Look Hard, Then Steer Slightly, by Robin Hanson

Humanity can best steer its future by working hard to clearly see the future it will have if we do nothing. Because most likely we will do little to steer our future. Yes, this answer frustrates our hunger for inspiring visions. Even so, it seems right. Let me explain.

Imagine you are holding on to a log, floating down the rapids of a wide fast murky river at night. You hear rough water ahead. How should you steer yourself?

You should **not** try to figure out what river you'd most rather be on, or what landscape you wished the river flowed through. Instead, you should focus on details of the actual river in front of you. You should also not just swim for the best looking spot in the river ahead; in a wide fast river you probably can't get most places.

What you should do is, keeping in mind your limited stamina and abilities, look to see the places ahead where you could plausibly swim. See them as clearly as possible, and try to infer what might be just under the water where you cannot see. Don't immediately swim before you look, but also don't wait too long before starting a plan.

Steering humanity's future is like swimming this river. It is way too fun and easy to assume that we can create any future world we can imagine. Yes the future is made by the sum total of all our actions, but we actually have very limited abilities to coordinate those actions, abilities that get worse on larger space and time scales. We don't have a world government, and won't anytime soon. The organizations we do have, they rarely plan more than a decade ahead.

Given our limited abilities to influence the future, our first priority must be to see as clearly as possible the likely outcomes if we do absolutely nothing. After all, the world today is very nearly what it would be if our distant ancestors had done nothing to try to influence it. And the future world will likely be similar.

Yes, science fiction is full of stories of a few foresighted heroes swinging the tide of their civilization. And yes, inspiring speakers often rouse audiences to cheer by framing their causes as ways to help the future. But honestly, people are mostly moved to action by the world around them, not the distant future.

Seeing the future in enough detail does seem the hard part; deciding what to do given any specific vision seems easier. For example, if you see in the river ahead a sharp rock a bit off to the left, you should swim to the right. Seeing the rock is hard; deciding which way to swim is easy.

True, it may feel more inspiring to think about how you'd want to restructure the whole river landscape. But focusing on the rocks straight ahead is the best way to avoid smashing against them.

Yes, seeing the future is hard, but it is hardly impossible. Recently a set of 1018 published tech forecasts were collected and scored for accuracy, by comparing the forecasted date of each tech milestone with the date it actually happend. Forecasts were significantly more accurate than random, even forecasts 10 to 25 years ahead, and even forecasts made by many different methods. On average events tended to occur a few years before their forecasted date, and sometimes forecasters were unaware that the event had already happened when they made their forecast (Mullins 2012).

It has also been possible to foresee tech on longer time scales. For example, in 1900 engineer John Watkins did a good job of forecasting many basic features of society a century later (Watkins 1900). In 1911, Konstantin Tsiolkovsky foresaw the basic issues and possibilities of space travel. More recently, K. Eric Drexler used basic physical science to envision the outlines of nanotech-based manufacturing (Drexler 1992). Others have used basic physical science to envision the outlines of starships (Benford & Benford 2013). While real nanotech and starships will surely differ in many ways from these outlines, these efforts are far better than random guides to future nanotech and starships, and better than the vast majority of science fiction movies.

Many say that while physical technologies can be foreseen, their social consequences cannot. But most such folks trained in engineering or physical sciences; they don't get that we social scientists do in fact know many useful things. For example, some say no one could have anticipated the recent big changes due to the arrival of the world wide web. Yet the Xanadu hypertext project in which I was involved from 1984 to 1993 correctly anticipated many key aspects of the web. And a 1999 business book used basic economics to accurately forecast key internet business issues (Shapiro & Varian 1999).

In my current book project (Hanson 2008; 2014), I offer an unprecedented quantity of credible detail on the social implications of a particular future tech: brain emulations. I hope that this quantity of detail can help to more clearly debunk this problematic myth that we can't foresee big tech-driven social change.

Brain emulations, also known as "uploads" or "ems", have been a staple of science fiction and tech futurism for decades. To make a brain emulation, one takes a particular human brain, scans it to record its particular cell features and connections, and then builds a computer model that processes signals according to those same features and connections. A good enough em has very close to the same overall input-output signal behavior as the original human. One might talk with it, and convince it to do useful jobs.

Like humans, ems would remember a past, are aware of a present, and anticipate a future. Ems can be happy or sad, eager or tired, fearful or hopeful, proud or shamed, creative or derivative, compassionate or cold. Ems can learn, and have friends, lovers, bosses, and colleagues. While em psychological features may differ from the human average, they are usually near the range of human variation.

I focus on ems because I think they will probably be the first kind of human level general artificial intelligence [AI]. Yes some AI experts are optimistic that they can write human-level AI software within a few decades. But overall, AI experts are much less optimistic when asked about what they should know best: recent progress in the AI subfield where they have the most expertize.

I was a professional AI researcher for nine years (1984-1993), and when I meet other such experienced AI experts informally, I am in the habit of asking them how much progress they have seen in their specific AI subfield in the last twenty years. They typically say they have only seen 5-10% of the progress required to achieve human level AI. They have also typically seen no noticeable acceleration over this period (Hanson 2012). At this rate of progress, it would take two to four centuries for most AI subfields to on average reach human level abilities.

In contrast, the three technologies required to create ems, scanning, modeling, and computing, all seem likely to be ready within a century, at which time we would be less than one quarter to one half of the way to writing human-level AI software. This is why I say that ems will probably make our first human-level AI.

My analysis of the early em era paints a picture that is disturbing and alien to many. The population of ems would quickly explode toward trillions, driving em wages down to near em subsistence levels, work hours up to fill most waking hours, and economic doubling times down to a month or less. Most ems would be copies of less than a thousand very smart, conscientious, and productive humans. Most ems would be near a subjective peak productivity age of sixty or more, and most would also be copies make to do a short term task and then end when that tasks is done.

Ems would cram in a few very tall cities packed dense with hot computer hardware. Ems would leisure in virtual reality, and most ems would work there as well. Em virtual reality would be of a spectacular quality, and ems would have fantastic virtual bodies that never need feel hunger, cold, grime, pain, or sickness. Since the typical em would run a thousand or more times faster than humans, their world would seem more stable to them than ours seems to us, and Earth would seem vast.

Ems would often spin off copies to do short term tasks and then end when those tasks are done. After a subjective career lasting perhaps a century or two, em minds would become less flexible and no longer compete well with younger minds. Such ems would then retire to an indefinite life of leisure at a slower speed.

The ease of making copies of ems would make preparation easier. One em could conceive of a software or artistic design and vision, and then split into an army who execute that vision. Big projects could be completed more often on time if not on budget by speeding up the ems who work on lagging parts. One em could be trained to do a job, with many copies then made of that trained em. Em labor markets would thus be more like our product markets, dominated by a few main suppliers.

Ems would be more unequal than we are, both because em speeds could vary, and because longer lifespans let unequal outcomes accumulate. Ems would split by speed into status castes, with faster ems being higher status. Em democracies would probably use speed-weighted voting, and em rulers would usually run faster than subordinates, to more easily coordinate bigger organizations. Em organizations may also use new governance methods like decision markets and combinatorial auctions.

Each em would feel strongly attached to its clan of copies all descended from the same original human. Em clans may self-govern and negotiate with other clans for the legal rules to apply to disputes with them. Clans may give members continual advice based on the life experiences of similar clan members.

To allow romantic relations when there is unequal demand for male vs. female em workers, the less demanded gender may run slower, and periodically speed up to meet with faster mates. Fast ems with physical robotic bodies would have proportionally smaller bodies; a typical thousand times human speed em would stand two millimeters tall. Most long distance physical travel would be via "beam me up" electronic travel, done with care to avoid mind theft.

Em cities are likely inhospitable to ordinary humans, who, controlling most of the rest of the Earth, mostly live a comfortable retirement on their em-economy investments. While ems could easily buy the rest of the Earth, they don't care enough to bother, beyond ensuring energy, raw materials, and cooling for em cities. Just as we rarely kill our retirees and take their stuff, there is a reasonable hope that ems might leave retired humanity in peace.

Over the longer run, the main risk to humans and nature is probably em civilization instabilities like wars or revolutions. Ems running a thousand times faster than humans might fit a few millennia of history into a few objective years. Slow em retirees would be allies to help humans to promote stability.

Okay, now that we have a clearer vision of the river ahead, how can we realistically steer it, given our very limited sphere of influence?

The easiest advice to give helps individuals and small groups, and requires little larger coordination. First, diversify both your financial and social assets, especially away from your abilities to earn wages, and toward the industries and locations most likely to host the new em economy. Second, teach your descendants to hope to start one of the most copied em clans. Realizing the odds are greatly against you, take great risks to achieve and show your high and reliable productivity and flexibility, in em-world-like tasks and environments. Learn to get along well with people much like yourself, and to value life when it is hard and alien.

To those who hope to coordinate on larger scales to influence the em world, the most solid advice is that it is probably easier for you to influence the early transition period than the equilibrium that will follow it. Work to make computers, and not brain cell modeling, be the last enabling technology to be ready. Help make the transition smooth, peaceful, and anticipated. Help ems to share their financial, legal, and political institutions with ordinary humans, and help make ems grateful for human support, instead of resentful against human resistance. Help all humans to be tested early for their suitability as founders of em clans.

Yes these are rather modest ambitions, and we many only achieve them to tiny degrees. But the parable of the river warns against thinking we can remake our river and its landscape; better to focus on the tiny feasible swims one could actually manage.

Perhaps my description here horrifies you so much that you want to ignore my warning and work to prevent this em world from arriving, and in essence disown these your em children. I can't say you are wrong, but I can say that parents who feel inclined to disown their kids should be encouraged to study their kids carefully before making such a choice. Please try 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. See them clearly, warts and all. Then decide.

Gregory Benford, James Benford (2013) *Starship Century: Toward the Grandest Horizon*, April.

K. Eric Drexler (1992) *Nanosystems: Molecular Machinery, Manufacturing, and Computation,* Wiley, October.

Robin Hanson (2008) Economics of the Singularity, *IEEE Spectrum*, 37-42, June 2008.

Robin Hanson (2012) AI Progress Estimate, *Overcoming Bias* blog, August 27, http://www.overcomingbias.com/2012/08/ai-progress-estimate.html .

Robin Hanson (2014) When the Economy Transcends Humanity, *The Futurist*, 27-30, January-February.

Carie Mullins (2012) Retrospective Analysis of Technology Forecasting: In-Scope Extension, Final Report by Tauri Group to U.S. Office of Secretary of Defense, August 13. http://www.dtic.mil/get-tr-doc/pdf?AD=ADA568107

Carl Shapiro, Hal Varian (1999). *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press.

John Watkins Jr. (1900) What May Happen In The Next Hundred Years. *Ladies' Home Journal*, p.8, December.