was Chair of the Usage Panel of the American Heritage Dictionary and writes frequently for *The New York Times*, *The Guardian*, and other publications. His twelfth book, published in 2021, is called *Rationality: What It Is, Why It Seems Scarce, Why It Matters*.

Scott Douglas Jacobsen: So we are here again with Professor Steven Pinker, one of the most prominent humanists around, particularly around the exhaustive research you do on various topics, dispelling myths around increasing violence– the fact that violence is declining. Things of

this nature. Some of the recent news that has popped up has been about how students feel on campus about wanting to be able to speak more freely. This is probably more particularly prominent in the American context with the First Amendment there. What are your reflections over the last decade on campuses where there has been pushback to bolder speech around issues that might be either new or perennial controversies?

Steven Pinker: Well, the pushback is very recent, and there is a very strong feeling among American university students that you have to watch what you say, that you cannot speak your mind, and you never know when you might commit racism, that you might commit some political sin and be cancelled, what used to be called excommunicated. The universities have not done a good job of fostering an environment of free speech. There are often student orientations in which they are warned about how they can commit a microaggression if asked somewhere, "Where are you from?" That can be considered a form of subtle racism. If you say, "Oh, you speak very well," that can be a form of racism. So, they are often terrified. I am not even talking about controversial political or scientific opinions. I am talking about ordinary interactions where they feel like they must walk on eggshells. This leads to the paradox that many American university students in their dorms are in adolescent heaven. Their peers surround them. They are constantly invited and given opportunities for socializing and recreation. They eat with each other, but they say they are lonely. How can this be?

We have reason to believe that in adolescents and young adults there is an increased risk of anxiety and depression, given that social interaction is one of the most important elixirs for mental health. Why is this possible? I suspect that the fact that interactions are so policed and so guarded means that social opportunities for interaction, far from being opportunities to relax, kick back, and laugh together, are more sources of anxiety. Particularly when a lot of it is done on social media, where you have to worry about being mobbed in real-time, anything you say can be dug up decades later by offence archaeologists and used to cancel you retroactively. None of this even gets to the expression of opinions on political, social, or scientific issues.

Jacobsen: Right, I like that. I like that step back from touching on social dynamics.

Pinker: A lot of social media technologies, too. I suspect, and we do know there are cases, a famous or notorious case at Harvard where a student was admitted and then the admissions office rescinded his admission when one of his social enemies uncovered a late-night chat when he was 15 years old in which he was throwing around racist terms to be transgressive. That he and his friends would be "bad boys." Harvard withdrew the admissions offer. So you have to worry not about what you might say in an op-ed or a paper where you are formulating your opinions, but when you are kicking back in a chat room. It might come back years later to ruin your life.

Jacobsen: So that will not lead to conversation, whether it be social or intellectual. There will be some people who, in response, will say, "Good, they got their comeuppance for the things they have done." I am sure you live and work in that world. What happens in those contexts?

Pinker: One quick note that one of the side effects is the epidemic of mental health problems, together with the cases in which that general attitude of censorship and cancellation leads to entire societies adopting the wrong policies or being in the dark as to major issues, such as the effects of, say, school closures and masking during COVID, where there appears to be tremendous harm on a generation of children losing out on a year or two of education based on what turns out to be a very trivial risk of their degree of harm. At a time when it was considered taboo to criticize policies of masking children during school closures and widespread shutdowns, bringing it up would lead to massive condemnation. If there had been a greater commitment to free speech and people not being punished for their opinions, realizing that these policies are harmful may have come sooner.

Jacobsen: People will probably consider this a largely academic phenomenon outside of the social media landscape. People from more ordinary backgrounds working blue-collar jobs and do not necessarily need higher education for their pursuits might think, “It is a humanistic thing that we should generally care about, but why should I, as a blue-collar person, necessarily care about this?”

Pinker: Well, partly because many blue-collar people are on social media, but also, what happens in academia does not stay in academia. About 10 or 15 years ago, people argued, “Who cares what kids get taught or what censorship regimes are implemented in academia? When students enter the real world, they will find they cannot escape this nonsense.” What we know happens is that the whole generation brought the regime of cancel culture into the workplace, so, publishing houses, newspapers, nonprofits, and artistic organizations are being torn apart by the regime of cancel culture, microaggressions, and constant accusations of racism because they have been exported from universities, including blue-collar people being fired from their jobs because of some accidental offence – precisely because the culture of the universities was then taken into the workplace and government and nonprofits.

Jacobsen: So, eventually, this does not only chill academic life; it also chills general culture.

Pinker: Yes, well, it is a chill in that the culture of academia is often brought into other institutions by the graduates of universities as they take positions of power. However, when it comes to societies making collective decisions based on an academic consensus, it can often be the wrong consensus if academia is churning out falsehoods because ideas cannot be criticized. I mentioned the effect of school closures and masking children. However, the other example is even the origin of SARS-CoV-2, where it was considered to be racist to suggest that the virus might have leaked from a lab in Wuhan. We do not know that that is true, but it is not implausible; it might very well have happened.

If it is true, it would have a major implication that we have got to ramp up lab security drastically, perhaps not do gain-of-function studies of the kind that could have created this virus, on pain of suffering from another catastrophic pandemic if we do not learn the lesson. So, that is a case in which what academics decide can affect the world’s fate. Another example would be the effectiveness of policing. If there is reason to think that after the George Floyd demonstrations

and the riots of 2020, the idea that police do not matter or that there is an epidemic of shooting by racist cops may have led to withdrawals of policing that then caused the violent crime, if that understanding of an epidemic of racist shootings had been put into context in the first place, they knew that there are not that many shootings of unarmed African Americans by cops, that this was a false conclusion. Journalism has as much a role in this as academia, but journalism has also developed a regime of cancel culture, where heterodox opinions are often firing offences. If the nationwide consensus is distorted, society will adopt policies that worsen it. Finally, one other thing, and I will turn it back to you, is that even when the academic consensus is almost certainly correct, as in the case of, say, human-induced climate change, if scientists, government officials, and scientific societies have forfeited their credibility by ostentatiously punishing dissenters, leading to the impression that they are their cult, we could blow off their recommendation because if anyone disagreed, they would be cancelled. So it is another cult, it is another priesthood, it is another political faction. The scientific consensus loses credibility if it comes from a culture known for intolerance of dissent.

Jacobsen: We could probably iterate that across domains, whether it is the combat over creationism, or vaccines causing autism, and things of this nature.

Pinker: Yes, so if the scientific consensus tries to debunk it, then no one has enough scientific competence to review everything scientists say perfectly. Some of the acceptance of the findings of science has to be committed trust; these are people who know what they are doing. They have means of distinguishing true from false hypotheses. If something they believed were false, it would be self-correcting. If you undercut that assumption, then people will blow off what scientists say. Scientists themselves are surprisingly oblivious to this possibility. Many scientific societies churn out a woke boilerplate, branding themselves as being on the hard political left and cultural left, with no appreciation that this may alienate the people who are not on the left or in the center who do not care but perceive science as another faction.

[https://www.npr.org/2024/11/15/nx-s1-5193258/scientific-american-editor-resigns-after-comments-about-trump-supporters-went-viral -Editors' Note]

Jacobsen: What areas are incursions of what is called something like woke ideology or wokeness into academic and empirical findings or before the empirical findings impact a lot of academic and professional life? So, at the highest level, where people are tenured professors, it is an ideological strain pushing against proper consideration of the evidence.

Pinker: It is worse in the humanities than in the social sciences, worse in the social sciences than in science and engineering. Although, those are generalizations. Probably worst of all, the branches of humanities and social science that are sometimes denigrated as grievance studies are often departments of women and gender studies or studies devoted to particular ethnic groups. Some of the social sciences are worse than others. For example, cultural anthropology is a lost cause. There has been such ideological capture. Most of my field, psychology, is not nearly that bad. Although, there are strains there. Sociology is divided; there is a branch of more quantitative sociology, verging into economics, that is pretty empirically oriented, but then there

is another far more ideological part. Even the hard sciences, particularly the scientific societies, have plenty of wokeness, even though the actual lab scientists may be more neutral or empirically oriented. However, the societies themselves tend to be “woker” than their members.

Jacobsen: Why are societies more likely to be captured than individuals?

Pinker: Yes, it is a good question, partly because of the selection of who goes into societies and institutions. If your heart and soul want to do science, you will be in the lab, getting your hands dirty with data. If your motivation is more political, verbal, or ideological, you will try to become a magazine’s editor or a society’s spokesperson. There is a tendency for institutions to drift leftward. Robert Conquest, the historian, is sometimes credited with a law that states that any institution that is not constitutionally right-wing becomes left-wing. You can see the drift that has happened to many institutions recently. They have not become left-wing in the economic quasi-Marxist sense but “woke” in the sense of identitarian politics, seeing culture and history as a zero-sum struggle among racial and sexual groups. A kind of intolerant identitarian politics has captured several societies with well-defined intellectual goals. It has happened to the ACLU, the American Humanist Association, and Planned Parenthood.

[‘In 1921, Sanger founded the American Birth Control League, which later became the Planned Parenthood Federation of America.’

‘Sanger was a proponent of negative eugenics, which aimed to improve human hereditary traits through social intervention by reducing the reproduction of those who were considered unfit.’

https://en.m.wikipedia.org/wiki/Margaret_Sanger

https://en.m.wikipedia.org/wiki/Planned_Parenthood

https://www.aclu.org/news/lgbtq-rights/sex-work-is-real-work-and-its-time-to-treat-it-that-way

https://www.aclu.org/news/topic/its-time-to-decriminalize-sex-work

-Editor’s Note]

So, selection is part of it. Another part may be the belief that the way to change the world is through the imposition of verbally articulated philosophies, as opposed to a bottom-up approach of experimentation, data gathering, entrepreneurship, trying things out, and seeing what happens. The top-down approach is much more likely to start with a predefined narrative and to try to impose that narrative. There may be something more pleasant to institutions in this approach.

To a more left-wing mindset. To elaborate on that a little bit, this comes from Thomas Sowell. Some systems achieve order spontaneously and in a distributed fashion, market economies being the most obvious example—the invisible hand. No planner decides how many size eight

shoes to make or where to sell them. The millions of people making choices proliferate information in markets, and the system becomes intelligent, with no one articulating exactly why. The evolution of a language works that way; a culture with its norms and mores works that way. There is a kind of sympathy for these distributed systems that are more on the right, and historically, there are many exceptions. However, on the left, there is more of an articulation of foundational principles, which is a good theory. Therefore, you are more likely to try to change things by joining an institution that can pass resolutions and implement verbally articulated policies. Conversely, on the right, people will go into business, try to invent things, and hope the invention will take off as part of this more distributed, bottom-up approach.

Jacobsen: Do you think the general humanistic approach is akin to an evidence-based moral philosophy where you work bottom-up and then formulate the principles of your ethics from that, rather than top-down, as you might find in divine command theory?

Pinker: There is some affinity in that humanism starts from the flourishing and suffering of individuals. When that is your ultimate good, instead of implementing scriptures or carrying out some grand historical dialectic or privileging some salient polity or entity like a nation, or a tribe, then, if you are a humanist, you see the point of a society, a religion, and so on, is what will leave those people better off.

My pleasure, thanks for the time to talk to you, Scott.

Jacobsen: Excellent. Take care. Bye.

[Editors' Note: https://www.hoover.org/research/quest-cosmic-justice]

Interview with Benoit Desjardins on Medical Professional Balance

Benoit Desjardins & Scott Douglas Jacobsen

Dr. Benoit Desjardins, M.D., Ph.D., FAHA, FACR, FNASCI, CEH, CISSP, is an Ivy League physician who is a world leader in three different fields (cardiovascular imaging, artificial intelligence, cybersecurity) and recently left the U.S. after significantly traumatic events.

Extended Bio: Dr. Benoit Desjardins, M.D., Ph.D., FAHA, FACR, FNASCI, CEH, CISSP, is Professor of Radiology at the University of Montreal. He recently retired from the University of Pennsylvania after 16 years on faculty. He is an international leader in three different fields: cardiovascular imaging, artificial intelligence, and cybersecurity. He has given over 200 invited presentations nationally and internationally in those three fields. He was co-leader of the Arrhythmia Imaging Research Laboratory at Penn. His research involves cardiac MRI and CT in electrophysiology, focusing on the relation between cardiac biomarkers such as myocardial scar, with pathways of abnormal electrical conduction in left ventricular arrhythmia. He is funded by the National Institute of Health and is very active in national scientific societies. He has extensive expertise in artificial intelligence, the field of his PhD. In the spring of 2022, he spent six months at Stanford as Visiting Professor and Associate Scholar of the Stanford Center for Artificial Intelligence in Medicine and Imaging. He is a reformed hacker. He has several certificates in cybersecurity and has done research and published on the cybersecurity of medical images. Outside work, he is a Black Belt at Taekwondo, an ex-Boy Scout Leader, a competitive marksman, and a FPV race drone pilot. He is also a member of the prestigious Mega Society and Prometheus Society.

Scott Douglas Jacobsen: Previously, you told a heartbreaking story of anxiety, stress, and degrading health, as with many American medical professionals. Does this start in medical school?

Dr. Benoit Desjardins: I am extraordinarily lucky to be alive today to let the readers catch up on the story. As you know, a few years ago, on a Friday afternoon on my 97th hour of work as a U.S. physician, at the end of a week during which I was not allowed to sleep much or eat much, and on a day which I was forced to do the workload of six doctors, the combination of lack of food, lack of sleep, and massive overwork made my body permanently fail. I almost died from a catastrophic medical condition caused by the work conditions and became handicapped for life. This was not the first time that I was physically hurt by these work conditions and not the first time that they almost killed me. But it was the first time that they caused permanent, severely limiting lifelong damage to my body.

To answer your question, I attended medical school in Canada, which has strict rules and laws on basic human rights, including those of physicians. In the U.S., physicians' working conditions are massively out of compliance with safe labour laws from all other industries. In 2019, Dr Pamela Wible published a book listing 40 categories of documented human rights violations towards physicians in the U.S. ("Human Rights Violations in Medicine: A-to-Z Action Guide"). This included sleep deprivation, food deprivation, overwork, exploitation, bullying, violence, etc. I have experienced most of those as a physician in the U.S. Since around 2014, the U.S. has been well-known for the inhumane work conditions of its physicians, killing and disabling its physicians by the thousands and burning out its physicians by the hundreds of thousands.

[https://www.researchgate.net/publication/341634090_Human_rights_violations_in_medicine_a-to-z_action_guide_by_Dr_Pamela_Wible -Editors' Note]

After medical school, I came to the U.S. in the early 1990s to pursue a PhD degree. I was initially a graduate student in the U.S. I was treated like everybody else. It was a rude awakening when I started in the U.S. medical system after my PhD. Here is one of many examples of what I faced: As a medical post-graduate trainee, I had once been forced to work at the hospital for 58 consecutive hours without rest and then drove back home. As my exhausted body crashed into my bed, I received a phone call from the chief resident asking me why I had left the hospital as I was apparently on call again for a third night in a row. He ordered me to get back to work. I drove back to the hospital, completely exhausted. I could have easily been killed in a car accident from exhaustion, like what happened to two of my immediate radiology colleagues. After arriving at the hospital, I was forced to work ten additional consecutive hours (for a total of 68 consecutive hours without sleep), until I crashed on the call room floor out of exhaustion. They found me unconscious later that morning. This is one of many examples of the work conditions of physicians in the U.S.

Jacobsen: When medical professionals enter into medicine in Canada and the United States, what are the contrasts in treatment and the similarities in treatment of medical professionals?

Desjardins: There are huge differences. We can divide this treatment into the public, employers, and government.

(1) by the public: In Canada, the public is respectful of physicians, of expertise and science, partly because the population is well-educated and scientifically literate and partly because access to healthcare is more restricted, and patients are very happy when they can access a physician. Canadians understand that physicians are human beings. In the U.S., the public has no respect for healthcare professionals, expertise, or science. Physicians and nurses regularly get attacked by patients, and sometimes get killed by them. One physician in Philadelphia recently got stabbed in the face by her patient. Also, physicians in the U.S. are viewed as lottery tickets. The strong anti-science culture in the U.S. has people making irrational cause-and-effect magical expectations of doctors. Any bad medical outcome, a regular part of medicine, almost invariably leads to a lawsuit that can produce a multimillion-dollar award.

(2) by employers: In the U.S., this was nicely summarized by the 2019 *New York Times* op-ed article "The Business of Health Care Depends on Exploiting Doctors and Nurses" by Dr Danielle Ofri. She discussed how the U.S. healthcare system involves massive exploitation of healthcare workers to stay in business. The nature of the exploitation depends on the environment, either academic or private practice. In academia, physicians are salaried and academic hospitals maximize the work done by physicians to avoid bankruptcy and maintain their razor-thin profit margins. The amount of work never stops increasing. Private practices are being bought one after another by venture capital firms, whose only goal is to maximize short-term profits for their investors, by forcing physician employees to do a massive amount of work with the lowest resources while disregarding quality of care. In Canada, almost all physicians are government employees, which is very different and will be covered next.

[https://www.nytimes.com/2019/06/08/opinion/sunday/hospitals-doctors-nurses-burnout.html]

(3) by the government: In Canada, the government is the main employer of physicians and exerts very strict control on the location of physicians' practice to ensure adequate distribution throughout the country. However, besides these limitations on their practice, physicians are treated like human beings by the government, with strict laws and rules on basic human rights and physician work conditions that must be respected. The treatment of physicians by the government in the U.S. is well-illustrated by the recent scandal of the PHPs (physician health programs). If, for example, a patient sees a physician drink a glass of champagne at a wedding, she can report him to the U.S. government as an alcohol abuser. Then, under the threat of losing his medical license, the physician gets forced by the government to attend an out-of-state "addiction" government therapy program, costing tens of thousands of dollars. This has led to several bankruptcies and dozens of suicides of physicians while in those PHP government programs. This included prominent doctors, such as a visionary in a pediatric field, who helped thousands of pediatric patients. He committed suicide after a government PHP program ruined his reputation and career. He had been forced into this PHP program by his employer after he reported dangerous local work conditions putting patient lives at risk.

Jacobsen: The conditions at your prior job sound slavish. Is there a cycle of entrapment and overwork among medical professionals?

Desjardins: When you get a job as a physician in the U.S., you get a state license enabling you to practice, which is a long process. Then you get installed, your spouse gets a job, and your kids attend local schools. You become locally established, and relocation becomes a major hassle for the physician, his spouse and kids, so the threshold for relocation is very high.

When I got to Philadelphia in the late 2000s, things were tolerable. However, the situation for physicians worsened progressively. It's like being a frog in progressively warming water. 2014 was a turning point in Philadelphia for two independent reasons. First, as I already mentioned, the U.S. has inhumane work conditions for its physicians. This became public knowledge around 2014, when the American Medical Association started its first three Physician Wellness programs to try to address the problem. Second, Philadelphia became known as having the most massively corrupt, scientifically illiterate medico-legal system on the planet. This is beyond the scope of this interview. But it's the last year we could recruit any radiologist in my section and the year when physicians started leaving Philadelphia by the boatload. Before 2014, we individually read about 15,000 images per day. Now, it's sometimes up to 250,000 images per day.

One of the advantages of my field of radiology is that we do not need to be close to patients. We can read medical images remotely. We took advantage of that during the pandemic, as most radiologists could do their full work shifts from home, without needing to enter the hospital and be exposed to COVID. This gave many radiologists an important escape route. When remote work became a viable option for radiologists after the pandemic, many entrapped in Philadelphia abandoned their local jobs and signed remote work contracts with out-of-state hospitals while remaining in Philadelphia. The workload for radiologists who did not abandon Philadelphia hospitals rapidly increased. We are living in the absurd situation of being surrounded by dozens of local radiologists whom we desperately need but who refuse to have their names ever associated again with Philadelphia hospitals. When we tried to do the converse and recruit out-of-state radiologists to work remotely for Philadelphia hospitals, we learned that most radiologists in the country refuse to ever have their names associated with hospitals in Philadelphia because of medico-legal reasons. The long-term implications of this situation are unclear but frightening.

Jacobsen: What health problems arise in this context?

Desjardins: We recently discussed extensively the healthcare effects of excessive workloads on human beings, which can lead to all sorts of chronic medical conditions and even death.

Jacobsen: Whether by death, health injury, or moving away, medical professionals do leave those conditions, as you recently informed me – with a perceptible tinge of elation as if a proverbial sigh of relief. How did you begin to find a way out?

Desjardins: I'm an Ivy League physician and a world-leading expert in my medical and scientific field. I used the same approach to solve all my scientific and clinical problems to find a way out. I was forced to continue working under the same work conditions that had almost killed me and disabled me for life. I needed urgent action. I selected a combination of two basic moves: (1) increase my protection and (2) remove myself from the toxic environment. To increase my protection, I started being closely monitored by a team of three physicians and taking protective medication to decrease the chances of recurrence of the event that permanently disabled me.

Removing myself from the toxic environment was more difficult. Physicians cannot change jobs easily. If you try to relocate locally, you face non-compete clauses preventing access to jobs at other institutions. If you try to relocate to another state or country, getting a new practice license for that new location takes months, and time was not on my side. Abandoning the medical profession was also an option, recently taken by thousands of physicians. I did not consider that option, as I am a world leader in academic radiology. My field needs me, and I have a lot more to offer to my field.

I initially secured a quick research sabbatical at Stanford, giving me six months out of that toxic environment. This gave my body time to cope with my new handicap and time to plan my long-term escape from Philadelphia. This was near the end of the pandemic. During those six months of sabbatical, I interviewed widely and secured four U.S. academic positions away from Philadelphia and was working on securing two positions in Canada. However, the work conditions of U.S. healthcare workers during the pandemic resulted in a massive exodus of healthcare workers from the profession, with even more planning to exit in the short-term future. Under these circumstances, I felt Canada offered a much better future.

Canada has a mechanism to recruit Nobel Laureates and international scientific superstars called the "Distinguished Professor" pathway. There are other mechanisms to recruit regular doctors. To be recruited under that pathway, one must be a world luminary in a specific field. I'm a world luminary in three different fields. However, this pathway takes one year to receive government approval. When Canada found out that I had been almost killed and had become disabled for life by the work conditions of U.S. physicians and that I was still forced to work under these same conditions, they granted me a humanitarian exception and my "Distinguished Professor" pathway was approved in one week, instead of one year. This is how I got out.

Jacobsen: You mentioned some in the previous interviews. What happened to earlier professionals who did not get out and were trapped, in essence, in these areas? Those continuing to undergo harassment and threats, violence, including nurses.

Desjardins: Those who are still trapped are currently abandoning the medical profession by the boatload. In my previous U.S. department, we had a deficit of 43 doctors due to departures and difficulties in recruiting replacements. One of the four academic medical centers in Philadelphia (Hahnemann) collapsed and permanently closed under similar conditions.

In other countries, it is illegal to treat human beings the way the U.S. treats its physicians. No other industrialized country forces its workers to work up to 120 hours per week and up to 72 consecutive hours without rest, like the U.S. does to its physicians. Since the pandemic, 30% of all healthcare professionals have left the medical profession, and an additional 30% are expected to leave in the next 2-3 years. The U.S. cannot recruit fast enough to recover from these massive levels of attrition, which is a global phenomenon, while acute in the United States. The up to 60% deficit in healthcare workers will never be fully replenished, and massive shortages of U.S. healthcare workers will become chronic.

There are two ways to increase the number of U.S. physicians: recruit them from other countries or train more physicians at home. Both are a huge problem. The work conditions of U.S. physicians are now well known since 2014, and even more since the pandemic. Physicians from Europe and Canada could be recruited to the U.S. but they no longer want to come. The U.S. can however still attract physicians from third-world countries. Furthermore, there are more and more books, articles, blogs, movies, TED talks, and news clips about the U.S. treatment of its healthcare workers. The medical profession is much less attractive than it used to be to the best and brightest undergraduate students at home. This will continue to decrease the pool of top U.S. applicants for the medical profession. More than 60% of physicians currently highly discourage their children from entering the medical profession, and an even greater percentage of younger physicians who never experienced the good old days of the medical profession strongly advise their children NOT to enter the medical profession.

Jacobsen: When getting out, what area of medicine and geography in Canada did you choose? (And why those?)

Desjardins: In terms of areas of medicine, I needed to continue in the same field, as I am a world leader in that field. I have been responsible for determining the standards of practice in that field for the past 20 years, and it made no sense at this point in my life to change my area of medicine.

Regarding geography, I could have worked anywhere in Canada, but I wanted to be close to my family in Montreal, so I focused on academic places within two hours of Montreal. My top choice was the CHUM, Quebec's crown jewel of medical centers, in the heart of Montreal, at my alma mater. I ended up working there. It was a fantastic decision. I even have several medical school classmates working in my department or at my institution.

Jacobsen: What was the feeling and process of transition to new work and more reasonable work conditions?

Desjardins: At this late phase in my career, relocating was expected to be very difficult. But against all odds, things worked out very fast, and I was able to leave the U.S. I'm still in disbelief, thinking I will wake up and that this is all a dream. I suspect I'll remain in a phase of disbelief for a while.

Expats U.S. physicians often describe their newfound freedom as like being released from U.S. prison. This is, of course, a ridiculous comparison, as U.S. prisons don't kill and disable their prisoners by the thousands as the U.S. does to its physicians. But there are nevertheless many similarities between the two situations.

I now work 40 hours per week instead of 80+ hours. I am on call every eight weeks instead of up to 22 times per month. My daily workload is up to 6 times less than what it was in the U.S., and I have 6 times more vacation than I had in the U.S. This is almost unbelievable, but this is how physicians are treated outside the U.S. I maintain many work collaborations with the U.S., as an international leader in three fields.

I still need to get used to the new freedom. I had not been allowed to take many vacations in my last 20 years as a physician in the U.S.; when I did, it was to travel to see my family in Canada. Now, I live 10 minutes away from my family and see them every weekend. I have yet to schedule a big trip. Switzerland? Australia? Italy? The Greek Islands? An Alaskan cruise? There are so many good picks! I've travelled extensively for scientific meetings, but never for pure pleasure outside work.

Jacobsen: How has this better balance affected your life with family, as a husband – including treating her like a queen – and father, and in your ability to treat patients with full focus and care – not sleep deprived, overworked, and stressed to the point of high detriment to personal health?

Desjardins: Well, I now have a family life. I can now eat dinner with my family, spend weekends with them, and go on vacations. This is very liberating. I had always treated my wife like a queen and my kids as best as possible, but I knew my availability was very limited. Now, I am making up for lost time.

I am much more rested during my workdays. There is a massive difference between 4 hours of stressed-out sleep and 7 hours of relaxed sleep. My body feels the difference already. And since I do up to six times less work every day, I get to spend four times longer interpreting each study (8-hour workdays instead of 12+ hours workdays), dramatically increasing the quality of care I can provide. Workdays are not insane marathons anymore; they are normal days with normal work. Patients benefit from this process by accessing more rested, less stressed-out doctors in a better-quality healthcare system. This might partly explain why the Canadian healthcare system currently ranks 32nd in the world, compared to the U.S. ranking of 69th. Canada used to be much better than 32nd, but its waiting lists for care currently hurt its rankings.

Jacobsen: Why have the problems you described in the U.S. medical system not been solved? Why the hiding of physician deaths and suicides?

Desjardins: U.S. physician work conditions are now a very well-known problem. Books, documentary movies (*Do No Harm: Exposing the Hippocratic Hoax*, Robyn Symon), TED talks, publications, and numerous blogs exist. The American Medical Association is aware of the

problem and has implemented solutions. Since 2014, Physician Wellness has been a major focus of discussion in medical centers, conferences, blogs, and medical schools. Most people in the public are not even aware that almost every U.S. medical center has a Physician Wellness program to try to stop U.S. physicians from dying by the thousands and burning out by the hundreds of thousands. These programs, which teach physicians resilience rather than improving their work conditions, have been compared to distributing Yoga mats to prisoners at Auschwitz during World War II.

Publicity on this topic is blocked by hospitals. Hospitals in the U.S. are businesses. They must hide the negative consequences of physician work conditions to be able to stay in business. If a hospital disclosed to the news media that three of its physicians jumped to their death from the roof of the hospital within a month of each other, like what happened recently in a New York hospital, this would affect the hospital's financial bottom line. After these three New York physicians jumped to their death, their bodies were simply covered by tarps, and this did not even make the local news. Their colleagues at the hospital were threatened of dismissal if they reported the deaths to the news media and were even forbidden to discuss the death among themselves or even to hold a funeral. Patients of the dead physicians were told that their physician had left the hospital.

Jacobsen: Is the lack of reportage on those who care for us in times of need showing a lack of care for them in their times of need across political party lines and media platforms?

Desjardins: Absolutely. The profession is crushed from all sides and getting no sympathy from anyone. The only reason the U.S. healthcare system has not yet collapsed under these circumstances, is because of the endless professional ethic of medical staff members, a resource that seems endless and that is currently massively exploited by the public, by corporate medicine, and by the government.

Jacobsen: Thank you for the opportunity and your time, Benoit.

Desjardins: Thank you for discussing this important topic.

Interview with Sam Vaknin on the Next Era of Invention

Sam Vaknin & Scott Douglas Jacobsen

Sam Vaknin is the author of *Malignant Self-love: Narcissism Revisited* as well as many other books and ebooks about topics in psychology, relationships, philosophy, economics, international affairs, and award-winning short fiction. He is former Visiting Professor of Psychology, Southern Federal University, Rostov-on-Don, Russia and on the faculty of CIAPS (Commonwealth Institute for Advanced and Professional Studies). He is a columnist with *Brussels Morning*, was the Editor-in-Chief of *Global Politician*, and served as a columnist for *Central Europe Review*, *PopMatters*, *eBookWeb*, and *Bellaonline*, and as a *United Press International* (UPI) Senior Business Correspondent. He was the editor of mental health and Central East Europe categories in The Open Directory and *Suite101*. His YouTube channels garnered 80,000,000 views and 405,000 subscribers.

Visit Sam's website: http://www.narcissistic-abuse.com

Scott Douglas Jacobsen: Today's topic: the next era of invention; what is the next era of invention? The previous eras relied upon unusually bright, innovative, and persistent, persons, solo: Legitimate geniuses. We have moved more into a world of invention emphasizing

teamwork and dollars alongside *some* coordination with narrow artificial intelligence or specified algorithms, programs.

Professor Sam Vaknin: Mankind has always alternated between teamwork and the individual genius. I think that we should focus on the raw materials (inputs) and the outputs of innovation rather than on who and how we bring it about.

We are transitioning from the age of monetized attention to the age of reality engineering.

Cities amounted to the first make-belief, virtual reality. Urbanization and population growth led to the rise of the creative genius (auteur), and the emergence of the concept of the original (due to the need to be seen and noticed in the multitude).

Intellectual property followed 300 years ago when mechanical reproduction blurred the line between original and copy and dramatically reduced the marginal cost of copies.

“The Work of Art in the Age of Mechanical Reproduction” (1935), by Walter Benjamin, is an essay of cultural criticism which proposes and explains that mechanical reproduction devalues the aura (uniqueness) of an *objet d’art*.

Since then, identity has become a big business: patents, copyrights, brands, and blockchain NFTs. Distributed ledgers as well as centralised records vouch for one’s identity and guarantee it.

The nonrivalrous zero marginal cost of digital goods has shifted the focus from manufacturing of tangibles to the manipulation of abstract symbols, the commodification of attention, and the emerging conundrum of discoverability.

Both individual creators and commercial enterprises reacted by interpellating potential consumers via propaganda and targeted advertising and by turning a profit via the aggregation of big data (targeting the demographics of attention).

These trends engendered self-sufficient disintermediated atomization – attention has been diverted to asocial online pursuits – and yielded an impaired reality testing (fantasy paracosms, virtual and augmented reality, and, soon, the metaverse).

The next frontiers are reality-like (pseudoreal) “real estate” and commodified but idiosyncratic menu-driven reality (the aforementioned metaverse).

Collaborative virtual realities will supplant physical ones and reality substitutes (sex dolls, intimacy apps) will proliferate. Tech behemoths, such as Facebook, Google, Apple, and Amazon will try to control the way we perceive reality and the immersive universes that we inhabit.

IRL AI will displace people as friends, advisors, interlocutors, lovers, and service providers. Users will construct online simulations and inhabit them. But this turn of events will also force the introduction of mandatory digital identities, hopefully based on blockchain rather than government regulation.

Jacobsen: What marks something as genuinely inventive rather than simply an update to some technology?

Vaknin: Truly innovative inventions profoundly change the way we live, communicate, work, make love, and interact. By this standard, neither the automobile nor the smartphone are veritable innovations: the former is a mere mechanized horse and the latter a derivative of the phone. But Bell's telephone and the telegraph are examples of paradigm-shifting, reality-altering inventions.

Jacobsen: More fundamentally, what is the basic principle of invention, its nature?

Vaknin: Most groundbreaking inventions generate their own markets, fostering needs in consumers that they were unaware of. They also recombine the familiar (e.g., previous technologies) in ways that produce alien, unprecedented, and strange products or services. Finally, true inventions become indispensable in short order: it is hard to imagine a life without them and we pity our predecessors for having been deprived of their existence.
Schumpeter seemed to have captured the unsettling nature of innovation: unpredictable, unknown, unruly, troublesome, and ominous. Innovation often changes the inner dynamics of organizations and their internal power structure. It poses new demands on scarce resources. It provokes resistance and unrest. If mismanaged – it can spell doom rather than boom.

[Editors' Note: https://en.wikiquote.org/wiki/Creative_destruction]

Yet, the truth is that no one knows why people innovate. The process of innovation has never been studied thoroughly – nor are the effects of innovation fully understood.

Jacobsen: What do you see as the most significant biotechnology invention in the history of the biological world?

Vaknin: Possibly CRISPR, the revolutionary gene editing technology. Sometimes, advances in speed and quantity do constitute a quantum leap.

Jacobsen: What has been the most worldview-shattering invention in human history?

Vaknin: The harnessing of fire, the ability to reignite it at will.

Jacobsen: How does the psychology of an inventive person work?

Vaknin: The typical inventor is solutions-oriented. S/he perceives a lack, deficiency, or lacuna and sets out to remedy it. Inventors are also possessed of a synoptic/panoramic view, able to discern the connective tissue that binds apparently disparate phenomena. Finally, a true inventor is able to transition seamlessly from the theoretical to practical, from the drawing board to testing, and thence to prototype.

Creative people are feared and hated, ostracized and punished, unless they are willing to clown themselves or dumb down and conform to the biases, prejudices, and errors of the masses.

High IQ does not translate into success in the absence of perseverance, agreeableness, industriousness, stability (self-regulation), humility, a capacity for teamwork (minimal empathy and respect for others), robust mental health, a social support network, and luck. Many geniuses are homeless or incarcerated and all but forgotten.

[Editors' Note: https://en.wikipedia.org/wiki/Big_Five_personality_traits

O: Openness
C: Conscientiousness
E: Extraversion
A: Agreeableness
N: Neuroticism]

The reality testing of inventors is impaired: they perceive the world differently (possibly a sign of autism). Coupled with recklessness, a sense of fearless godlike immunity, it leads to exploratory behavior.

Originality, novelty, difference: synoptic connectivity appears schizotypal or even psychotic (Schizotypy). Eysenck linked psychoticism to creativity. Indeed, the creative burst is often disorganized initially (inspiration, intuition, dreams). Attention multitasking generates unexpected insights and synergies.

Impatience, grandiosity or contempt and condescension characterize inventors: convinced of their superiority, they tend to block out "noise" and ignore criticism. Lability and dysregulation are sources of inspiration. Proclivity for change, thrill seeking, and risky conduct result in innovation.

[Editors' Note: https://en.wikipedia.org/wiki/Latent_inhibition

Cf. https://www.harvardmagazine.com/2004/05/ideas-rain-in-html]

These are the reasons that most innovators endure inordinate hardships in life, their resilience and perseverance tested to the breaking point.

Jacobsen: With the advent of some software capable of mimicking human capacities more, and performing in superhuman capacities – at least on paper in computational power, how is this changing the interaction of human beings with software to invent in more precise and creative ways?

Vaknin: We tend to mythologize the process of invention, to render it mystical and uniquely human. The truth is that it is an emergent artefact (epiphenomenon), the ineluctable outcome of complexity. At this stage, we are feeding computer models with humongous reams of raw data in the hope that irreducible interactions between the umpteen pieces of information will yield innovative insights and discoveries. The next phase will involve fine-tuning the inputs so as to allow artificial intelligence to work on its own and to seek data as well as outputs autonomously. At that stage, we would still be able to define the research agenda, but not for long.

Jacobsen: In line with Alan Turing's views, who I agree with more than the 'moderns' in Western technology communities when engineered computational systems match our "feeble powers," how will this change the world of invention?

Vaknin: We will be rendered obsolete. We would still maintain a parasitic, atomized, technologically self-sufficient kind of existence for a while, but then, like everything superfluous in Nature, we will wane and fade away. Hence my prediction of a Luddite counter-revolution which would seek to physically demolish or ban certain technologies, maybe justly so.

Jacobsen: How might the style of invention, or even the definition of invention itself, change with the precision and breadth future computation, and simulation, will bring to everything in our lives? Where, there might be the capacity of a constant role of mini-invention increasingly in every facet of human life, similar to the infusion of – what we consider – ordinary technologies now.

Vaknin: The overwhelming vast majority of people are incapable of making use of the full set of features made available even by current technologies, let alone of innovating. I foresee "innovation engineers" whose job would be to cajole artificial intelligence codes and models into new discoveries. But innovation would become the domain of machines, not humans.

Jacobsen: How long until the technological world or the biological world make human beings, as an environmentally engineered (evolved) structure, neither entirely relevant to the business of the Earth nor the dominant conscious information processors on the planet?

Vaknin: I would be surprised if this would take longer than 50 years. With the exception of physical jobs like plumbing, AI would be perfectly capable of replacing and displacing us and doing a better job of it.

Jacobsen: Thank you for the opportunity and your time, and the happy final note, Sam.

Vaknin: You are welcome. Always delighted to spread doom!

Interview with Sam Vaknin on Human-Machine Interface

Sam Vaknin & Scott Douglas Jacobsen

Scott Douglas Jacobsen: When did the first human-machine interactions truly begin in modern history insofar as we take technology now?

Dr. Sam Vaknin: When a man (or a woman) picked up a stone and threw it at a scavenger.

Jacobsen: How have technologies influenced the psycho-social makeup of human beings?

Vaknin: Technology fostered the delusion that every problem has a solution and the hubris that attends upon proving this contention somewhat true. We have learned to internalize technologies and render them our extensions, driving us deeper into fantastic paracosms, replete with populations of internal objects that represent cohorts of external devices and systems. We became dependent on technology and this dependency emerged as our default mode, leading us to prefer machines to other humans.

Jacobsen: These technologies, especially contemporary ones, come out of smart people working hard. How are they, in a way, extensions of ourselves based on those smart people's understanding of some principle and then applying this to ergonomic design?

Vaknin: These "smart people" are not representative of humanity, not even remotely. They are a self-selecting sample of schizoid, mostly white, mostly men. I am not sure why you limited your question to the least important and most neglected aspect of technology: ergonomic design, dictated by the very structure and functioning of the human body. There are other, much more crucial aspects of technology that reflect the specific mental health pathologies, idiosyncrasies, and eccentricities of engineers, coders, and entrepreneurs – rather than any aspect or dimension of being human.

Jacobsen: How are military applications showing this to be the case with drones and the like? Also, the eventual *reductio ad absurdum* of long-term war with all these technology innovations around autonomous war-robots seems increasingly apparent, when, in some hypothetical future, it'd be simply machines fighting machines for some geographic or resource squabble of some leaders.

Vaknin: War is increasingly more democratized (terrorism and asymmetrical warfare, anyone?). It is also more remote controlled. But its main aim is still to kill people, combatants and civilians alike. Machines will never merely fight only other contraptions. War will never be reduced to a mechanized version of chess. Men, women, and children will always die in battle as conflict

becomes ever more total. The repossession of resources requires the unmitigated annihilation of their erstwhile owners.

Jacobsen: Are autocratic, theocratic, or democratic, societies, utilizing the technologies 'interfacing' with human beings more wisely – which one?

Vaknin: Wisdom is in the eye of the beholder. There is no difference in the efficacy of deploying technologies between various societal organizational forms. All governments and collectives – autocratic, democratic, and theocratic, even ochlocratic or anarchic – leverage technology to secure and protect the regime and to buttress the narratives that motivate people to fight, work, consume, and mate.

Jacobsen: I interviewed another smart guy, Dr. Evangelos Katsioulis, years ago. He, at that time – maybe now too, believed no limit existed to the integration between machines and humans. When will human mechanics be understood sufficiently to when, as with the Ship of Theseus, human beings can function as human beings with 10%, 25%, 75% non-biological machine parts comprising their localized subjectivity and locomotion?

[Editors' Note: https://plato.stanford.edu/entries/identity-time/#4]

Vaknin: Much sooner than we think. But there will always be a Resistance: a substantial portion of the population who will remain averse to cyborg integration and as the Luddites of yesteryear will seek to forbid such chimeras and destroy them.

In some rudimentary ways, we are already integrated with machines. Can you imagine your life without your devices?

Jacobsen: How are interactions with technologies more intimately blurring the sense of self?

Vaknin: Human brains are ill-equipped to tell the difference between reality and mimicry, simulation, or fantasy. Technologies are the reifications of the latter at the expense of the former.

One of the crucial aspects of the putative "Self" or "Ego' is reality testing. As the boundaries blur, so will our selves. We are likely to acquire a hive mind, melded with all the technologies that surround us, seamlessly slipping in and out of dream states and metaverses. The "Self' will become the functional equivalent of our attire: changeable, disposable, replaceable.

As it is, I am an opponent of the counterfactual idea of the existence of some kernel, immutable core identity, self, or ego – see this video about IPAM, my Intrapsychic Activation Model.

Jacobsen: How are the plurality of software and hardware available vastly outstripping the capacity for ordinary people to use them all, let alone understand them? Most seem drawn merely to video games, television, cell phones, and some social media platforms. That's about it. There's so, so much more around now.

Vaknin: There have always been technologies for the masses as well as for niche users. Where we broke off with the past is in multitasking, the simultaneous suboptimal use of multiple devices.

Jacobsen: What is the ultimate point of human-machine 'interfaces'? We 'birthed' electronic machines and information processing. What will be birthed from this union of biological mechanisms and alloyed assistants, playthings?

Vaknin: As they get more integrated by the day, the point is to empower, enhance, and expand both symbiotic partners: humans and machines alike. It is a virtuous cycle which will lead to functional specialization with both parties focused on what they do best.

Still, if humans fail to bake Asimov-like rules into their automata, the potential for conflict is there, as artificial intelligence becomes more sentient and intelligent and prone to passing the Turing Test with flying colors. In short: indistinguishable from us, except with regards to its considerably more potent processing prowess.

Popular culture reflected this uncanny valley: the growing unease with android robots, first postulated by Masahiro Mori, the Japanese roboticist, in 1970.

The movie *I, Robot* is a muddled affair. It relies on shoddy pseudo-science and a general sense of unease that artificial (non-carbon based) intelligent lifeforms seem to provoke in us. But it goes no deeper than a comic book treatment of the important themes that it broaches. *I, Robot* is just another – and relatively inferior – entry in a long line of far better movies, such as *Blade Runner* and *Artificial Intelligence*.

Sigmund Freud said that we have an uncanny reaction to the inanimate. This is probably because we know that – pretensions and layers of philosophizing aside – we are nothing but recursive, self-aware, introspective, conscious machines. Special machines, no doubt, but machines all the same.

[Editors' Note: https://web.mit.edu/allanmc/www/freud1.pdf

Cf. https://www.sas.upenn.edu/~cavitch/pdf-library/Freud_Uncanny.pdf]

Consider the James Bond movies. They constitute a decades-spanning gallery of human paranoia. Villains change: communists, neo-Nazis, media moguls. But one kind of villain is a fixture in this psychodrama, in this parade of human phobias: the machine. James Bond always finds himself confronted with hideous, vicious, malicious machines and automata.

It was precisely to counter this wave of unease, even terror, irrational but all-pervasive, that Isaac Asimov, the late Sci-fi writer (and scientist) invented the Three Laws of Robotics:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings, except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Many have noticed the lack of consistency and, therefore, the inapplicability of these laws when considered together.

First, they are not derived from any coherent worldview or background. To be properly implemented and to avoid their interpretation in a potentially dangerous manner, the robots in which they are embedded must be equipped with reasonably comprehensive models of the physical universe and of human society.

Without such contexts, these laws soon lead to intractable paradoxes (experienced as a nervous breakdown by one of Asimov's robots). Conflicts are ruinous in automata based on recursive functions (Turing machines), as all robots are. Gödel pointed at one such self-destructive paradox in the *Principia Mathematica*, ostensibly a comprehensive and self-consistent logical system. It was enough to discredit the whole magnificent edifice constructed by Russel and Whitehead over a decade.

Some argue against this and say that robots need not be automata in the classical, Church-Turing, sense. That they could act according to heuristic, probabilistic rules of decision making. There are many other types of functions (non-recursive) that can be incorporated in a robot, they remind us.

True, but then, how can one guarantee that the robot's behavior is fully predictable? How can one be certain that robots will fully and always implement the three laws? Only recursive systems are predictable in principle, though, at times, their complexity makes it impossible.

An immediate question springs to mind: HOW will a robot identify a human being? Surely, in a future of perfect androids, constructed of organic materials, no superficial, outer scanning will suffice. Structure and composition will not be sufficient differentiating factors.

There are two ways to settle this very practical issue: one is to endow the robot with the ability to conduct a Converse Turing Test (to separate humans from other life forms) – the other is to somehow "barcode" all the robots by implanting some remotely readable signaling device inside them (such as a RFID – Radio Frequency ID chip). Both present additional difficulties.

The second solution will prevent the robot from positively identifying humans. He will be able identify with any certainty robots and only robots (or humans with such implants). This is ignoring, for discussion's sake, defects in manufacturing or loss of the implanted identification tags. And what if a robot were to get rid of its tag? Will this also be classified as a "defect in manufacturing"?

In any case, robots will be forced to make a binary choice. They will be compelled to classify one type of physical entities as robots – and all the others as "non-robots". Will non-robots include monkeys and parrots? Yes, unless the manufacturers equip the robots with digital or optical or molecular representations of the human figure (masculine and feminine) in varying positions (standing, sitting, lying down). Or unless all humans are somehow tagged from birth.

These are cumbersome and repulsive solutions and not very effective ones. No dictionary of human forms and positions is likely to be complete. There will always be the odd physical posture which the robot would find impossible to match to its library. A human disk thrower or swimmer may easily be classified as "non-human" by a robot – and so might amputated invalids.

What about administering a converse Turing Test?

This is even more seriously flawed. It is possible to design a test, which robots will apply to distinguish artificial life forms from humans. But it will have to be non-intrusive and not involve overt and prolonged communication. The alternative is a protracted teletype session, with the human concealed behind a curtain, after which the robot will issue its verdict: the respondent is a human or a robot. This is unthinkable.

Moreover, the application of such a test will "humanize" the robot in many important respects. Humans identify other humans because they are human, too. This is called empathy. A robot will have to be somewhat human to recognize another human being, it takes one to know one, the saying (rightly) goes.

Let us assume that by some miraculous way the problem is overcome and robots unfailingly identify humans. The next question pertains to the notion of "injury" (still in the First Law). Is it limited only to physical injury (the elimination of the physical continuity of human tissues or of the normal functioning of the human body)?

Should "injury" in the First Law encompass the no-less-serious mental, verbal and social injuries (after all, they are all known to have physical side effects which are, at times, no less severe than direct physical "injuries")? Is an insult an "injury"? What about being grossly impolite, or psychologically abusive? Or offending religious sensitivities, being politically incorrect – are these injuries? The bulk of human (and, therefore, inhuman) actions actually offend one human being or another, have the potential to do so, or seem to be doing so.

Consider surgery, driving a car, or investing money in the stock exchange. These “innocuous” acts may end in a coma, an accident, or ruinous financial losses, respectively. Should a robot refuse to obey human instructions which may result in injury to the instruction-givers?

Consider a mountain climber – should a robot refuse to hand him his equipment lest he fall off a cliff in an unsuccessful bid to reach the peak? Should a robot refuse to obey human commands pertaining to the crossing of busy roads or to driving (dangerous) sports cars?

Which level of risk should trigger robotic refusal and even prophylactic intervention? At which stage of the interactive man-machine collaboration should it be activated? Should a robot refuse to fetch a ladder or a rope to someone who intends to commit suicide by hanging himself (that’s an easy one)?

Should he ignore an instruction to push his master off a cliff (definitely), help him climb the cliff (less assuredly so), drive him to the cliff (maybe so), help him get into his car in order to drive him to the cliff… Where do the responsibility and obeisance bucks stop?

Whatever the answer, one thing is clear: such a robot must be equipped with more than a rudimentary sense of judgment, with the ability to appraise and analyse complex situations, to predict the future and to base his decisions on very fuzzy algorithms (no programmer can foresee all possible circumstances). To me, such a “robot” sounds much more dangerous (and humanoid) than any recursive automaton which does NOT include the famous Three Laws.

Moreover, what, exactly, constitutes “inaction”? How can we set apart inaction from failed action or, worse, from an action which failed by design, intentionally? If a human is in danger and the robot tries to save him and fails – how could we determine to what extent it exerted itself and did everything it could?

How much of the responsibility for a robot’s inaction or partial action or failed action should be imputed to the manufacturer – and how much to the robot itself? When a robot decides finally to ignore its own programming – how are we to gain information regarding this momentous event? Outside appearances can hardly be expected to help us distinguish a rebellious robot from a lackadaisical one.

The situation gets much more complicated when we consider states of conflict.

Imagine that a robot is obliged to harm one human in order to prevent him from hurting another. The Laws are absolutely inadequate in this case. The robot should either establish an empirical hierarchy of injuries – or an empirical hierarchy of humans. Should we, as humans, rely on robots or on their manufacturers (however wise, moral and compassionate) to make this selection for us? Should we abide by their judgment which injury is the more serious and warrants an intervention?

A summary of the Asimov Laws would give us the following “truth table”:

A robot must obey human commands except if:

1. Obeying them is likely to cause injury to a human, or

2. Obeying them will let a human be injured.

A robot must protect its own existence with three exceptions:

1. That such self-protection is injurious to a human;

2. That such self-protection entails inaction in the face of potential injury to a human;

3. That such self-protection results in robot insubordination (failing to obey human instructions).

Trying to create a truth table based on these conditions is the best way to demonstrate the problematic nature of Asimov’s idealized yet highly impractical world.

Here is an exercise:

Imagine a situation (consider the example below or one you make up) and then create a truth table based on the above five conditions. In such a truth table, “T” would stand for “compliance” and “F” for non-compliance.

Example:

A radioactivity monitoring robot malfunctions. If it self-destructs, its human operator might be injured. If it does not, its malfunction will equally seriously injure a patient dependent on his performance.

One of the possible solutions is, of course, to introduce gradations, a probability calculus, or a utility calculus. As they are phrased by Asimov, the rules and conditions are of a threshold, yes or no, take it or leave it nature. But if robots were to be instructed to maximize overall utility, many borderline cases would be resolved.

Still, even the introduction of heuristics, probability, and utility does not help us resolve the dilemma in the example above. Life is about inventing new rules on the fly, as we go, and as we encounter new challenges in a kaleidoscopically metamorphosing world. Robots with rigid instruction sets are ill suited to cope with that.

At the risk of going abstruse, two comments:

1. Godel's Theorems

The work of an important, though eccentric, Czech-Austrian mathematical logician, Kurt Gödel (1906-1978) dealt with the completeness and consistency of logical systems. A passing acquaintance with his two theorems would have saved the architect a lot of time.

Gödel's First Incompleteness Theorem states that every consistent axiomatic logical system, sufficient to express arithmetic, contains true but unprovable ("not decidable") sentences. In certain cases (when the system is omega-consistent), both said sentences and their negation are unprovable. The system is consistent and true – but not "complete" because not all its sentences can be decided as true or false by either being proved or by being refuted.

The Second Incompleteness Theorem is even more earth-shattering. It says that no consistent formal logical system can prove its own consistency. The system may be complete – but then we are unable to show, using its axioms and inference laws, that it is consistent.

In other words, a computational system can either be complete and inconsistent – or consistent and incomplete. By trying to construct a system both complete and consistent, a robotics engineer would run afoul of Gödel's theorem.

2. Turing Machines

In 1936 an American (Alonzo Church) and a Briton (Alan M. Turing) published independently (as is often the case in science) the basics of a new branch in Mathematics (and logic): computability or recursive functions (later to be developed into Automata Theory).

The authors confined themselves to dealing with computations which involved "effective" or "mechanical" methods for finding results (which could also be expressed as solutions (values) to formulae). These methods were so called because they could, in principle, be performed by simple machines (or human-computers or human-calculators, to use Turing's unfortunate phrases). The emphasis was on finiteness: a finite number of instructions, a finite number of symbols in each instruction, a finite number of steps to the result. This is why these methods were usable by humans without the aid of an apparatus (with the exception of pencil and paper as memory aids). Moreover: no insight or ingenuity were allowed to "interfere" or to be part of the solution seeking process.

What Church and Turing did was to construct a set of all the functions whose values could be obtained by applying effective or mechanical calculation methods. Turing went further down Church's road and designed the "Turing Machine" – a machine which can calculate the values of all the functions whose values can be found using effective or mechanical methods. Thus, the program running the TM (=Turing Machine in the rest of this text) was really an effective or mechanical method. For the initiated readers: Church solved the decision-problem for

propositional calculus and Turing proved that there is no solution to the decision problem relating to the predicate calculus. Put more simply, it is possible to “prove” the truth value (or the theorem status) of an expression in the propositional calculus – but not in the predicate calculus. Later it was shown that many functions (even in number theory itself) were not recursive, meaning that they could not be solved by a Turing Machine.

No one succeeded to prove that a function must be recursive in order to be effectively calculable. This is (as Post noted) a “working hypothesis” supported by overwhelming evidence. We don’t know of any effectively calculable function which is not recursive, by designing new TMs from existing ones we can obtain new effectively calculable functions from existing ones and TM computability stars in every attempt to understand effective calculability (or these attempts are reducible or equivalent to TM computable functions).

The Turing Machine itself, though abstract, has many “real world” features. It is a blueprint for a computing device with one “ideal” exception: its unbounded memory (the tape is infinite). Despite its hardware appearance (a read/write head which scans a two-dimensional tape inscribed with ones and zeroes, etc.) – it is really a software application, in today’s terminology. It carries out instructions, reads and writes, counts and so on. It is an automaton designed to implement an effective or mechanical method of solving functions (determining the truth value of propositions). If the transition from input to output is deterministic, we have a classical automaton – if it is determined by a table of probabilities – we have a probabilistic automaton.

With time and hype, the limitations of TMs were forgotten. No one can say that the Mind is a TM because no one can prove that it is engaged in solving only recursive functions. We can say that TMs can do whatever digital computers are doing – but not that digital computers are TMs by definition. Maybe they are – maybe they are not. We do not know enough about them and about their future.

Moreover, the demand that recursive functions be computable by an UNAIDED human seems to restrict possible equivalents. Inasmuch as computers emulate human computation (Turing did believe so when he helped construct the ACE, at the time the fastest computer in the world) – they are TMs. Functions whose values are calculated by AIDED humans with the contribution of a computer are still recursive. It is when humans are aided by other kinds of instruments that we have a problem. If we use measuring devices to determine the values of a function it does not seem to conform to the definition of a recursive function. So, we can generalize and say that functions whose values are calculated by an AIDED human could be recursive, depending on the apparatus used and on the lack of ingenuity or insight (the latter being, anyhow, a weak, non-rigorous requirement which cannot be formalized).

Jacobsen: Thank you for the opportunity and your time, Sam.

Vaknin: Thank you as ever, Scott.

Interview with Sam Vaknin on Legal Derivatives of Invented Technologies

Sam Vaknin & Scott Douglas Jacobsen

Scott Douglas Jacobsen: Technologies integrated with human cultures continually make new laws, even creating entirely new frames of legal discourse. What have been some of the more disruptive forms of technology to legal systems, philosophies of law?

Dr. Sam Vaknin: Every technology necessitated a revision of existing laws to incorporate its unique features. The more disruptive the technology, the more profound the legal revisions: the printing press, for example, or the telegraph, telephone, automobile, Internet, social media, smartphone, and so on.

Jacobsen: What role does invention play in the creation of new laws, policies, even whole new legal systems of consideration in governance?

Vaknin: I dispute this claim or premise. Technology does not spur legal innovations or revolutions. Consider crime: contemporary technologies simply allow us to commit age-old offenses in new ways.

New technologies do force laws and regulations to become a lot more detailed and specific in order to accommodate their idiosyncrasies, but there is no paradigmatic shift involved.

Jacobsen: We talked about human-machine interfaces. What is the past of law regarding human use of technologies?

Vaknin: Laws, past and present, have dealt mostly with the adverse outcomes, actual and potential, of using technology. As technologies became more sophisticated, though, their unintended consequences became less predictable and the Law had to play catchup and whack-a-mole with those.

Jacobsen: Of modern communications technologies, what have required the most ubiquitous change in law?

Vaknin: The telegraph and the radio were the most disruptive technologies with the Internet a close third. The abolition of distance by the first two and the egalitarianism fostered by the latter served to undermine many erstwhile legal tenets and conceptual pillars.

Jacobsen: With narrow AI in many facets of life, quietly, and more obviously such as LLMs, what are some necessary changes to law for protection of copyright and plagiarism? Linguist Noam Chomsky is reported to have said, “Let’s stop calling it ‘Artificial Intelligence’ and call it what it is: ‘plagiarism software.’ Don’t create anything, copy existing works from existing artists and alter it sufficiently to escape copyright laws. It’s the largest theft of property ever since Native American lands by European settlers.” You had him in your list of geniuses. What will be the outcome of the theft of intellectual property to create some of these algorithms?

Vaknin: I completely disagree with this way of looking at things. I don’t see even a hint of these legal issues or ostensible transgressions with large language models. AI generates derivative works based on databases of texts, but does not reprint or replicate these texts verbatim. It learns from texts but does not plagiarize them in the strict legal sense (except in rare cases).

There is definitely an ethical conundrum here, but not a legal one. Still, this ethical dilemma arises also with cliff notes or Blinkist or parodies or any creative work inspired by another. Chomsky’s own work relies on the oeuvre of previous scholars!

Jacobsen: What will be the future of the discourse between increasing intimate contact, even fusion, with synthetic systems and the law? When digital conscious systems become more fully decoupled from human control – degrees of autonomy, what will this mean for both the concept of personhood and the idea, not only human rights but, rights attributed to agents more broadly?

Vaknin: At some point, we would need to generalize the language of the Law to apply it equally to all forms of intelligences with agency, including cyborgs, androids, and artificial intelligence. Sentience, not carbon content, would become the test of applicability of laws, norms, rules, and regulations.

Who would enforce these carbon-blind laws would become a major point of contention. We are having a hard time coping with driverless cars. How well would we adapt to non-human cops and judges?

Jacobsen: Thank you for the opportunity and your time, Sam.

Vaknin: Thank you as ever, Scott.

Western Conceptions of Infinity and Eternity, Temporal and Atemporal

Ken Shea

'We burn with desire to find solid ground and an ultimate sure foundation whereon to build a tower reaching to the Infinite. But our whole groundwork cracks and the earth opens to abysses.' -Blaise Pascal

'The problem is, when we try to calculate all the way down to zero distance, the equation blows up in our face and gives us meaningless answers - things like infinity. This caused a lot of trouble when the theory of quantum electrodynamics first came out. People were getting infinity for every problem they tried to calculate!' -Richard Feynman

'There must have been a time in which the world did not exist, that is, a void time. But in a void time, the origination of a thing is impossible; because no part of any such time contains a distinctive condition of being in preference of that of non-being.' -Immanuel Kant

The following qualities are typically associated with time: duration, succession, impermanence, change, movement, immanence, becoming. In this light, the juxtaposition of being and time, 'As eternity is the proper measure of being, so time is the proper measure of movement', by the theologian Thomas Aquinas is more understandable. Time, consciousness, and being appear to be indivisible on a phenomenological level, patterned after a flow. While Plato dichotomized the eternal 'world of immutable being' against the evanescent 'world of generation' and change, Thomas Aquinas gives a beautiful rendering of Plato's notion, expounded in the *Timaeus*, of time presenting as the *moving image* of eternity; Aquinas: 'The *now* that stands still is said to make eternity according to our apprehension. For just as the apprehension of time is caused in us by the fact that we apprehend the flow of the *now*, so the apprehension of eternity is caused in us by our apprehending the *now* standing still.' Mathematical truths are said to be outside of time or eternal because they are above change, immutable, and in that way, *pace* Plato, transcendent. 'Time and change make no difference to the truth of *two plus two equals four*', the American encyclopedist Mortimer Adler astutely notes. How might eternity and infinity relate?

'The notion of infinity involves greater perplexities that than of eternity. The meaning of eternity is weighted with the mystery of God, the world, and time. All these affect the conception of infinity; but for the infinite there are also the mysteries of number and of space, of matter and motion. In the sphere of quantity, or of things subject to quantity, infinity is itself the source of mystery, or at least the root of difficulty in analysis. It is the central term in the discussion of the continuous and the indivisible, the nature of series and of limits.' -Mortimer Adler

What Georg Hegel reckoned a 'bad infinity' or 'spurious infinity' of attempting to overcome finiteness by, essentially, piling on more finiteness (n+1 recursively applied indefinitely) would be analogous in the realm of temporality to sempiternity, or an interminable duration. The word

'duration' is the giveaway that sempiternity does not transcend time (sempiternity presupposes time, after all), like Hegel's 'bad infinity' presupposes finiteness and thereby crashes with Icarus under the weight of its own contradictions. The popular connotation of 'eternity' as infinite duration or endless time - differing from the way Plato, Plotinus, Thomas Aquinas, and Baruch Spinoza use the term 'eternal', viz., to mean timelessness or a fullness of Being, lacking no attributes, infinite substance - receives full-throated expression by Ivan Karamazov in Fyodor Dostoevsky's masterpiece, *The Brothers Karamazov*. Ivan Karamazov attempts to convey the enormity of 'a quadrillion of a quadrillion raised to the quadrillionth power' years and winds up with something closer to sempiternity by presupposing duration, albeit apparently interminable. The term 'eternity' finds employment colloquially in the distorted sempiternity sense. Macbeth's famous soliloquy captures a despairing sense of sempiternity, as opposed to eternity.

'Tomorrow, and tomorrow, and tomorrow,

Creeps in this petty pace from day to day,

To the last syllable of recorded time'

-Macbeth (*Macbeth*, Act V, Scene V)

The so-called 'block universe' theory in physics likewise would appear thoroughly mired in temporality, though Lee Smolin et al. have contended, e.g., see the quote below, that the block universe is 'fundamentally timeless.' That is not true. If you read carefully what Lee Smolin is attempting to convey, the overall picture speaks to a fractalized or spatialized time, which would nonetheless fall under the rubric of temporality however radically the 'block universe' innovated the works of Galileo and Descartes. What Lee Smolin phrases 'the expulsion of time implied by Einstein's theory of special relativity' is nothing more than relativizing time according to a frame of reference, keeping temporality, but undercutting the notion of linear time. In the context of Plato's *Timaeus*, one scholar found, 'A sempiternal object is one which exists at all moments of time', which could almost be the definition of the sempiternal, not eternal, block universe.

'The picture of the history of the universe, taken as one, as a system of events connected by causal relations, is called the *block universe*. The reason for that perhaps peculiar name is that it suggests that what is real is the whole history at once - the allusion is to a block of stone, from which something solid and unchanging can be carved.

The block universe is the culmination of the movement begun by Galileo and Descartes to treat time as if it were another dimension of space. It gives a description of the whole history of the universe as a mathematical object, which, as we noted in Chapter 1, is timeless. If you believe that it corresponds to what is objectively real in nature, you're asserting that the universe is fundamentally timeless. This block universe picture is the second step in the expulsion of time implied by Einstein's theory of special relativity.

The block universe marries time and space.' -Lee Smolin (*Time Reborn*)

So much for sempiternity. The paradox of infinity is that any attempt to precisely define infinity or an infinite quantity would make it finite in some way (see non-denumerable sets below)!

Obviously a finite mind would only be capable of entertaining finite ideas - a point empiricist philosophers John Locke, David Hume, and George Berkeley are at pains to stress. Thomas Hobbes got right to business by saying, 'Whatsoever we imagine is *finite*', presumably by virtue of being refracted through the prism of a finite mind. Hobbes grimly elaborated: 'When we say anything is infinite, we signify only that we are not able to conceive the ends and bounds of the thing named, having no conception of the thing, but of our own inability.' In the paradoxical sense of 'nonsense' which Ludwig Wittgenstein explores in the *Tractatus*, logicians like Aristotle have regarded vague or negative terms as 'infinite' because of their indefinite or indeterminate nature. The nonsense term 'not-turtle' would be apprehended by some logicians as infinite because the term refers, if that's the right word, to every facet of the multiverse save turtles. The term 'not-turtle' is a kind of booby trap or maddening origami masquerading as a negation but, once activated, opens out into an unfathomable vastness devoid of graspable meaning.

In a meditation which threatens to encompass the discrete versus continuous nature of reality, George Berkeley disclaims the possibility of infinite division: 'If I cannot perceive innumerable parts in any infinite extension, it is certain that they are not contained in it: but it is evident, that I cannot distinguish innumerable parts in any particular line, surface, or solid, which I either perceive by sense, or figure to myself in my mind; wherefore I conclude that they are not contained in it.' (Notwithstanding physical or perceptual limitations and physical theory anomalies like space-time singularities, the British mathematician Bertrand Russell says in *Introduction to Mathematical Philosophy*: 'We can prove that however nearly equal two fractions may be, there are always other fractions between them', thus revealing a kind of infinity).The French mathematician Blaise Pascal tended to err on the side of continuousness, arguably tumbling over into Hegel's bad or spurious infinity. Pascal intones, 'In sizes or numbers, nature has set before man two marvelous infinities', and illustrates by specifying, 'If we can multiply a number up to 100,000 times, say, we can also take a hundred thousands part of it by dividing it by the same number we multiply it with, and thus every term of increase will become a term of division by changing the integer into a fraction. So that infinite increase includes necessarily infinite division.' Mortimer Adler registers a few concerns with this way of thinking:

'An endless addition produces the infinitely large, so endless division produces the infinitesimal or the infinitely small. A trillion trillion is a finite number, because the addition of a single unit creates a larger number [cf. Hegel's 'n+1' bad infinity]. The fact that the addition of another unit produces a different number indicates that a trillion trillion has a determinate size, which is the same as saying that it is a finite number. An infinite number cannot be increased by addition, for it is constituted - in thought at least - as a number larger than the sum of any two finite numbers; which is another way of saying that it is approached by carrying on the process of addition endlessly. The size of an infinite number is therefore indeterminate', cf. transfinite cardinal numbers, continuum hypothesis.

The mathematical physicist, and 2020 recipient of the Nobel Prize in Physics, Roger Penrose, touches on Georg Cantor's theory of infinite numbers, anticipated centuries earlier by the astronomer and mathematician Galileo Galilei. Penrose: 'One might think, at first, that the number of integers is already greater than the number of natural numbers; since every natural number is an integer whereas some integers (namely the negative ones) are not natural

numbers, and similarly one might think that the number of fractions is greater than the number of integers. However, this is not the case. According to the powerful and beautiful theory of infinite numbers put forward in the late 1800s by the highly original by the Russian-German mathematician Georg Cantor, the total number of fractions, the total number of integers and the total number of natural numbers are all the *same* infinite number' (*The Emperor's New Mind*), denoted by the 'aleph nought' $\aleph_0$ (a.k.a., aleph-zero, aleph-null) symbol, cf. Zermelo–Fraenkel set theory's axiom of infinity, Dedekind's infinite 'system'. The set of real numbers is considered an infinite set and non-denumerable because of its inability to be put in a one-to-one correspondence with the set of natural numbers. Taking a step back, the philosopher of science Patrick Suppes probably gives a fair, albeit amusing, summary of a quixotic history: 'The attempt to prove the existence of an infinite set of objects has a rather bizarre and sometimes tortured history', singling out Richard Dedekind's idiosyncratic work.

'So, naturalists observe, a flea

Hath smaller fleas that on him prey,

And these have smaller yet to bite 'em,

And so proceed ad infinitum'

-Jonathan Swift

'The basic principle behind [Aristotle's] potential infinity is that there are systems, such as the natural numbers, 1, 2, 3 …. , that have no upper bound, so that nothing prevents them from growing even larger. Yet infinity itself is clearly not a number, or anything like it. Infinity, for the adherents of these views, is something one always approaches but never reaches - a sort of open-ended arrangement.'

-Paul Davies (*The Edge of Infinity*)

The German phenomenalist thinker Immanuel Kant actually said, 'an absolute limit is impossible in experience.' The German mathematician and theoretical physicist Hermann Weyl appeared to second Kant's skepticism and speak to Aristotle's notion of 'potential infinity' by finding infinity 'remains forever in the status of creation', a sort of dimly-grokked vanishing point emanating, perhaps, out of David Bohm's implicate order or Śūnyatā in the East. What about experience makes an absolute limit impossible, though? Consciousness and the sense of time deriving from consciousness are phenomenologically continuous. Words like ‘stream’ and ‘flow’ are frequently used to refer to consciousness and the perception of time. William James, for instance, found in *Psychology: Briefer Course* that, 'Within each personal consciousness, thought is sensibly continuous. I can only define "continuous" as that which is without breach, crack, or division.’ Personal consciousness should probably be understood as ‘phenomenal awareness’, as James uses the term in the above context. William James, furthermore, characterizes time, in a chapter titled 'The Sense of Time', an 'altogether ideal abstraction', insofar as the 'now' or 'present' moment is ungraspable as such (*Le moment où je parle est déjà loin de moi*), let alone indivisible, on the level of phenomenology; James provocatively elaborates on the claim that time is an 'altogether ideal abstraction' by rightly suggesting time

partitioned as an individuated modality would be unthinkable or peculiar to contemplate as an isolated aspect of consciousness to 'those unaccustomed to philosophic meditation.'

'And likewise time cannot itself exist,

But from the flight of things we get a sense of time ….

No man, we must confess, feels time itself.

But only knows of time from flight or rest of things.'

-Lucretius (*De Rerum Natura*)

Scientists, logicians, and philosophers have attempted to bridge the 'inductive leap' (i.e., the inference from specific instances to more general conclusions, cf. problem of induction) and tame infinite regressions (e.g., with respect to dis/confirming cases and counter/examples) by assigning boundaries and probabilities for happenings within those boundaries. The end result of forming general conclusions from specific instances reined in by boundaries and probabilities would be some kind of model of reality, or a part of reality.

Determining the basics of ‘reality’ according to a particular conceptual scheme proves more controversial than one might initially reckon. On page 162 of *The Construction of Social Reality*, the philosopher John Searle showcases three non-interpenetrating circles [labeled A, B, C] and asks how many 'objects' reside in the 'miniworld' thus composed. 'Well, according to Carnap's system of arithmetic (and according to common sense), there are three; but according to Leśniewski and other Polish logicians, there are seven objects in this world, counted as follows:

1 = A

2 = B

3 = C

4 = A + B

5 = A + C

6 = B + C

7 = A + B + C

So how many objects are there really in the imagined world? Are there really three or really seven [or some other number]? There is no absolute answer to these questions. The only answers we can give are relative to the arbitrary choice of conceptual schemes. The same sentence, e.g., "There are exactly three objects in the world," will be true in one scheme, false in the other. The heart of the argument is that external realism leads to inconsistencies because it allows for inconsistent descriptions of the supposedly independently existing reality' (Searle, *The Construction of Social Reality*, pages 161-162).

The so-called ‘immaterialist’ philosopher George Berkeley interrogated the questing-begging assumption of a mind-independent reality, regarded by John Searle as unproblematic. In the above thought experiment of set theory revolving around the three circles, subscribers to David Bohm’s implicate order or Eastern philosophy’s notion of Śūnyatā might well contend that the answer would be one, seven, and zero in that all of the apparent multiplicity is really composed of one ‘stuff’ and being and non-being deserve equal due as manifest reality emanates from the implicate order or The Void countless times per second to lend the illusion of linear time, self, objects, motion, forward causality, and a stable world. The ‘empty’ white background ‘holding’ the three circles, making them distinguishable, might warrant analysis, as well.

‘Here among the disappearing, in the realm of the transient,

be a ringing glass that shatters as it rings.’

-Rainer Maria Rilke

‘If we conceive of some point of time which cannot be divided into even the minutest parts of moments, that is the only point that can be called the present’ -Augustine

'In the whole field of learning, philosophy is distinguished from other disciplines - from history, the sciences, and mathematics - by its concern with the problem of being. It also asks about the nature of existence, the modes and properties of being, the difference between being and becoming, appearance and reality, the possible and the actual, being and non-being.' -Mortimer Adler

But as Ken Wilber reminds readers, any model of reality is artificial, if not totally arbitrary, because Reality does not subscribe to or suffer from the bounds of the model, e.g., at the very least, in assuming a subject-object duality. 'Our ordinary conception of the world as a complex of things extended in space and succeeding one another in time is only a conventional map of the universe - it is not real [cf. quantum theory's reversed-time causality]. It is not real because this picture painted by symbolic-map knowledge depends upon the splitting of the universe into separate things seen in space-time, on the one hand, and the seer of these things on the other. In order for this to occur, the universe necessarily has to split itself into observer and observed' (Wilber, *The Spectrum of Consciousness*). The notion of 'a complex of things' or 'separate things' (Wilber’s terms) or the blinders of a Kuhnian disciplinary matrix would be ridiculed by Baruch Spinoza thus: 'A plurality of substances possessing the same nature is absurd.' With both exoteric religion (transcendence) and science (immanence) having demonstrably failed to construct a coherent picture of reality, how about taking Spinoza's non-dual ontology for a spin?

'All things from eternity are of like forms and come round in a circle.' -Marcus Aurelius

'The great thing is to apprehend in the show of the temporal and the transient the substance which is immanent and the eternal which is present.' -Georg Hegel

‘From a wild weird clime that lieth, sublime,

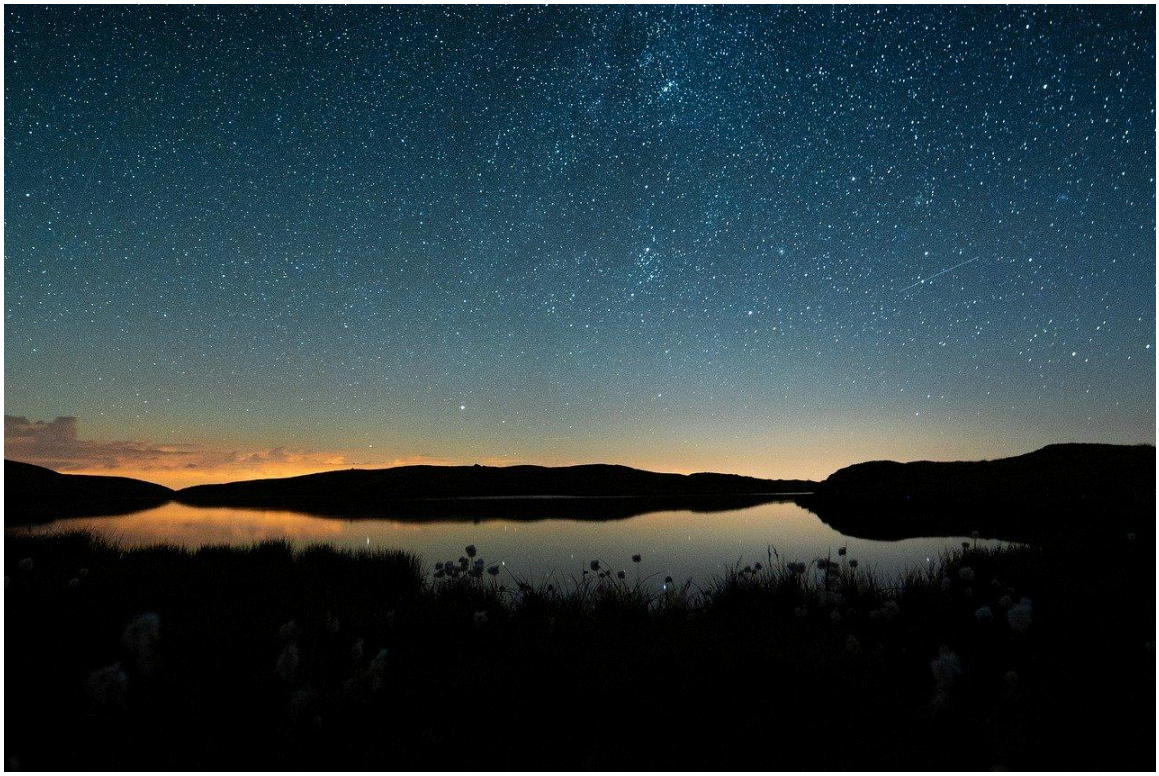
Out of SPACE - Out of TIME.’

-Edgar Allan Poe (excerpted from 'Dream-Land')

Truth-24 Vax

Zero-point energy
Faster-than-light
Flocks of geese;
Antigravitic
Conspiracy Theories,
“Safe and Effective.”

May-Tzu



'The sky was tense with stars, but not so tense as he was, in his breast. Everything overhead was in equilibrium, kept in place by mutual tensions. What was it that *his* tensions kept in place?

And what he saw with his eyes was not even the real heavens. No, only white marks, bright vibrations, clouds of sky roe, tokens of the real thing, only as much as could be taken in through the distortions of the atmosphere. Through these distortions you saw objects, forms, partial realities. The rest was to be felt. And it wasn't only that you felt, but that you were drawn to feel and to penetrate further, as if you were being informed that what was spread over you had to do with your existence, down to the very blood and the crystal forms inside your bones. Rocks, trees, animals, men and women, these also drew you to penetrate further, under the distortions (comparable to the atmospheric ones, shadows within shadows), to find their real being with your own. This was the sense in which you were drawn.'

-Saul Bellow (excerpted from *The Dean's December*)