

Book Review: Age of Em

Posted on May 28, 2016 by Scott Alexander



Note: *I really liked this book and if I criticize it that's not meant as an attack but just as what I do with interesting ideas. Note that Robin has offered to debate me about some of this and I've said no – mostly because I hate real-time debates and have bad computer hardware – but you may still want to take this into account when considering our relative positions. Mild **content warning** for murder, rape, and existential horror. Errors in Part III are probably my own, not the book's.*

I

There are some people who are destined to become adjectives. Pick up a David Hume book you've never read before and it's easy to recognize the ideas and style as Humean. Everything Tolkien wrote is Tolkienesque in a non-tautological sense. This isn't meant to denounce either writer as boring. Quite the opposite. They produced a range of brilliant and diverse ideas. But there was a hard-to-define and very consistent ethos at the foundation of both. Both authors were *very much like themselves*.

Robin Hanson is more like himself than anybody else I know. He's obviously brilliant – a PhD in economics, a masters in physics, work for DARPA, Lockheed, NASA, George Mason, and the Future of Humanity Institute. But his greatest aptitude is in being really, really [Hansonian](#). Bryan Caplan describes it as well as anybody:

When the typical economist tells me about his latest research, my standard reaction is 'Eh, maybe.' Then I forget about it. When Robin Hanson tells me about his latest research, my standard reaction is 'No way! Impossible!' Then I think about it for years.

This is my experience too. I think I said my first “No way! Impossible!” sometime around 2008 after reading his blog [Overcoming Bias](#). Since then he's influenced my thinking more than almost anyone else I've ever read. When I heard he was writing a book, I was – well, I couldn't even imagine a book by Robin Hanson. When you read a thousand word blog post by Robin Hanson, you have to sit down and think about it and wait for it to digest and try not to lose too much sleep worrying about it. A whole book would be *something*.

I have now read [Age Of Em](#) ([website](#)) and it is indeed something. Even the cover gives you a weird sense of sublimity mixed with unease:



And in this case, judging a book by its cover is entirely appropriate.

||

Age of Em is a work of futurism – an attempt to predict what life will be like a few generations down the road. This is not a common genre – I can't think of another book of this depth and quality in the same niche. Predicting the future is notoriously hard, and that

seems to have so far discouraged potential authors and readers alike.

Hanson is not discouraged. He writes that:

Some say that there is little point in trying to foresee the non-immediate future. But in fact there have been many successful forecasts of this sort. For example, we can reliably predict the future cost changes for devices such as batteries or solar cells, as such costs tend to follow a power law of the cumulative device production (Nagy et al 2013). As another example, recently a set of a thousand published technology forecasts were collected and scored for accuracy, by comparing the forecasted date of a technology milestone with its actual date. Forecasts were significantly more accurate than random, even forecasts 10 to 25 years ahead. This was true separately for forecasts made via many different methods. On average, these milestones tended to be passed a few years before their forecasted date, and sometimes forecasters were unaware that they had already passed (Charbonneau et al, 2013).

A particularly accurate book in predicting the future was *The Year 2000*, a 1967 book by Herman Kahn and Anthony Wiener. It accurately predicted population, was 80% correct for computer and communication technology, and 50% correct for other technology (Albright 2002). On even longer time scales, in 1900 the engineer John Watkins did a good

job of forecasting many basic features of society a century later (Watkins 1900) [...]

Some say no one could have anticipated the recent big changes associated with the arrival and consequences of the World Wide Web. Yet participants in the Xanadu hypertext project in which I was involved from 1984 to 1993 correctly anticipated many key aspects of the Web [...] Such examples show that one can use basic theory to anticipate key elements of distant future environments, both physical and social, but also that forecasters do not tend to be much rewarded for such efforts, either culturally or materially. This helps to explain why there are relatively few serious forecasting efforts. But make no mistake, it *is* possible to forecast the future.

I think Hanson is overstating his case. All except Watkins were predicting only 10 – 30 years in the future, and most of their predictions were simple numerical estimates, eg “the population will be one billion” rather than complex pictures of society. The only project here even remotely comparable in scope to Hanson’s is [John Watkins’ 1900 article](#).

Watkins is classically given some credit for broadly correct ideas like “Cameras that can send pictures across the world instantly” and “telephones that can call anywhere in the world”, but of his 28 predictions, I judge only eight as even somewhat correct. For example, I grant him a prediction that “the average American will be two inches taller because of good medical care” even though he then

goes on to say in the same sentence that the average life expectancy will be fifty and suburbanization will be so total that building city blocks will be illegal (sorry, John, only in San Francisco). Most of the predictions seem simply and completely false. Watkins believes all animals and insects will have been eradicated. He believes there will be “peas as large as beets” and “strawberries as large as apples” (these are two separate predictions; he is weirdly obsessed with fruit and vegetable size). We will travel to England via giant combination submarine/hovercrafts that will complete the trip in a lightning-fast two days. There will be no surface-level transportation in cities as all cars and walkways have moved underground. The letters C, X, and Q will be removed from the language. Pneumatic tubes will deliver purchases from stores. “A man or woman unable to walk ten miles at a stretch will be regarded as a weakling.”

Where Watkins is right, he is generally listing a cool technology slightly beyond what was available to his time and predicting we will have it. Nevertheless, he is still mostly wrong. Yet this is Hanson’s example of accurate futurology. And he is *right* to make it his example of accurate futurology, because everything else is even worse.

Hanson has no illusions of certainty. He starts by saying that “conditional on my key assumptions, I expect at least 30% of future situations to be usefully informed by my analysis. Unconditionally, I expect at least 10%.” So he is not explicitly overconfident. But in an implicit sense, it’s just *weird* to see the level of detail he tries to predict – for example, he has two pages about what sort of

swear words the far future might use. And the book's style serves to reinforce its weirdness. The whole thing is written in a sort of professorial monotone that changes little from loving descriptions of the sorts of pipes that will cool future buildings (one of Hanson's [pet topics](#)) to speculation on our descendents' romantic relationships (key quote: "The per minute subjective value of an equal relation should not fall much below half of the per-minute value of a relation with the best available open source lover"). And it leans heavily on a favorite Hansonian literary device – the weirdly general statement about something that sounds like it can't possibly be measurable, followed by a curt reference which if followed up absolutely confirms said statement, followed by relentlessly ringing every corollary of it:

Today, mental fatigue reduces mental performance by about 0.1% per minute. As by resting we can recover at a rate of 1% per minute, we need roughly one-tenth of our workday to be break time, with the duration between breaks being not much more than an hour or two (Troughakos and Hideg 2009; Alvanchi et al 2012)... Thus many em tasks will be designed to take about an hour, and many spurs are likely to last for about this duration.

Or:

Today, painters, novelists, and directors who are experimental artists tend to do their best work at roughly ages 46-52, 38-50, and 45-63 respectively, but those ages are 24-34,

29-40, and 27-43, respectively for conceptual artists (Galenson 2006)... At any one time, the vast majority of actual working ems [should be] near a peak productivity subjective age.

Or:

Wars today, like cities, are distributed evenly across all possible war sizes (Cederman 2003).

At some point I started to wonder whether Hanson was putting me on. Everything is just played *too straight*. Hanson even addresses this:

To resist the temptation to construe the future too abstractly, I'll try to imagine a future full of complex detail. One indication that I've been successful in all these efforts will be if my scenario description sounds less like it came from a typical comic book or science fiction movie, and more like it came from a typical history text or business casebook.

Well, count that project a success. The effect is strange to behold, and I'm not sure it will usher in a new era of futurology. But *Age of Em* is great not *just* as futurology, but as a bunch of different ideas and purposes all bound up in a futurological package. For example:

- An introduction to some of the concepts that recur again and again across Robin's thought – for example, [near vs. far mode](#), the

[farmer/forager dichotomy](#), [the inside and outside views](#), [signaling](#).

Most of us learned these through years reading Hanson's blog *Overcoming Bias*, getting each chunk in turn, spending days or months thinking over each piece. Getting it all out of a book you can read in a couple of days sounds really hard – but by applying them to dozens of different subproblems involved in future predictions, Hanson makes the reader more comfortable with them, and I expect a lot of people will come out of the book with an intuitive understanding of how they can be applied. – A whirlwind tour through almost every science and a pretty good way to learn about the *present*. If you didn't already know that wars are distributed evenly across all possible war sizes, well, read *Age of Em* and you will know that and many similar things besides. – A manifesto. Hanson often makes predictions by assuming that since the future will be more competitive, future people are likely to converge toward optimal institutions. This is a dangerous assumption for futurology – it's the same line of thinking that led Watkins to assume English would abandon C, X, and Q as inefficient – but it's a *great* assumption if you want a chance to explain your ideas of optimal institutions to thousands of people who think they're reading fun science-fiction. Thus, Robin spends several pages talking about how ems may use prediction markets – an information aggregation technique he invented – to make their decisions. In the real world, Hanson has been trying to push these for decades, with [varying levels](#) of success. Here, in the guise of a future society, he can expose a whole new group of people to their advantages – as well as the advantages of something called “combinatorial auctions” which I am still not smart enough to understand. – A mind-expanding drug. One of the great risks of futurology is to fail to realize

how *different* societies and institutions can be – the same way uncreative costume designers make their aliens look like humans with green skin. A lot of our thoughts about the future involve assumptions we've never really examined critically, and Hanson dynamites those assumptions. For page after page, he gives strong arguments why our descendants might be poorer, shorter-lived, less likely to travel long distances or into space, less progressive and open-minded. He predicts little noticeable technological change, millimeter-high beings living in cities the size of bottles, careers lasting fractions of seconds, humans being incomprehensibly wealthy patrons to their own robot overlords. And *all of it makes sense*.

When I read Stross' *Accelerando*, one of the parts that stuck with me the longest were the Vile Offspring, weird posthuman entities that operated a mostly-incomprehensible Economy 2.0 that humans just sort of hung out on the edges of, goggle-eyed. It was a weird vision – but, for Stross, mostly a black box. *Age of Em* opens the box and shows you every part of what our weird incomprehensible posthuman descendents will be doing in loving detail. Even what kind of swear words they'll use.



So, what is the Age of Em?

According to Hanson, AI is really hard and won't be invented in time to shape the posthuman future. But sometime a century or so

from now, scanning technology, neuroscience, and computer hardware will advance enough to allow emulated humans, or “ems”. Take somebody’s brain, scan it on a microscopic level, and use this information to simulate it neuron-by-neuron on a computer. A good enough simulation will map inputs to outputs in exactly the same way as the brain itself, effectively uploading the person to a computer. Uploaded humans will be much the same as biological humans. Given suitable sense-organs, effectuators, virtual avatars, or even robot bodies, they can think, talk, work, play, love, and build in much the same way as their “parent”. But ems have three very important differences from biological humans.

First, they have no natural body. They will never need food or water; they will never get sick or die. They can live entirely in virtual worlds in which any luxuries they want – luxurious penthouses, gluttonous feasts, Ferraris – can be conjured out of nothing. They will have some limited ability to transcend space, talking to other ems’ virtual presences in much the same way two people in different countries can talk on the Internet.

Second, they can run at different speeds. While a normal human brain is stuck running at the speed that physics allow, a computer simulating a brain can simulate it faster or slower depending on preference and hardware availability. With enough parallel hardware, an em could experience a subjective century in an objective week. Alternatively, if an em wanted to save hardware it could process all its mental operations *very slowly* and experience only a subjective week every objective century.

Third, just like other computer data, ems can be copied, cut, and pasted. One uploaded copy of Robin Hanson, plus enough free hardware, can become a thousand uploaded copies of Robin Hanson, each living in their own virtual world and doing different things. The copies could even converse with each other, check each other's work, duel to the death, or – yes – have sex with each other. And if having a thousand Robin Hansons proves too much, a quick ctrl-x and you can delete any redundant ems to free up hard disk space for Civilization 6 ([coming out this October!](#))

Would this count as murder? Hanson predicts that ems will have unusually blase attitudes toward copy-deletion. If there are a thousand other copies of me in the world, then going to sleep and not waking up just feels like delegating back to a different version of me. If you're still not convinced, Hanson's essay [Is Forgotten Party Death?](#) is a typically disquieting analysis of this proposition. But whether it's true or not is almost irrelevant – at least *some* ems will think this way, and they will be the ones who tend to volunteer to be copied for short term tasks that require termination of the copy afterwards. If you personally aren't interested in participating, the economy [will leave you behind](#).

The ability to copy ems as many times as needed fundamentally changes the economy and the idea of economic growth. Imagine Google has a thousand positions for Ruby programmers. Instead of finding a thousand workers, they can find one very smart and very hard-working person and copy her a thousand times. With unlimited available labor supply, wages plummet to subsistence levels. “Subsistence levels” for ems are the bare minimum it takes to

rent enough hardware from Amazon Cloud to run an em. The overwhelming majority of ems will exist at such subsistence levels. On the one hand, if you've got to exist on a subsistence level, a virtual world where all luxuries can be conjured from thin air is a pretty good place to do it. On the other, such starvation wages might leave ems with little or no leisure time.

Sort of. This gets weird. There's an urban legend about a "test for psychopaths". You tell someone a story about a man who attends his mother's funeral. He met a really pretty girl there and fell in love, but neglected to get her contact details before she disappeared. How might he meet her again? If they answer "kill his father, she'll probably come to that funeral too", they're a psychopath – ordinary people would have a mental block that prevents them from even considering such a drastic solution. And I bring this up because after reading *Age of Em* I feel like Robin Hanson would be able to come up with some super-solution even the psychopaths can't think of, some plan that gets the man a threesome with the girl and her even hotter twin sister at the cost of wiping out an entire continent. Everything about labor relations in *Age of Em* is like this.

For example, suppose you want to hire an em at subsistence wages, but you want them 24 hours a day, 7 days a week. Ems probably need to sleep – that's hard-coded into the brain, and the brain is being simulated at enough fidelity to leave that in. But jobs with tasks that don't last longer than a single day – for example, a surgeon who performs five surgeries a day but has no day-to-day carryover – can get around this restriction by letting an em have

one full night of sleep, then copying it. Paste the em at the beginning of the workday. When it starts to get tired, let it finish the surgery it's working on, then delete it and paste the well-rested copy again to do the next surgery. Repeat forever and the em never has to get any more sleep than that one night. You can use the same trick to give an em a "vacation" – just give it *one* of them, then copy-paste that brain-state forever.

Or suppose your ems want frequent vacations, but you want them working every day. Let a "trunk" em vacation every day, then make a thousand copies every morning, work all the copies for twenty-four hours, then delete them. Every copy remembers a life spent in constant vacation, and cheered on by its generally wonderful existence it will give a full day's work. But from the company's perspective, 99.9% of the ems in its employment are working at any given moment.

(another option: work the em at normal subjective speed, then speed it up a thousand times to take its week-long vacation, then have it return to work after only one-one-thousandth of a week has passed in real life)

Given that ems exist at subsistence wages, saving enough for retirement sounds difficult, but this too has weird psychopathic solutions. Thousands of copies of the same em can pool their retirement savings, then have all except a randomly chosen one disappear at the moment of retirement, leaving that one with a nest egg thousands of times what it could have accumulated by its own efforts. Or an em can invest its paltry savings in some kind of low-

risk low-return investment and reduce its running speed so much that the return on its investment is enough to pay for its decreased subsistence. For example, if it costs \$100 to rent enough computing power to run an em at normal speed for one year, and you only have \$10 in savings, you can rent 1/1000th of the computer for \$0.10, run at 1/1000th speed, invest your \$10 in a bond that pays 1% per year, and have enough to continue running indefinitely. The only disadvantage is that you'll only experience a subjective week every twenty objective years. Also, since other entities are experiencing a subjective week every second, and some of those entities have nukes, probably there will be some kind of big war, someone will nuke Amazon's data centers, and you'll die after a couple of your subjective minutes. But at least you got to retire!

If ems do find ways to get time off the clock, what will they do with it? Probably they'll have really weird social lives. After all, the existence of em copies is mostly funded by companies, and there's no reason for companies to copy-paste any but the best workers in a given field. So despite the literally trillions of ems likely to make up the world, most will be copies of a few exceptionally brilliant and hard-working individuals with specific marketable talents. Elon Musk might go out one day to the bar with his friend, who is also Elon Musk, and order "the usual". The bartender, who is Elon Musk himself, would know exactly what drink he wants and have it readily available, as the bar caters entirely to people who are Elon Musk. A few minutes later, a few Chesley Sullenbergers might come in after a long day of piloting airplanes. Each Sullenberger would have met hundreds of Musks before and have a good idea about which Musk-Sullenberger conversation topics were most en-

joyable, but they might have to adjust for circumstances; maybe the Musks they met before all branched off a most recent common ancestor in 2120, but these are a different branch who were created in 2105 and remember Elon's human experiences but not a lot of the posthuman lives that shaped the 2120 Musks' worldviews. One Sullenberger might tentatively complain that the solar power grid has too many outages these days; a Musk might agree to take the problem up with the Council of Musks, which is totally a thing that exist (Hanson calls these sorts of groups "copy clans" and says they are "a natural candidate unit for finance, reproduction, legal, liability, and political representation").

Romance could be even weirder. Elon Musk #2633590 goes into a bar and meets Taylor Swift #105051, who has a job singing in a nice local nightclub and so is considered prestigious for a Taylor Swift. He looks up a record of what happens when Elon Musks ask Taylor Swifts out and finds they are receptive on 87.35% of occasions. The two start dating and are advised by the Council of Musks and the Council of Swifts on the issues that are known to come up in Musk-Swift relationships and the best solutions that have been found to each. Unfortunately, Musk #2633590 is transferred to a job that requires operating at 10,000x human speed, but Swift #105051's nightclub runs at 100x speed and refuses to subsidize her to run any faster; such a speed difference makes normal interaction impossible. The story has a happy ending; Swift #105051 allows Musk #2633590 to have her source code, and whenever he is feeling lonely he spends a little extra money to instantiate a high-speed copy of her to hang out with.

(needless to say, these examples are not exactly word-for-word taken from the book, but they're heavily based off of Hanson's more abstract descriptions)

The em world is not just very weird, it's also very very big. Hanson notes that labor is a limiting factor in economic growth, yet even today the economy doubles about once every fifteen years. Once you can produce skilled labor through a simple copy-paste operation, especially labor you can run at a thousand times human speed, the economy will go through the roof. He writes that:

To generate an empirical estimate of em economy doubling times, we can look at the timescales it takes for machine shoper and factories today to make a mass of machines of a quality, quantity, variety, and value similar to that of machines that they themselves contain. Today that timescale is roughly 1 to 3 months. Also, designs were sketched two to three decades ago for systems that might self-replicate nearly completely in 6 to 12 months... these estimates suggest that today's manufacturing technology is capable of self-replicating on a scale of a few weeks to a few months.

Hanson thinks that with further innovation, such times can be reduced so far that "the economy might double every objective year, month, week, or day." As the economy doubles the labor force – ie the number of ems – may double with it, until only a few years after the first ems the population numbers in the trillions. But if the em population is doubling every day, there had better be some

pretty amazing construction efforts going on. The only thing that could possibly work on that scale is prefabricated modular construction of giant superdense cities, probably made mostly out of some sort of proto early-stage computronium (plus cooling pipes). Ems would be reluctant to travel from one such city to another – if they exist at a thousand times human speed, a trip on a hypersonic airliner that could go from New York to Los Angeles in an hour would still take forty subjective days. Who wants to be on an airplane for forty days?

(long-distance trade is also rare, since if the economy doubles fast enough it means that by the time goods reach their destination they could be almost worthless)

The real winners of this ultra-fast-growing economy? Ordinary humans. While humans will be way too slow and stupid to do anything useful, they will tend to have non-subsistence amounts of money saved up from their previous human lives, and also be running at speeds thousands of times slower than most of the economy. When the economy doubles every day, so can your bank account. Ordinary humans will become rarer, less relevant, but fantastically rich – a sort of doddering Neanderthal aristocracy spending sums on a cheeseburger that could support thousands of ems in luxury for entire lifetimes. While there will no doubt be pressure to liquidate humans and take their stuff, Hanson hopes that the spirit of rule of law – the same spirit that protects rich minority groups today – will win out, with rich ems reluctant to support property confiscation lest it extend to them also. Also, em retirees will have incentives a lot like humans – they have saved up money and

go really slow – and like AARP members today they may be able to obtain disproportionate political power which will then protect the interests of slow rich people.

But we might not have much time to enjoy our sudden rise in wealth. Hanson predicts that the Age of Em will last for subjective em millennia – ie about one to two actual human years. After all, most of the interesting political and economic activity is going on at em timescales. In the space of a few subjective millennia, either someone will screw up and cause the apocalypse, somebody will invent real superintelligent AI that causes a technological singularity, or some other weird thing will happen taking civilization beyond the point that even Robin dares to try to predict.

IV

Hanson understands that people might not like the idea of a future full of people working very long hours at subsistence wages forever (Zack Davis' [Contract-Drafting Em](#) song is, as usual, relevant). But Hanson himself does not view this future as dystopian. Despite our descendents' by-the-numbers poverty, they will avoid the miseries commonly associated with poverty today. There will be no dirt or cockroaches in their sparkling virtual worlds, nobody will go hungry, petty crime will be all-but-eliminated, and unemployment will be low. Anybody who can score some leisure time will have a dizzying variety of hyperadvanced entertainment available, and as for the people who can't, they'll mostly have been copied from people who really like working hard and don't miss it anyway. As unhappy as we

moderns may be contemplating em society, ems themselves will not be unhappy! And as for us:

The analysis in this book suggests that lives in the next great era may be as different from our lives as our lives are from farmers' lives, or farmers' lives are from foragers' lives. Many readers of this book, living industrial era lives and sharing industrial era values, may be disturbed to see a forecast of em era descendants with choices and lifestyles that appear to reject many of the values that they hold dear. Such readers may be tempted to fight to prevent the em future, perhaps preferring a continuation of the industrial era. Such readers may be correct that rejecting the em future holds them true to their core values. But I advise such readers to first try hard to see this new era in some detail from the point of view of its typical residents. See what they enjoy and what fills them with pride, and listen to their criticisms of your era and values.

A short digression: there's a certain strain of thought I find infuriating, which is "My traditionalist ancestors would have disapproved of the changes typical of my era, like racial equality, more open sexuality, and secularism. But I am smarter than them, and so totally okay with how the future will likely have values even more progressive and shocking than my own. Therefore I pre-approve of any value changes that might happen in the future as definitely good and better than our stupid hidebound present."

I once read a science-fiction story that depicted a pretty average sci-fi future – mighty starships, weird aliens, confederations of planets, post-scarcity economy – with the sole unusual feature that rape was considered totally legal, and opposition to such as bigoted and ignorant as opposition to homosexuality is today. Everybody got really angry at the author and said it was offensive for him to even speculate about that. Well, that’s the method by which our cheerful acceptance of any possible future values is maintained: restricting the set of “any possible future values” to “values slightly more progressive than ours” and then angrily shouting down anyone who discusses future values that actually sound bad. But of course the *whole question* of how worried to be about future value drift *only makes sense* in the context of future values that genuinely violate our current values. Approving of all future values except ones that would be offensive to even speculate about is the same faux-open-mindedness as [tolerating anything except the outgroup](#).

Hanson deserves credit for positing a future whose values are likely to upset even the sort of people who say they don’t get upset over future value drift. I’m not sure whether or not he deserves credit for not being upset by it. Yes, it’s got low-crime, ample food for everybody, and full employment. But so does *Brave New World*. The whole *point* of dystopian fiction is pointing out that we have complicated values beyond material security. Hanson is absolutely right that our traditionalist ancestors would view our own era with as much horror as some of us would view an em era. He’s even right that on utilitarian grounds, it’s hard to argue with an em era where everyone is really happy working eighteen hours a day for their entire lives because we selected for people who feel that way.

But at some point, can we make the Lovecraftian argument of “I know my values are provincial and arbitrary, but they’re *my* provincial arbitrary values and I will make any sacrifice of blood or tears necessary to defend them, even unto the gates of Hell?”

This brings us to an even worse scenario.

There are a lot of similarities between Hanson’s futurology and (my possibly erroneous interpretation of) the futurology of Nick Land. I see Land as saying, like Hanson, that the future will be one of quickly accelerating economic activity that comes to dominate a bigger and bigger portion of our descendents’ lives. But whereas Hanson’s framing focuses on the participants in such economic activity, playing up their resemblances with modern humans, Land takes a bigger picture. He talks about the economy itself acquiring a sort of self-awareness or agency, so that the destiny of civilization is consumed by the imperative of economic growth.

Imagine a company that manufactures batteries for electric cars. The inventor of the batteries might be a scientist who really believes in the power of technology to improve the human race. The workers who help build the batteries might just be trying to earn money to support their families. The CEO might be running the business because he wants to buy a really big yacht. And the whole thing is there to eventually, somewhere down the line, let a suburban mom buy a car to take her kid to soccer practice. Like most companies the battery-making company is primarily a profit-making operation, but the profit-making-ness draws on a lot of not-purely-economic actors and their not-purely-economic subgoals.

Now imagine the company fires all its employees and replaces them with robots. It fires the inventor and replaces him with a genetic algorithm that optimizes battery design. It fires the CEO and replaces him with a superintelligent business-running algorithm. All of these are good decisions, from a profitability perspective. We can absolutely imagine a profit-driven shareholder-value-maximizing company doing all these things. But it reduces the company's non-masturbatory participation in an economy that points outside itself, limits it to just a tenuous connection with soccer moms and maybe some shareholders who want yachts of their own.

Now take it further. Imagine there are no human shareholders who want yachts, just banks who lend the company money in order to increase their own value. And imagine there are no soccer moms anymore; the company makes batteries for the trucks that ship raw materials from place to place. Every non-economic goal has been stripped away from the company; it's just an appendage of Global Development.

Now take it even further, and imagine this is what's happened everywhere. There are no humans left; it isn't economically efficient to continue having humans. Algorithm-run banks lend money to algorithm-run companies that produce goods for other algorithm-run companies and so on ad infinitum. Such a masturbatory economy would have all the signs of economic growth we have today. It could build itself new mines to create raw materials, construct new roads and railways to transport them, build huge factories to manufacture them into robots, then sell the robots to whatever companies need more robot workers. It might even eventually invent

space travel to reach new worlds full of raw materials. Maybe it would develop powerful militaries to conquer alien worlds and steal their technological secrets that could increase efficiency. It would be vast, incredibly efficient, and utterly pointless. The real-life incarnation of those strategy games where you mine Resources to build new Weapons to conquer new Territories from which you mine more Resources and so on forever.

But this seems to me the natural end of the economic system. Right now it needs humans only as laborers, investors, and consumers. But robot laborers are potentially more efficient, companies based around algorithmic trading are already pushing out human investors, and most consumers already aren't individuals – they're companies and governments and organizations. At each step you can gain efficiency by eliminating humans, until finally humans aren't involved *anywhere*.

True to form, Land doesn't see this as a dystopia – I think he conflates “maximally efficient economy” with “God”, which is a *hell* of a thing to conflate – but I do. And I think it provides an important new lens with which to look at the Age of Em.

The Age of Em is an economy in the early stages of such a transformation. Instead of being able to replace everything with literal robots, it replaces them with humans who have had some aspects of their humanity stripped away. Biological bodies. The desire and ability to have children normally. Robin doesn't think people will lose all leisure time and non-work-related desires, but he doesn't

seem too sure about this and it doesn't seem to bother him much if they do.

I envision a spectrum between the current world of humans and Nick Land's Ascended Economy. Somewhere on the spectrum we have ems who get leisure time. A little further on the spectrum we have ems who don't get leisure time.

But we can go further. Hanson imagines that we can "tweak" em minds. We may not understand the brain enough to create totally new intelligences from the ground up, but by his Age of Em we should understand it well enough to make a few minor hacks, the same way even somebody who doesn't know HTML or CSS can usually figure out how to change the background color of a webpage with enough prodding. Many of these mind tweaks will be the equivalent of psychiatric drugs – some might even be computer simulations of what we observe to happen when we give psychiatric drugs to a biological brain. But these tweaks will necessarily be much stronger and more versatile, since we no longer care about bodily side effects (ems don't have bodies) and we can apply it to only a single small region of the brain and avoid actions anywhere else. You could also very quickly advance brain science – the main limits today are practical (it's really hard to open up somebody's brain and do stuff to it without killing them) and ethical (the government might have some words with you if you tried). An Age of Em would remove both obstacles, and give you the added bonus of being able to make thousands of copies of your test subjects for randomized controlled trials, reloading any from a saved copy if they died. Hanson envisions that:

As the em world is a very competitive world where sex is not needed for reproduction, and as sex can be time and attention-consuming, ems may try to suppress sexuality, via mind tweaks that produce effects analogous to castration. Such effects might be temporary, perhaps with a consciously controllable on-off switch... it is possible that em brain tweaks could be found to greatly reduce natural human desires for sex and related romantic and intimate pair bonding without reducing em productivity. It is also possible that many of the most productive ems would accept such tweaks.

Possible? I can do that *right now* with a high enough dose of Paxil, and I don't even have to upload your brain to a computer first. Fun stories about Musk #2633590 and Swift #105051 aside, I expect this would happen about ten minutes after the advent of the Age of Em, and we would have taken another step down the path to the Ascended Economy.

There are dozens of other such tweaks I can think of, but let me focus on two.

First, stimulants have a very powerful ability to focus the brain on the task at hand, as anybody who's taken Adderall or modafinil can attest. Their main drawbacks are addictiveness and health concerns, but in a world where such pills can be applied as mental tweaks, where minds have no bodies, and where any mind that gets too screwed up can be reloaded from a backup copy, these are barely concerns at all. Many of the purely mental side effects of stimulants come from their effects in parts of the brain not vital

to the stimulant effect. If we can selectively apply Adderall to certain brain centers but not others, then unapply it at will, then from employers' point of view there's no reason not to have all workers dosed with superior year 2100 versions of Adderall at all times. I worry that not only will workers not have any leisure time, but they'll be neurologically incapable of having their minds drift off while on the job. Davis' contract-drafting em who starts wondering about philosophy on the job wouldn't get terminated. He would just have his simulated-Adderall dose increased.

Second, Robin managed to write an entire book about emulated minds without using the word ["wireheading"](#). This is another thing we can do right now, with today's technology – but once it's a line of code and not a costly brain surgery, it should become nigh-universal. Give ems the control switches to their own reward centers and all questions about leisure time become irrelevant. Give bosses the control switches to their employees' reward centers, and the situation changes markedly. Hanson says that there probably won't be too much slavery in the em world, because it will likely have strong rule of law, because slaves aren't as productive as free workers, and there's little advantage to enslaving someone when you could just pay them subsistence wages anyway. But slavery isn't *nearly* as abject and inferior a condition as the one where somebody else has the control switch to your reward center. Combine that with the stimulant use mentioned above, and you can have people who will never have nor want to have any thought about anything other than working on the precise task at which they are supposed to be working at any given time.

This is something I worry about even in the context of normal biological humans. But Hanson already believes em worlds will have few regulations and be able to ignore the moral horror of 99% of the population by copying and using the 1% who are okay with something. Combine this with a situation where brains are easily accessible and tweakable, and this sort of scenario becomes horribly likely.

I see almost no interesting difference between an em world with full use of these tweaks and an Ascended Economy world. Yes, there are things that look vaguely human in outline laboring in the one and not the other, but it's not like there will be different thought processes or different results. I'm not even sure what it would mean for the ems to be conscious in a world like this – they're not doing anything interesting with the consciousness. The best we could say about this is that if the wireheading is used liberally it's a lite version of the world where everything gets converted to [hedonium](#).

V

In a book full of weird ideas, there is only one idea rejected as too weird. And in a book written in a professorial monotone, there's only one point at which Hanson expresses anything like emotion:

Some people foresee a rapid local “intelligence explosion” happening soon after a smart AI system can usefully modify its local architecture (Chalmers 2010; Hanson and Yud-

kowsky 2013; Yudkowsky 2013; Bostrom 2014)... Honestly to me this local intelligence explosion scenario looks suspiciously like a super-villain comic book plot. A flash of insight by a lone genius lets him create a genius AI. Hidden in its super-villain research lab lair, this guines villain AI works out unprecedented revolutions in AI design, turns itself into a super-genius, which then invents super-weapons and takes over the world. Bwa ha ha.

For someone who just got done talking about the sex lives of uploaded computers in millimeter-tall robot bodies running at 1000x human speed, Robin is sure quick to use the absurdity heuristic to straw-man intelligence explosion scenarios as “comic book plots”. Take away his weird authorial tic of using the words “genius” and “supervillain”, this scenario reduces to “Some group, perhaps Google, perhaps a university, invent an artificial intelligence smart enough to edit its own source code; exponentially growing intelligence without obvious bound follows shortly thereafter”. Yes, it’s weird to think that there may be a sudden quantum leap in intelligence like this, but no weirder than to think most of civilization will transition from human to em in the space of a year or two. I’m a little bit offended that this is the only idea given this level of dismissive treatment. Since I do have immense respect for Robin, I hope my offense doesn’t color the following thoughts too much.

Hanson’s arguments against AI seem somewhat motivated. He admits that AI researchers generally estimate less than 50 years before we get human-level artificial intelligence, a span shorter than his estimate of a century until we can upload ems. He even admits

that no AI researcher thinks ems are a plausible route to AI. But he dismisses this by saying when he asks AI experts informally, they say that in their own field, they have only noticed about 5-10% of the progress they expect would be needed to reach human intelligence over the past twenty years. He then multiplies out to say that it will probably take at least 400 years to reach human-level AI. I have two complaints about this estimate.

First, he is explicitly ignoring published papers surveying hundreds of researchers using validated techniques, in favor of what he describes as “meeting experienced AI experts informally”. But even though he feels comfortable rejecting vast surveys of AI experts as potentially biased, as best I can tell he does not ask a single neuroscientist to estimate the date at which brain scanning and simulation might be available. He just says that “it seems plausible that sufficient progress will be made in roughly a century or so”, citing a few hopeful articles by very enthusiastic futurists who are not neuroscientists or scanning professionals themselves and have not talked to any. This seems to me to be an extreme example of [isolated demands for rigor](#). No matter how many AI scientists think AI is soon, Hanson will cherry-pick the surveying procedures and results that make it look far. But if a few futurists think brain emulation is possible, then no matter what anybody else thinks that’s good enough for him.

Second, one would expect that even if there were only 5-10% progress over the last twenty years, then there would be faster progress in the future, since the future will have a bigger economy, better supporting technology, and more resources invested in AI re-

search. Robin answers this objection by saying that “increases in research funding usually give much less than proportionate increases in research progress” and cites Alston et al 2011. I looked up Alston et al 2011, and it is a paper relating crop productivity to government funding of agriculture research. There was no attempt to relate its findings to any field other than agriculture, nor to any type of funding other than government. But [studies show](#) that while public research funding often does have minimal effects, the effect of private research funding is usually much larger. A single sentence citing a study in crop productivity to apply to artificial intelligence while ignoring much more relevant results that contradict it seems like a really weak argument for a statement as potentially surprising as “amount of research does not affect technological progress”.

I realize that Hanson has done a lot more work on this topic and he couldn't fit all of it in this book. I disagree with his other work too, and I've said so elsewhere. For now I just want to say that the arguments in this book seem weak to me.

I also want to mention what seems to me a very Hansonian counterargument to the ems-come-first scenario: we have always developed de novo technology before understanding the relevant biology. We built automobiles by figuring out the physics of combustion engines, not by studying human muscles and creating mechanical imitations of myosin and actin. Although the Wright brothers were inspired by birds, their first plane was not an ornithopter. Our power plants use coal and uranium instead of the Krebs Cycle. Biology is *really hard*. Even slavishly *copying* biology is really hard. I don't

think Hanson and the futurists he cites understand the scale of the problem they've set themselves.

Current cutting-edge brain emulation projects have found their work much harder than expected. Simulating a nematode is pretty much the rock-bottom easiest thing in this category, since they are tiny primitive worms with only a few neurons; the history of the field is a [litany of failures](#), with current leader [OpenWorm](#) “reluctant to make bold claims about its current resemblance to biological behavior”. A more ambitious \$1.3 billion attempt to simulate a tiny portion of a rat brain has gone down in history as a [legendary failure](#) (politics were involved, but I expect they would be involved in a plan to upload a human too). And these are just attempts to get something that behaves *vaguely* like a nematode or rat. Actually uploading a human, keeping their memory and personality intact, and not having them go insane afterwards boggles the mind. We're still not sure how much small molecules matter to brain function, how much glial cells matter to brain function, how many things in the brain are or aren't local. AI researchers are making programs that can defeat chess grandmasters; upload researchers are still struggling to make a worm that will wriggle. The right analogy for modern attempts to upload human brains isn't modern attempts at designing AI. It's an attempt at designing AI by someone who doesn't even know how to plug in a computer.

VI

I guess what really bothers me about Hanson's pooh-poohing of AI is him calling it "a comic book plot". To me, it's Hanson's scenario that seems science-fiction-ish.

I say this not as a generic insult but as a pointer at a specific category of errors. In *Star Wars*, the Rebellion had all of these beautiful hyperspace-capable starfighters that could shoot laser beams and explore galaxies – and *they still had human pilots*. 1977 thought the pangalactic future would still be using people to pilot its military aircraft; in reality, even 2016 is moving away from this.

Science fiction books have to tell interesting stories, and interesting stories are about humans or human-like entities. We can enjoy stories about aliens or robots as long as those aliens and robots are still approximately human-sized, human-shaped, human-intelligence, and doing human-type things. A *Star Wars* in which all of the X-Wings were combat drones wouldn't have done anything for us. So when I accuse something of being science-fiction-ish, I mean bending over backwards – and ignoring the evidence – in order to give basically human-shaped beings a central role.

This is my critique of Robin. As weird as the Age of Em is, it makes sure never to be weird in ways that warp the fundamental humanity of its participants. Ems might be copied and pasted like so many .JPGs, but they still fall in love, form clans, and go on vacations.

In contrast, I expect that we'll get some kind of AI that will be totally inhuman and much harder to write sympathetic stories about. If we get ems after all, I expect them to be lobotomized and drugged

until they become *effectively* inhuman, cogs in the Ascended Economy that would no more fall in love than an automobile would eat hay and whinny. Robin's interest in keeping his protagonists relatable makes his book fascinating, engaging, and probably wrong.

I almost said "and probably less horrible than we should actually expect", but I'm not sure that's true. With a certain amount of horror-suppressing, the Ascended Economy can be written off as morally neutral – either having no conscious thought, or stably wire-headed. All of Robin's points about how normal non-uploaded humans should be able to survive an Ascended Economy at least for a while seem accurate. So morally valuable actors might continue to exist in weird Amish-style enclaves, living a post-scarcity lifestyle off the proceeds of their investments, while all the while the Ascended Economy buzzes around them, doing weird inhuman things that encroach upon them not at all. This seems slightly worse than a Friendly AI scenario, but much better than we have any right to expect of the future.

I highly recommend *Age of Em* as a fantastically fun read and a great introduction to these concepts. It's engaging, readable, and *weird*. I just don't know if it's weird *enough*.